Conformational Study of Small Peptides Using Spectroscopy and Quantum Chemistry Calculations

A report submitted to Indian Institute of Science Education and Research Pune in partial fulfillment of the requirements for the Integrated BS-MS programme

Supervisor: Prof. Aloke Das (IISER, Pune)

By

Kaarthik RS

20151014

21/05/2020

Indian Institute of Science Education and Research Pune

Maharashtra, 411008, India

Certificate

This is to certify that this dissertation entitled “Conformational study of small peptides using Spectroscopy and quantum chemistry calculations" towards the partial fulfillment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by “Kaarthik RS at Indian Institute of Science Education and Research, Pune” under the supervision of “Prof. Aloke Das, Professor, Department of Chemistry” during the academic year 2019-2020.

21/05/2020

Prof. Aloke Das

(Thesis Supervisor)

Declaration

I hereby declare that the matter embodied in the report entitled “Conformational study of small peptides using Spectroscopy and quantum chemistry calculations” are the results of the work carried out by me at the Department of Chemistry, Indian Institute of Science Education and Research, Pune, under the supervision of Prof. Aloke Das and the same has not been submitted elsewhere for any other degree.

21/05/2020 **Kaarthik RS**

20151014

Acknowledgment

I would like to thank my parents for their love, kindness, and constant motivation, support, and blessings.

I express my gratitude towards my supervisor, Prof. Aloke Das, for providing me an opportunity to carry out the project in his lab, considering me a part of the lab and giving timely suggestions during the project.

I would like to thank my lab-seniors, Dr. Kamal Kumar Mishra, for teaching me the basics, explaining the concepts patiently. I would like to thank Mr. Satish Kumar, Mr. Kshetrimayum Borish, for teaching me all the experiment related concepts, clearing my doubts, and giving suggestions throughout my project. I thank Mr. Surajit Metya for helping me with the software issues and motivating me during my time in the lab. I also thank Mr. Prakash Panwaria for being more concerned about my work, always reminding me of my duties, and amazing me with his extensive expertise.

I would like to thank my friends Pawandeep Singh, Jaideep Mahajan, for being there for me, waiting, and motivating me towards my project.

Table of Content

Contents

List of Figures and Tables

Abstract

Structure-function relationship found in protein helps us to understand the function that a protein is capable of by knowing its structure. Understanding the secondary structure of a small peptide will give more information on the structure of the protein, which can be used to understand the functions of the protein in detail. Apart from covalent bonds, these secondary structures are shaped by non-covalent interactions like hydrogen bonds, hydrophobic interactions, and n→π* interactions. Though these non-covalent interactions are individually weak, many such interactions cumulatively contribute to stabilizing the conformation of a peptide. Gas-phase studies on small peptides are carried out to study such non-covalent interactions that are stabilizing the peptides when there is no effect from the solvent medium or any other nearby molecules.

1. Introduction

1.1. Proteins

The human body has carbohydrates, proteins, fats, and water as the major components.[2] Proteins act as the building block, giving shape to the human body.[3] They also carry other processes like i) catalyzing reactions (by acting as enzymes), ii) transporting signals in and out of the cell. Proteins are made of small building units called the amino acids.[4] The protein structure, on the whole, has many interactions involved in stabilizing the structure, hence understanding its 3-dimensional structure becomes tedious. Protein structure can be studied by studying its substructures, namely: i) primary structure, ii) secondary structure, iii) tertiary structure, and iv)

quaternary structure. The 20 essential amino acids and other unnatural amino acids bind to each other through amide bonds forming the linear primary structure of the amino acid sequence. The secondary structure is formed when the primary structure changes its orientation in three-dimensional space, bringing the side chain groups closer, forming hydrogen bonds. The geometry of the secondary structure is held throughout the protein structure. The secondary structure consists only of a small part of the whole amino acid sequence. When the complete sequence changes in the geometry orientation, they form the tertiary structure. Different tertiary structures come closer and form a quaternary structure that makes up the fully functional protein structure.

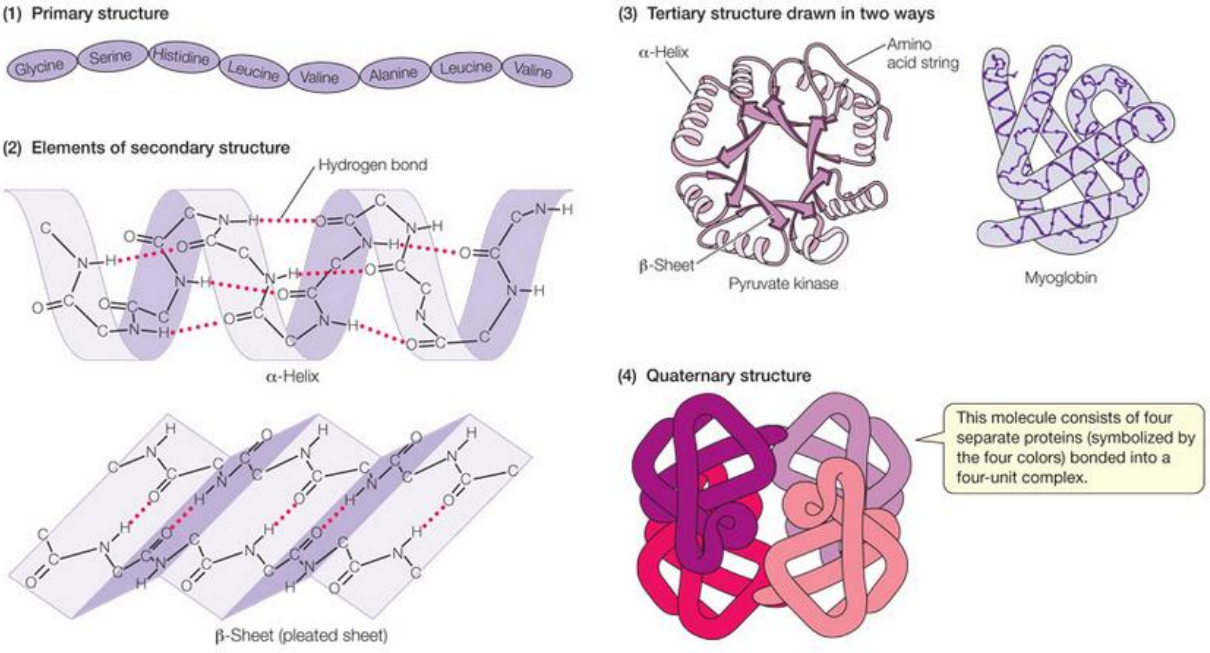

Figure 1: *Different levels of protein structure.*[1]

1.2 Secondary Structure of protein

The geometry of the secondary structure governs the geometry of the protein structure, which plays a vital role in the functions they carry out.[5,6] Understanding the geometry of secondary structure thus helps us in understanding the structure-function relationship.[7] The common types of secondary structure are the α-helix, the β-sheets,

and the turns.[8] The α-helix is the most common secondary structure,[9] where the amino acid sequence forms a helical geometry having hydrogen bonds between the carbonyl and the amide group (C=O…N-H) of every i and i+4th residue, respectively. 3^{10} helix[10,11] and n-helix[12] are the other common helix type of secondary structures observed.

The β-sheet is the second major secondary structure,[13] in which a linear sequence of amino acid residue forms multiple hydrogen bonds with another linear sequence, resulting in a sheet-like structure. The linear sequences can be parallel or antiparallel to each other. The terminal functional groups determine the direction of a sequence. A sequence starts with N-terminal (amide group) and ends at the C-terminal (carbonyl group). When two antiparallel sequences form a sheet, the secondary structure is called a β-hairpin.[14] The loop connecting the sequences is the turn. The sequence of a turn varies between proteins, but if they carry out a specific function, then the sequence will be maintained[15] along with the structure.

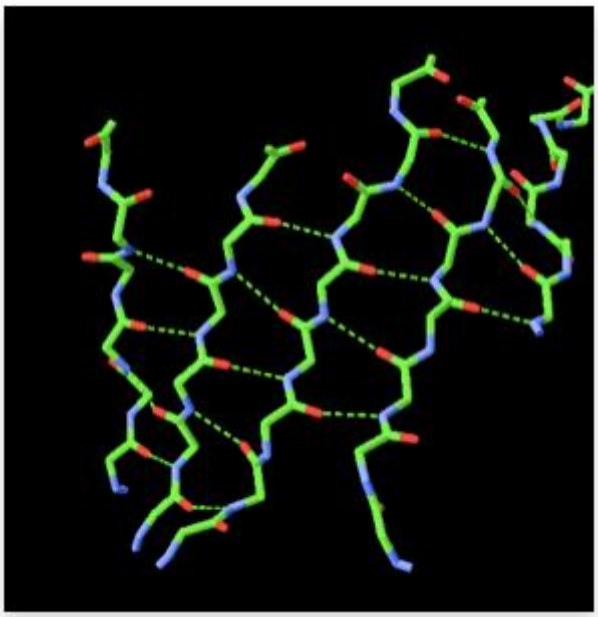

Figure 2: Secondary structure of protein showing Beta-sheet structure.[15]

Figure 3: Secondary structure showing Hairpin loop or tight turn.[15]

The turns are classified into δ-turn, γ-turn, β-turn, α-turn, and π-turn based on the hydrogen bond formed between the carbonyl group (C=O) in the i^{th} position and the amide group (N-H) in the i+1, i+2, i+3, i+4 and $i+5^{th}$ position of the amino acid sequence.[16] The β-turn is the most common turn observed and studied extensively due

to their effect in the function of different proteins and smaller peptides. They have nine sub-classes based on the dihedral angles (Ψ and Φ) of the residues in the i+1 and i+2 position of the peptide sequence.[15] The γ-turn is the second most observed[17] and studied turn after the β-turn and forms hydrogen bond between the carbonyl group of the ith residue and amide group of the i+2th residue. They have two sub-classes i) the classic and ii) the inverse, based on the dihedral angles. The inverse type is the commonly observed type of γ-turn. The classic type is responsible for the reversal of the direction of the peptide sequence.[18] The other types of secondary structure are not studied extensively but also have an effect on the structure and the functioning of the peptide. During the project, I worked on smaller peptides like di- and tri-peptide to find their structure and understand the different turns they form. The structure of the small peptides and their functional properties can be found in the results and discussions section.

1.3. Studying the secondary structure

The protein structure is governed by the orientation of all the amino acid residues and their side chains. The non-covalent interactions between the residues change the orientations, giving rise to different secondary structures (like the α-helix, β-sheets, and turns). These structures lead to the exposure of specific amino acid residues on the surface of the protein structure, enabling the protein to carry out a specific function. By studying the secondary structure of the small peptides in the protein, we get closer to understanding the complete structure of the protein. If the protein is observed to be involved in a specific function, then the small peptides in the protein and their orientation are essential for the protein to carry out that function. In such cases, by understanding the secondary structure of the small peptides, we can say that the structure plays a role in enabling the protein to carry out the function.

The secondary structure of the peptides is studied by different techniques like Nuclear Magnetic Resonance (NMR)[19], X-ray crystallographic technique[20], Electron Paramagnetic Resonance (EPR) spectroscopy[21]. Studying non-covalent interactions in

the secondary structure of the peptide in the absence of any external perturbations like the solvent molecules and intermolecular interactions can be done in the gas-phase spectroscopy.

We have studied the secondary structure of Boc-Gly-L-Pro-NH-Bn-OMe (a dipeptide) and Boc-Phe-SAA-Val-OMe (a tripeptide) using the gas-phase spectroscopic studies along with quantum chemical calculations. The dipeptide is found to be a part of an enzyme that is involved in treating Alzheimer's disease.[22] The function of the enzyme is studied in detail, but the structure of the enzyme is not fully understood. Understanding the structure of the dipeptide will be a step towards understanding the enzyme's structure. It is also reported that the intra-residue C5 hydrogen bonding shown by the dipeptide has a significant contribution towards stabilizing the peptide.[23] One of the low energy conformers obtained from the calculation shows intra-residue C5 hydrogen bond having bond distance comparable with the report. The tripeptide is found to show self-assembling property and capable of transporting ions across membrane barriers.[24] The intra-residue C5 hydrogen bond is observed from the calculations in case of the extended conformations of the tripeptide. The structure of the sugar-amino acid residue containing tripeptide is studied in solution-phase.[25] However, very less work is done in studying such structures in gas-phase, so gas-phase study on the tripeptide is attempted. A window of different conformations is possible, due to the conformational flexibility of the dihedral angle between the peptide bonds. All the possible relatively low energy conformers were screened down through quantum chemistry calculations and are expected to be observed in the experiment. Spectroscopic techniques like laser desorption, Resonantly Enhanced Multi-photon Ionization, and Resonant Ion Dip Infrared Spectroscopy are used for the study.[26,27]

2. METHODS

2.1. Experimental Setup

The gas-phase spectroscopy experiment was carried out in the home-built Time of flight-Mass spectrometer-Resonantly enhanced multi-photon Ionization setup (TOF-MS-REMPI) coupled with supersonic jet expansion technique.[28] The experimental setup has two vacuum chambers i) molecular beam chamber and ii) ionization chamber. These chambers are connected by a skimmer having a 2 mm diameter. The two chambers are pumped by separate diffusion pumps to generate a vacuum of range 10^{-8} Torr. The desorption setup is placed behind the skimmer in the molecular beam chamber just in front of the pulse valve. The ionization chamber is also connected to the 1 m long TOF chamber (Jordan TOF, USA), which has a Micro-channel plate (MCP) detector on the top. The schematic of the setup is shown in figure 5. The mixture of Gly-L-Pro and graphite powder (Sigma Aldrich, size ~20 µm) in the ratio 8:4 was made into a pellet of 2 mm thickness and 8 mm diameter using the KBr Press by applying ~3 tons of pressure.

2.1.1 Laser Desorption Technique

The pellet was attached to a 9 mm diameter graphite rod and attached to a motor that is fixed to the desorption sample holder. The motor rotates during the laser desorption bringing new surface for desorption. The peptide from the solid phase is brought into the gas phase by the Laser Desorption technique[29], where a 532 nm laser (150 µJ/pulse) from an Nd: YAG laser (continuum, Minilite-1, 10 Hz, 10 ns) is focused on a small part of the peptide with the help of lenses as shown in figure 4. In this technique, the peptide gets vaporized due to the rise in temperature by ~1000 K in ~10 ns on the sample surface, which is fast enough to bring the sample to the vapor phase without disrupting any peptide bonds. This technique succeeds in crossing the barrier of limitation in bringing peptide samples into the vapor phase, which is not possible by regular heating.

The peptide molecule has low vapor pressure and can decompose upon heating at high temperatures.

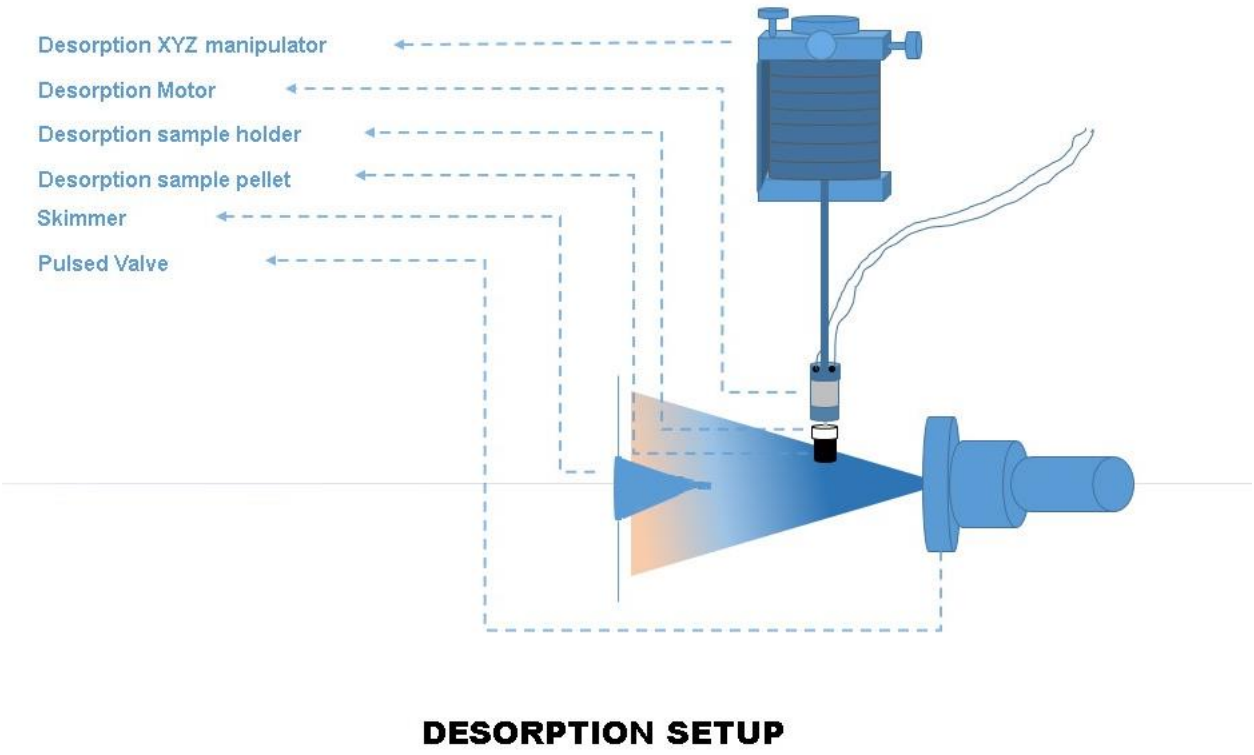

Figure 4: *The schematic explaining the desorption mechanism. The desorption laser comes perpendicular to the molecular beam and falls vertically on the surface of the desorption sample.*

2.1.2 Supersonic Jet Expansion

Argon (Ar) is used as the carrier gas, which undergoes supersonic jet expansion at the pulsed nozzle, cooling the vaporized sample and carrying the sample into the ionization chamber through the skimmer. The Argon gas is released from the cylinder at 4.1 x 10^3 Torr pressure, which comes out of the pulsed valve into the molecular beam chamber, which is at 10^{-8} Torr pressure. The vast difference in pressure results in the collision, followed by the expansion of the gas in the molecular beam chamber. This collision converts the internal energy of the molecules into a directed mass flow,[30] which causes

internal cooling. There is an increase in the Mach number of the gas (M = u/a, where M is the Mach number, the mass flow velocity is given by ‘u,’ and ‘a’ denotes the classical speed of sound). The gas flow is termed supersonic jet as M > 1, the speed of the gas is faster than the speed of the sound. The population of Ar atoms in higher translational, vibrational, and rotational energy levels is cooled down the ground vibrational and rotational energy level. Such concentration of the population in the ground state of the vibrational and rotational energy level is known as the internal cooling. This internally cooled Ar gas serves as a bath for the peptide sample resulting in cooling the peptide internally, concentrating the population of the peptide molecule in the ground translational, vibrational, and rotational energy level.[31] The molecular beam consists of the Argon gas and the sample vapor. The coldest region of the molecular beam is selected using the skimmer.

2.1.3 Time of flight (TOF) chamber

The molecular beam, reaches the ionization chamber, after passing the skimmer, where the beam is ionized upon interaction with the UV beam. The ionized molecular beam is repelled towards the detector by the voltage from the plate below the molecular beam. The repelled beam is extracted and accelerated towards the detector by the other two plates in the ionization chamber (above the molecular beam). These plates make sure that all the ions leave from the plates at the same time. The ions accelerated towards the detector travels through the field-free zone of the 1m long time of flight (TOF) chamber and reaches the detector after a specific time. The time the ion takes to reach the detector depends on only the mass of the ion. The detector records the intensity of ions reaching the detector at a different time. The intensity is amplified and recorded using a pre-amplifier and the oscilloscope, which is connected to a computer. Home-built LabView program (National Instruments), is used to collect the data and also to control the laser programs. The time to mass-scale conversion is done based on the time indole ion takes to reach the detector. The molecular mass of indole is 117.14 g/mol, and the time it takes to reach the detector is 15.66 µs.

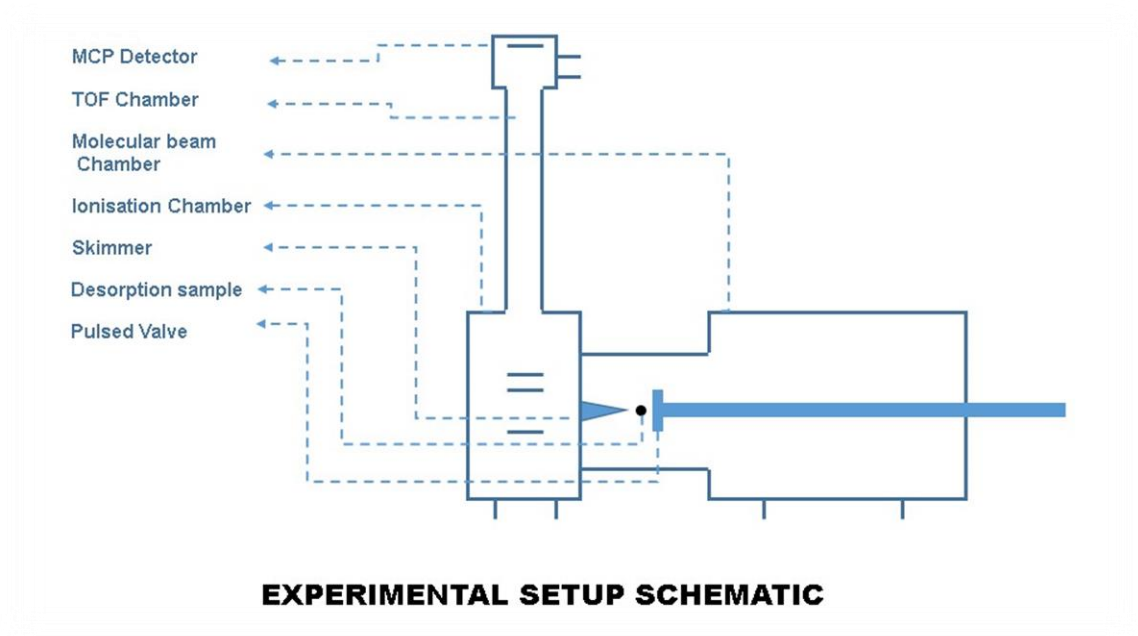

Figure 5: *Schematics of the experimental setup showing the Molecular beam, the ionization beam, and the TOF chamber.*

2.2. Spectroscopic Techniques

2.2.1. Resonant Two-Photon Ionization (R2PI)

Two photons are utilized in the spectroscopic technique Resonant Two-Photon Ionization (R2PI). The molecule from the ground state (S_0) is excited by one photon to the excited state (S_1). Subsequently, the molecule from the excited state is ionized to the ionized state continuum (D_0), by the second photon has the same wavelength as the first photon.[32] R2PI spectrum is recorded in the UV region. The UV laser of desirable wavelength (200 – 300 µJ/pulse) is produced using a tunable dye laser (ND6000, Continuum), which is pumped by the doubled output (532 nm) of Nd: YAG laser (10 ns, 10 Hz, Surelite II-10, Continuum).

2.2.2. Resonant ion-dip infrared Spectroscopy (RIDIRS)

In the case of the Resonant Ion-Dip Infra-Red Spectroscopy (RIDIRS) technique, a tunable IR laser is used to excite the molecule from the ground vibrational state (v_0) to the excited vibrational level (v_1) of the ground electronic level (S_0). The IR laser (Laser Vision, pulse energy ~4-5 mJ/pulse, resolution ~2.5 cm^{-1}), was pumped by an unseeded Nd: YAG laser (Continuum, Surelite II-10, 10 ns, 10 Hz). After 160 ns, a single photon from the UV laser of fixed energy excites the population in the ground vibrational level (v_0) of the ground electronic level (S_0) to the excited electronic level (S_1), which is followed by ionization due to another UV photon. The UV energy is fixed at one of the peaks obtained from the R2PI spectrum, which corresponds to the electronic transition from the ground state to a higher electronic level of a certain conformer. The energy of the IR laser beam is in the range of the N-H stretching frequency region (3100 to 3500 cm^{-1}). Since the ground state population is already excited by the IR laser, there will be less excitation by the UV laser resulting in less number of ions produced when the IR laser energy matches any vibrational transition of the molecule, which results in a dip in the ion signal recorded by the detector.[32] By this technique, we get the IR spectrum of a specific conformer in the NH stretching region.

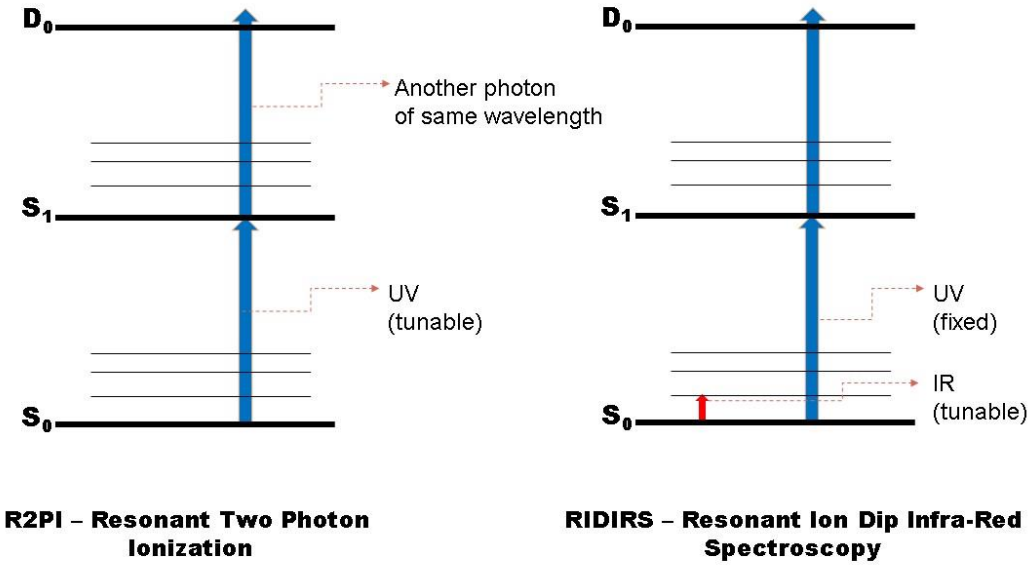

Figure 6: *Schematic of R2PI and RIDIRS, the mechanism is explained in the experimental section. The violet arrow indicates the UV laser, and the red arrow indicates the IR laser.*

2.3. Computational Methods

Different possible conformers were obtained through computation, which is compared with the experimental data to find the preferred conformation, and the interactions that stabilize the conformer is observed. Few conformers having energy within 20 kcal/mol from the lowest energy conformer is selected, and the geometry optimization is carried out at the HF/6-31G level of theory using the Gaussian 09 software[33]. The frequency of the conformers having relative energy less than 20 kJ/mol is calculated after geometry optimization using the M05-2X/6-31+G(d) level of theory using the Gaussian 09 software. Gibbs free energy was calculated for a range of temperatures to correct the electronic energy of the different conformers. The Natural Bond Orbital (NBO) analysis for the frequency calculated conformers was performed using the NBO6.0 program package[34] at the M05-2X/6-31+G(d) level of theory using the Gaussian 09 software. In NBO, localized natural bond orbitals are obtained from the delocalized molecular orbitals and are visualized in the Lewis structure of the molecule.

From NBO analysis, the interaction energy between the donor and the acceptor orbitals can be obtained from the second-order perturbation energy. All the interactions that stabilize the conformer can be seen observed from the NBO calculation.

3. Result and Discussion

3.1. Dipeptide

The dipeptide I worked on is the Boc-Gly-L-Pro-NH-Bn-OMe (BGLPOMe), which is made up of the amino acids Glycine (Gly) and L-Proline (L-Pro). The N-terminal of the dipeptide is protected using the tert-Butyloxycarbonyl (BOC) group, and the C-terminal is protected using NHBnOMe (Bn = Benzyl), where the benzyl group acts as a chromophore for the study. There are three complete amide bonds (C-N) in the dipeptide. The protected group is similar to -OMe substituted phenylalanine (Phe) amino acid residue, hence the end-protected dipeptide can be treated as a tripeptide model and the folding of the benzyl group could be studied. Glycine is the most straightforward amino acid sequence with no side chain at the C_α position. The maximum number of different conformers that the glycine can form on its own is less due to its size and non-availability of any side groups. The proline group is a rigid amino acid sequence and can also form a very less number of conformers[35] due to the ring constrains but gives rise to puckering effect.[36,37] The addition of the protecting groups increases the possible conformers by many folds.

Gly-L-Pro is found in nature as a cyclic dipeptide, inducing disease resistance in the Arabidopsis plant.[36] The cyclic dipeptide is also found to have acaricidal properties.[38] Weak intra-residue hydrogen bond being the major contributor towards stabilizing Z-Gly-Pro-OH dipeptide is reported.[23] The tripeptide Gly-L-Pro-Glu (GPE) is used in the treatment of Alzheimer's disease because of its anti-inflammatory effects.[39] Knowing the conformational preferences will help us understand the functions of the peptide due to the structure-function relationship.

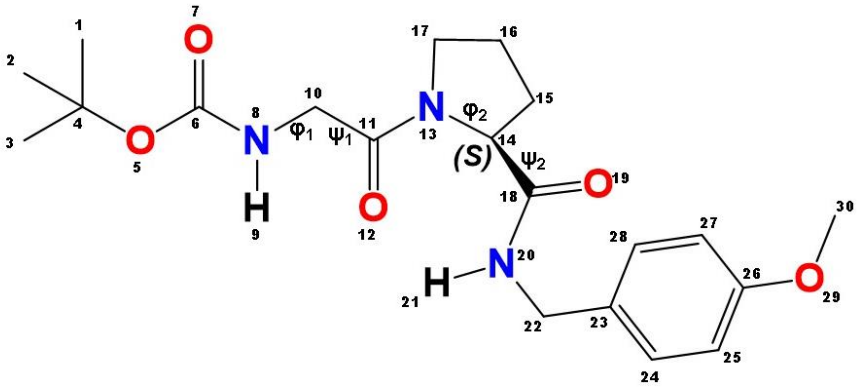

Figure 7: Structure of the dipeptide: Boc-Gly-L-Pro-NH-Bn-OMe and the numbering. The dihedral angles Φ_1 (C6-N8-C10-C11), Ψ_1 (N8-C10-C11-N13), Φ_2 (C11-N13-C14-C18) and Ψ_2 (N13-C14-C18-N20) are the Ramachandran angles.[40]

3.1.1. Conformational analysis of BGLPOMe

CONFLEX conformational analysis software gave all possible conformations (280 conformers) of BGLPOMe having potential energy within 25 kcal/mol of the lowest energy conformer, using the MMFF94s[41] force field. The conformers having similar structures and having an energy difference less than 0.1 kcal/mol were grouped, and the geometry of the conformers was optimized at the HF/6-31G level of theory using Gaussian 09 software. The geometry optimized conformers were grouped into 40 conformers, which were then optimized at M05-2X/6-31+G(d)[42] level of theory. The frequency was also calculated at the same level of theory. The calculation at a higher level of theory was not attempted as the computation at the level of the theory mentioned above took more time to complete (3 – 4 days). Eleven conformers having potential energy less than 20 kJ/mol from the lowest conformer, were obtained. Grouping the conformers based on the relative energy and the structural similarity gave seven conformers, which were grouped into three different types namely, Type A, Type B, and Type C, based on the type on interactions that are found to be similar in the groups. All the conformers obtained showed folded conformation[23], thus given a nomenclature as Fn/Ci/Cj where n tells the position of the conformer based on the

relative energy of the conformer at 0 K and i,j tells the number of atoms involved in the closed cycle formed due to the hydrogen bond. For example, the lowest energy conformer is named as F1/C5/C7, where F1 says that the conformer is folded and the lowest in energy at 0 K and C5, C7 says that due to hydrogen bonding, C5 and C7 cycle is observed in the F1 conformer. The relative energy of different conformers and the Ramachandran angles are shown in table 1.

Table 1: *The relative energies of the different conformers and the geometrical parameters observed at the M05-2x/6-31+G(d) level of theory.*

Conformers	ΔE_{rel} (0 K) (kJ/mol)	ΔG_{rel} (300 K) (kJ/mol)	Φ1 (C6-N8-C10-C11)	Ψ1 (N8-C10-C11-N13)	Φ2 (C11-N13-C14-C18)	Ψ2 (N13-C14-C18-N20)
F1/ C5/ C7	0.00	0.00	167.61	-176.44	-82.51	78.22
F2/ C5/ C7	7.21	8.62	170.94	-175.32	-80.15	83.17
E1/ C5/ C7	12.81	0.42	-179.41	178.94	-85.70	66.21
F3/ C10	7.49	7.00	59.58	-130.91	-78.46	-4.51
F6/ C10	13.75	13.54	61.64	-140.95	-65.98	-14.38
F4/ C7/ NH-π	9.02	9.23	123.37	-61.14	-85.77	76.35
F5/ C7/ NH-π	9.16	7.43	120.81	-53.62	-80.01	84.14

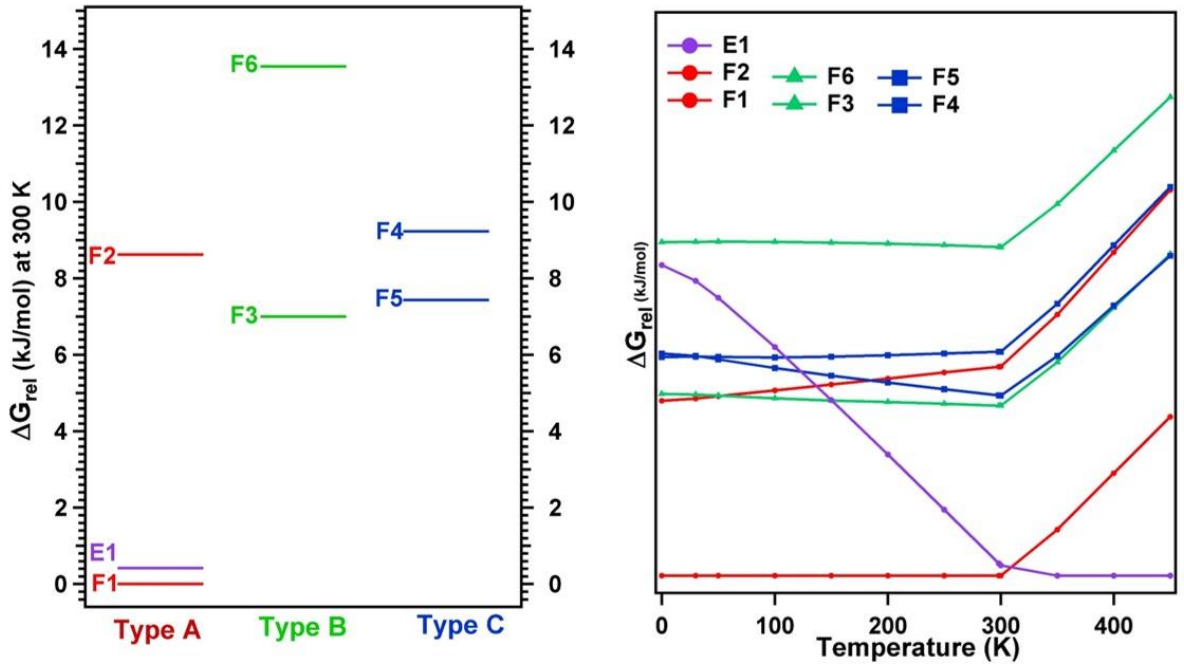

Figure 8*: The left figure shows the relative Gibbs free energy of the different conformers at 300K. The Type-A conformers (represented by red color for the folded and purple color for the extended) appear to be the preferred orientation due to their low Gibbs free energy. The figure on the right shows the change in the relative Gibbs free energy with an increase in temperature. The energy of the extended conformer (represented by purple color) decreases upon the increase in the temperature, showing that the peptide prefers extended conformation at the higher temperature.*

From the thermal analysis shown in Figure 8, it is observed that the type A conformer F1 is the lowest energy conformer till 300 K. The conformer E1 becomes lowest above 300 K, which may be because the rise in temperature affects the hydrogen bonds of other conformers, which reduces their stability at a higher temperature.[43] The stability of the extended conformation is enhanced by the entropic effects at higher temperatures. Such stabilization of the extended conformer over the folded conformer of different peptides has been reported in the literature.[44,45,46,23] Type A shows the C5 and C7 cycle formed due to the bifurcated hydrogen bond between the carbonyl group (C=O) of the Gly residue and the nearby amine group (N-H) groups. The C5 cycle is due to the intra-residue hydrogen bond formed by the glycine (Gly) residue. The C7 cycle is formed by the γ-turn between the carbonyl group (C=O) of the Gly residue, and the amine (N-H) group found next to the proline residue. The hydrogen bond shows a bifurcated

interaction sharing the same lone pair of the oxygen atom with the two amine groups. The lowest energy conformer also shows type A kind of structure. There is also a hydrogen bond between the carbonyl group (C=O) of the Boc group and the hydrogen from the C-H bond of the OMe group, which makes the structure into a closed structure. Six out of eleven conformers belong to the A-type, but only four of them have the C=O…H-C hydrogen bond. One of the structures of the group shows extended conformation despite having the bifurcated hydrogen bonding interaction. The orientation of the protecting group (benzyl group) determines if the structure is closed or extended, giving rise to different conformers. As the benzyl group act as the chromophore, the change in the orientation will be reflected in the electronic absorption spectrum.

The B-type shows C10 cycle conformation forming hydrogen bond between the carbonyl (C=O) of the Boc group and amine (N-H) group at the C-terminal of the proline residue resulting in a β-turn, the commonly observed turns in peptides. The orientation of the glycine residue plays a role in forming the C10 cycle. The NH group next to the benzyl is involved in the hydrogen bonding, which will induce an effect in the benzyl group, thus altering the S_0 to S_1 transition. It is expected to give rise to different conformers in the spectrum.

The C-type shows the C7 cycle, similar to the one observed in the type-A. Along with the C=O…H-N hydrogen bond, N-H…C=C (NH-π) hydrogen bond is observed between the NH group of the Gly residue and the benzyl group. Energetically lowest conformer of the group has a relative energy of 9 kJ/mol more than the lowest conformer of Type-A. Among all the groups, type-C has direct interaction with the benzene ring, so it is expected to find a conformer of this group in the electronic spectrum. The lowest energy conformers from different groups are shown in figure 9.

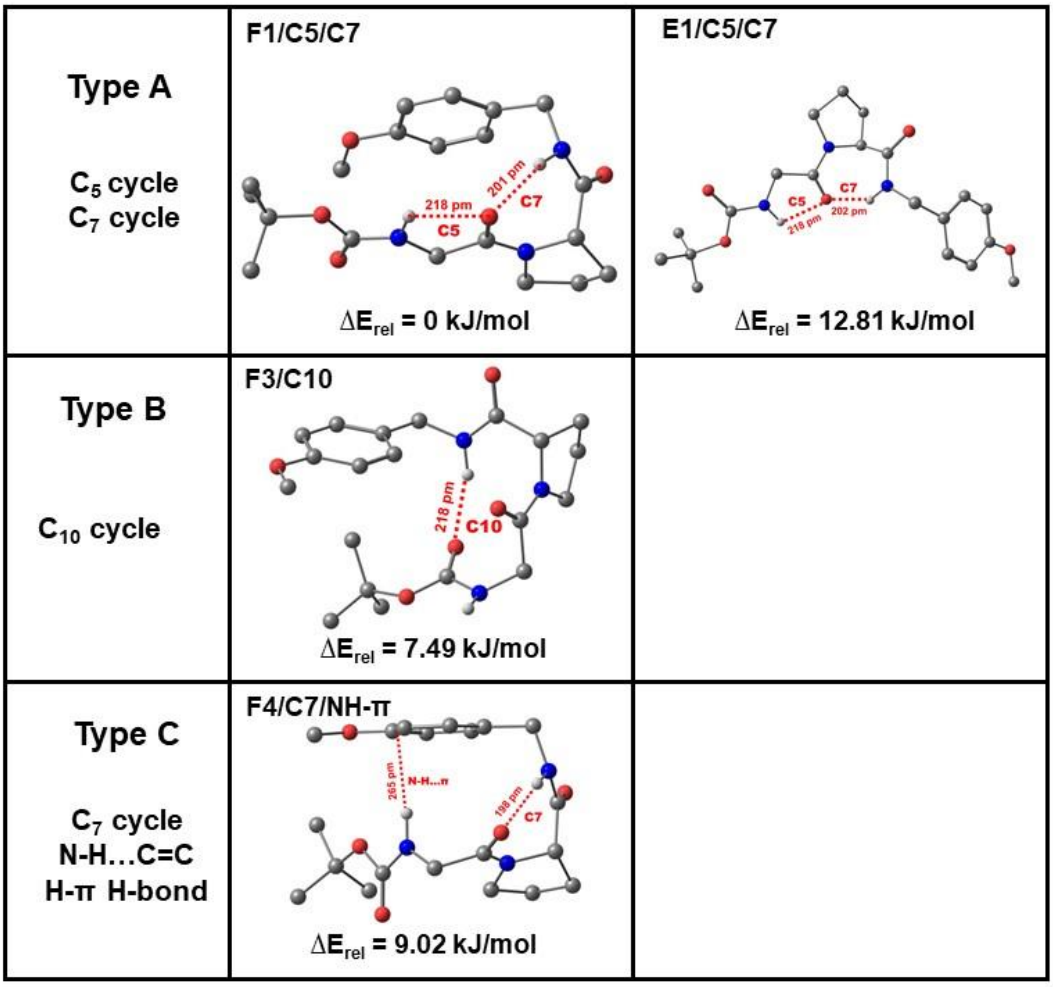

Figure 9*: Conformers representing different types and their relative energy at 0 K.*

The solution phase IR spectrum of the dipeptide was recorded in $CDCl_3$ solvent. The IR spectrum in the NH stretching frequency region (3200 - 3550 cm^{-1}) is shown in figure 10. The two distinct NH bands can be seen from the spectrum; the broad peak at 3338 cm^{-1} corresponds to the NH of the C-terminal protecting group, and the peak at 3434 cm^{-1} corresponds to the NH of the glycine residue. The glycine NH is in resonance with the carbonyl group with an electronegative oxygen group on the other side. The electronegativity of the oxygen results in an inductive effect towards the oxygen atom, strengthening the N-H bond and giving rise to a higher stretching frequency. The glycine NH is connected to an ester group (-COO), whereas the NH of the C-terminal protecting group is not, which results in the red-shift. The difference between the two bands (96

cm^{-1}) is closer to the difference of the unscaled frequencies of the type-B conformer F3 (98 cm^{-1}). It can be speculated that the conformer F3 is the most stable conformer in the solution phase. The stabilization of the type-B conformer is probably due to the intermolecular interaction and solvent effect.

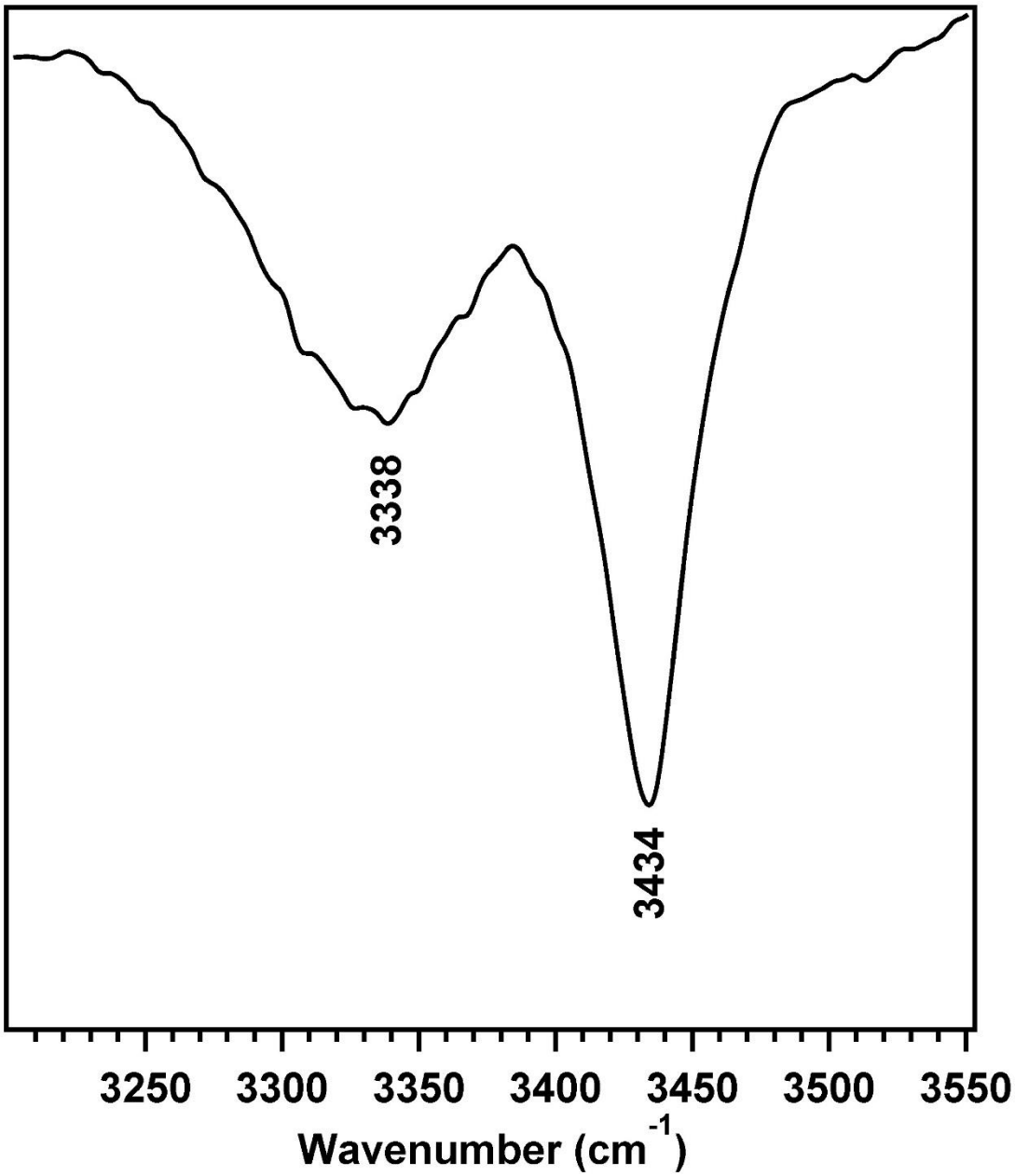

Figure 10: *Solution phase IR spectrum of the dipeptide in the NH stretching frequency region (3200 - 3500 cm-1).*

The scaled frequency obtained from the frequency calculation of different conformers is listed in table 2. The frequencies are scaled by 0.944, which was used as a scaling factor for a different Gly-Pro dipeptide.[23] From table 2, it is observed that the frequency for the groups is different and thus can be used to find the stable conformation along with the spectroscopic IR data obtained from the RIDIRS.

Table 2: *The scaled stretching frequency of the conformers of different groups.*

Conformers	$\nu_{N8\text{-}H9}$ (cm^{-1})	$\nu_{N20\text{-}H21}$ (cm^{-1})
F1/ C5/ C7	3456	3345
F2/ C5/ C7	3462	3376
E1/ C5/ C7	3466	3389
F3/ C10	3500	3407
F6/ C10	3490	3414
F4/ C7/ NH-π	3427	3356
F5/ C7/ NH-π	3410	3371

3.1.2. NBO Analysis

The Natural Bond Orbital Analysis of the different groups is performed to measure the interaction energy of all the interactions that are involved in stabilizing the structure. From the analysis, we can see all the non-covalent interactions that are found in the different groups from the figure 10, and their second-order perturbation energy is measured and is shown in the table. From the NBO analysis, the interaction energy of the hydrogen-bond forming the C5 cycle is observed to be in the range of ~15 – 18 kJ/mol and C7 cycle are observed in the range of ~10 – 20 kJ/mol and C10 cycle is observed in the range of ~14 – 15 kJ/mol. The NH-π hydrogen bond is equally strong as a C7 cycle having energy in the range of ~13 – 16 kJ/mol.

Both type-A (Fn/C5/C7) and type-C (Fn/C7/ NH-π) have a common hydrogen bond that shows the C7 cycle due to the γ-turn. However, it is observed that type-A conformers showing bifurcated H-bonding interaction, has less orbital overlap interaction energy

($E_{n\rightarrow\sigma*}^{(2)}$) in the C7 H-bond. This decrease in interaction energy is due to the bifurcated hydrogen bond in type-A, where the lone pair from the oxygen atom is shared with both σ* of the N-H bond of the glycine residue forming C5 cycle and the N-H bond of C-terminal protecting group forming C7 cycle. In the case of type-C, the oxygen lone pair is donated only to σ* of the N-H bond of the C-terminal protecting forming C7 cycle.

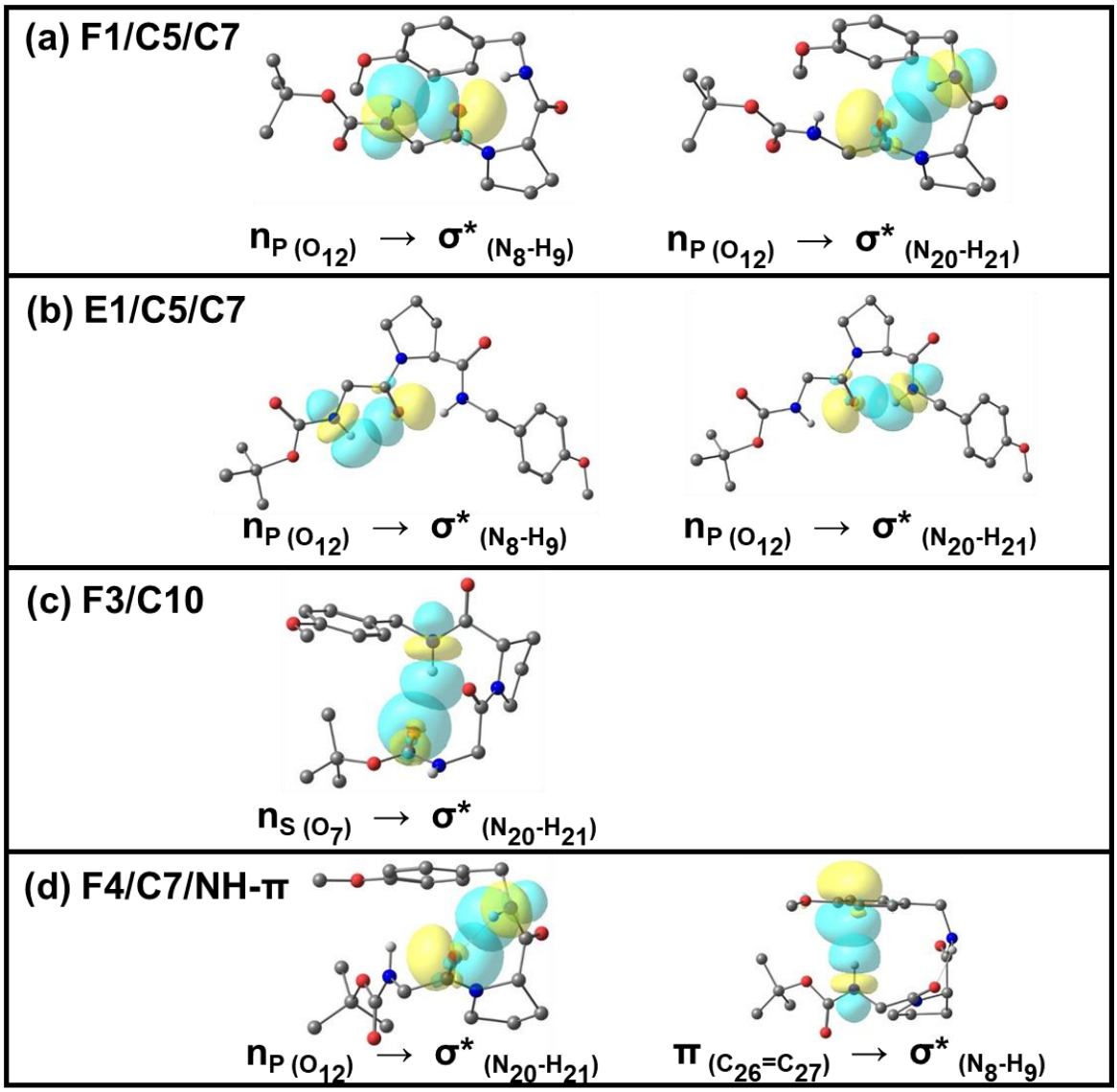

Figure 11*: NBO analysis of conformers of the different groups showing all the overlapping orbitals involved in forming hydrogen bonds.*

***Table 3**: The table shows the NBO analysis of the conformers calculated at the M05-2X/6-31+G(d) level of theory and the second-order perturbation energy of all the non-covalent interactions observed in different conformers.*

Conformers	Second-order perturbation Energy $E^2_{i\rightarrow j*}$ (kJ/mol)			
	C5 cycle	C7 cycle	C10 cycle	H-π H-bond
F1/ C5/ C7	18.12	9.04	-	-
F2/ C5/ C7	15.82	9.08	-	-
E1/ C5/ C7	18.20	8.70	-	-
F3/ C10	-	-	15.10	-
F6/ C10	-	-	14.23	-
F4/ C7/ NH-π	-	21.08	-	21.08
F5/ C7/ NH-π	-	17.24	-	13.39

3.1.3. Mass spectrum of BGLPOMe

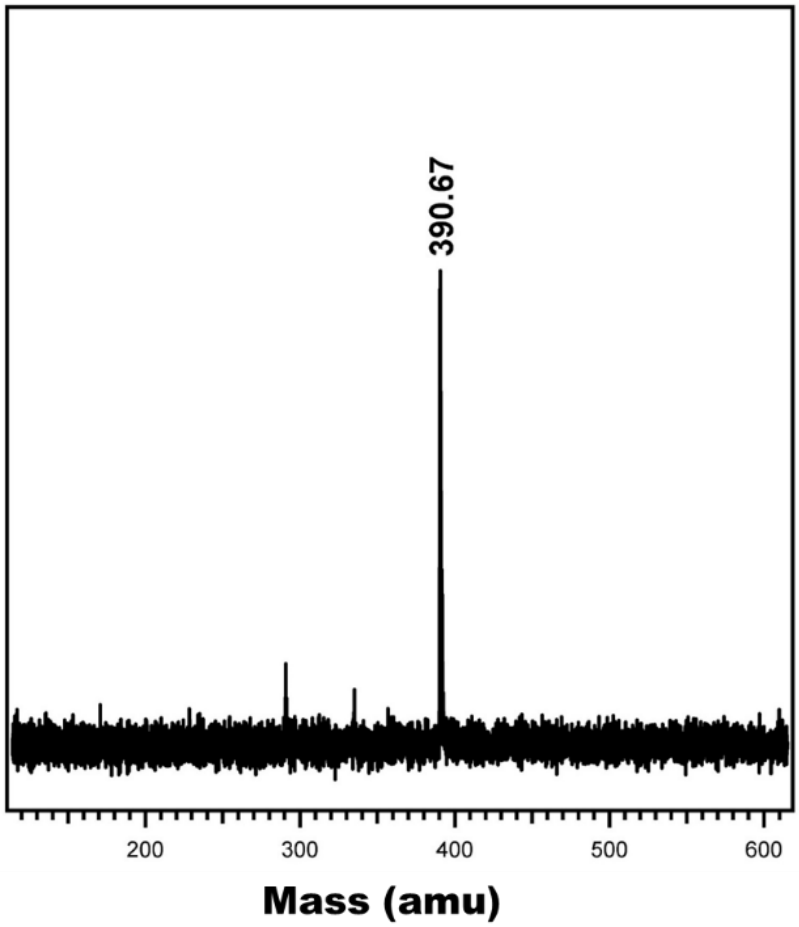

***Figure 12:** The mass scan of the dipeptide observed at the 28 µs mass channel in the oscilloscope.*

The signal of BGLPOMe was observed in the oscilloscope at 28 µs mass channel, which corresponds to the molecular mass of 391 amu.

3.1.4. Electronic Spectrum of BGLPOMe

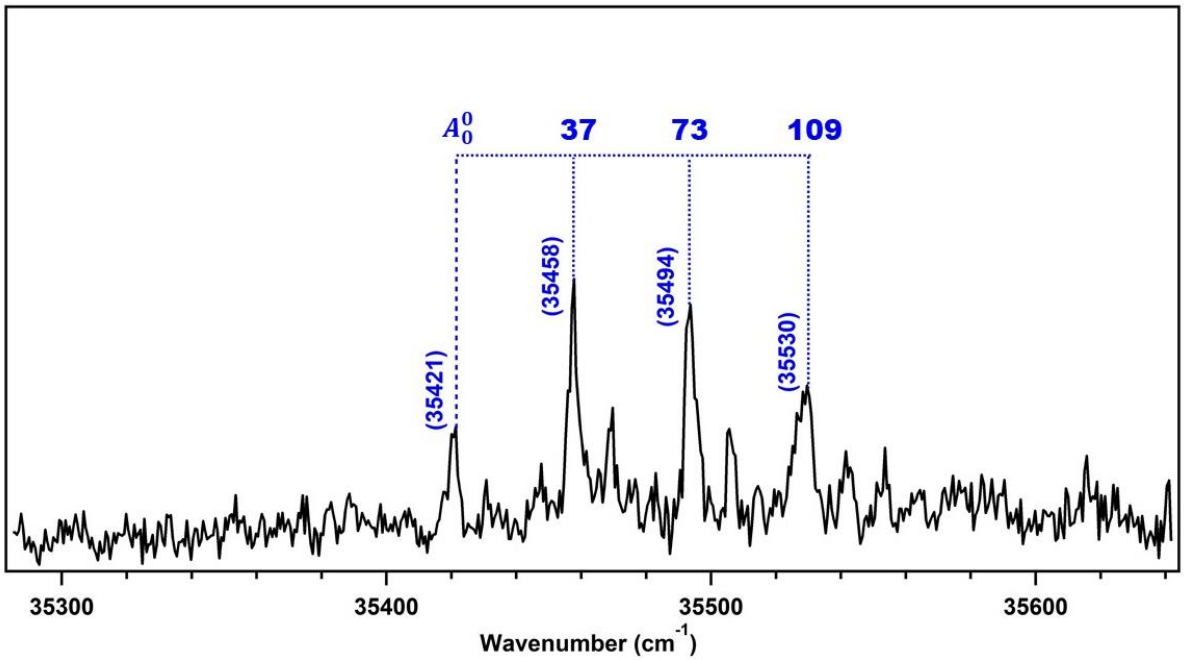

Figure 13:*1C-R2PI spectrum of the dipeptide showing A_0^0 transition at 35421 cm^{-1}.*

The R2PI of the BGLPOMe recorded for the range 35200 – 35700 cm^{-1} in the mass channel of the dipeptide (391.25 amu) is shown in the figure 13. The S_0-S_1 transition is observed at 35421 cm^{-1} and A_0^0 + 37, A_0^0 + 73, A_0^0 + 109 are considered to be the bands of the same conformer, tentatively. The other small bands may correspond to different conformers or belong to a single conformer, which can be concluded from the double resonance spectroscopic techniques like the hole-burning spectroscopy. From UV-UV and IR-UV hole-burning techniques, we can get the conformer specific UV and IR spectra of the BGLPOMe dipeptide.

3.2. Tripeptide

I have also worked on a tripeptide, Boc-Phe-SAA-Val-OMe, which has two natural amino acids, namely phenylalanine (Phe) and Valine (Val) and an unnatural amino acid called the Sugar amino acid (SAA). This sugar amino acid was made by adding amino acid functional groups to a sugar unit. The sugar amino acid is a newly synthesized unnatural amino acid unit; hence its properties and natural availability are not thoroughly studied. This amino acid has the rigidity of a sugar unit and versatile in function as an amino acid unit.[47] It is also a γ-amino acid which has a turn inducing property. The N-terminal is protected by the tert-Butyloxycarbonyl (BOC) group, and the OMe group protects the C-terminal. There are four amide bonds (C-N) in the tripeptide. Unlike other γ-amino acids, the sugar amino acid is rigid and induces a specific turn on the tripeptide, which is observed in all the conformers. The structure of the tripeptide in solution-phase and its self-assembling property is already studied.[24,25] Since this tripeptide has a Phe residue, which acts as the chromophore, this tripeptide can be studied in gas-phase. Very less work is done in studying sugar-amino-acid containing tripeptides in the gas phase,[48,49] as producing sugar-amino acid involves more synthesis, and to study the R2PI spectrum, the residue should be attached to a chromophore, and the sample should be stable for an extended period. There are six stereocenters which increase the possibility of having different conformers by many folds.

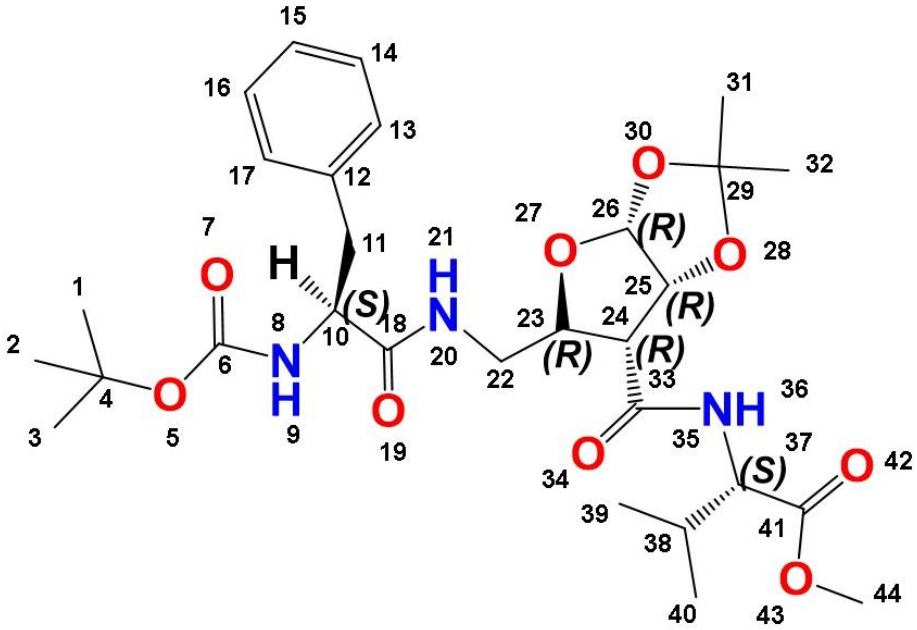

Figure 14: *Structure of the tripeptide Boc-Phe-SAA-Val-OMe*

3.2.1. Conformational analysis of the tripeptide

The structure of the Sugar Amino Acid (SAA) residue in the Gaussian software was made, with random bond length and bond angle, but only maintaining the valency of all the atoms in the residue. The structure was optimized at the HF/3-21G level of theory. Once the SAA residue is optimized, the tripeptide is constructed around this optimized geometry. The tripeptide is then optimized at the HF/3-21G level of theory. From the optimized tripeptide structure, 100 possible conformers were obtained using the MMFF94 force field available in the MarvinSketch software. The conformers were grouped, and 50 conformers having energy less than 10 kcal/mol were optimized at the HF/6-31G level of theory, which was grouped, and 28 conformers were obtained. Out of the 28 conformers, 16 conformers were obtained after optimization at the M05-2X/6-31+G(d) level of theory for which the frequency was then calculated. These conformers were grouped into three categories Hairpin shaped, S-Shaped, and extended, based on the shape of the backbone of the tripeptide. The Hairpin shaped conformer was found to

have the lowest energy of all the predicted conformers. All three categories, though, differ in shape, expressed turns at the SAA residue, implying that the turn-inducing effect of the residue is superior to the other interactions present in the different conformers.

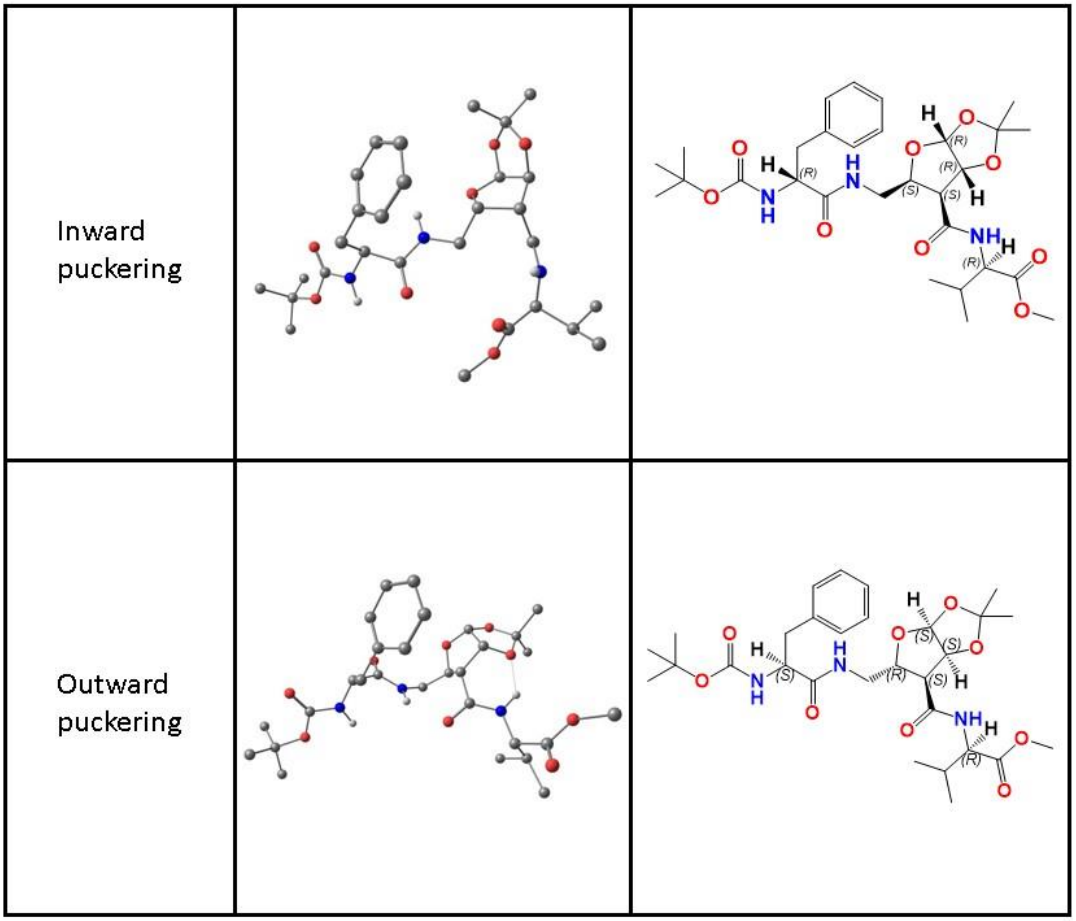

Figure 15: *The molecular structure and the 2D structure of conformers showing a difference in the puckering of the SAA residue. The turn inducing property can be seen in both cases.*

When the structure of the most stable conformer obtained computationally was compared with the structure reported in the solution phase, it is observed that the SAA residue was showing a different kind of bending (puckering). Another structure was made showing similar puckering effect as the structure in the solution phase and optimized at M05-2X/6-31+G(d) level of theory. It was observed that the newly found

conformer was far lower in energy, indicating that it is more stable than all the previously calculated conformers. Fifty new conformers were obtained using the MMFF94 force field in the MarvinSketch software, which was optimized in the HF/6-31G level of theory and gave 29 conformers after grouping. The 29 conformers at M05-2X/6-31+G(d) level of theory were optimized, and the frequency of the geometry optimized conformers was calculated. Seventeen conformers were obtained that were categorized into two groups the curved structure and the extended structure. The differentiation of the groups is based on the folding along the Cα of the Phe group. In the case of the extended conformations, there is a C5 intra-residue hydrogen bond formed in the Phe residue. Only three conformers showed extended conformation.

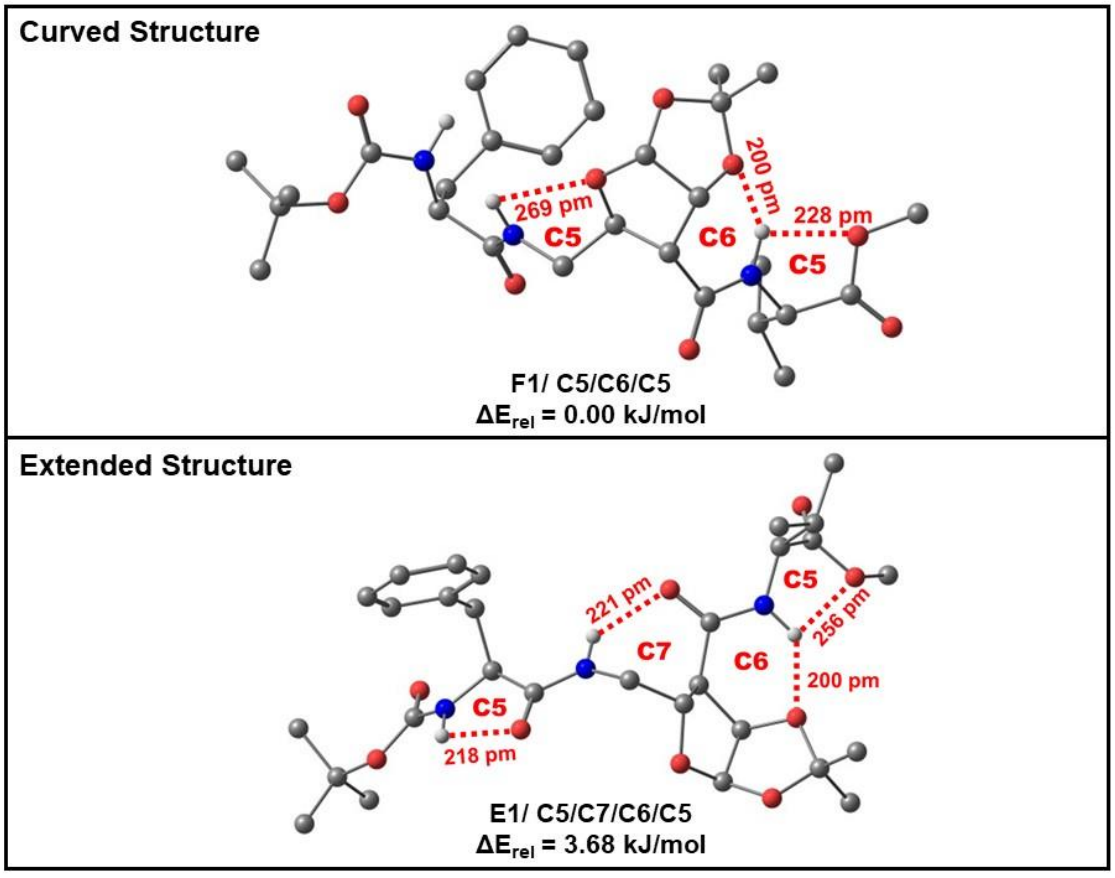

Figure 16: Classification of the conformers of the tripeptide into Curved and Extended conformation

All the conformers, similar to the previous conformers, showed a turn at the SAA residue again, showing that the turn inducing property of the SAA is strong enough to be present in all the conformers. The sugar amino acid has three oxygen atoms, making the residue electronegative and always ready to form a hydrogen bond with the N-H of the amino acid residues. Unlike the turn inducing property due to H-bond in the peptide chain, due to the tetrahedral geometry of the C-atom, the five-membered ring of SAA residue induces turns in the geometry of the peptide. The Phe residue is flexible, which results in forming different kinds of curved structures.

3.2.2. NBO Analysis

From the NBO analysis, it is observed that the hydrogen bonds forming C5 and C7 cycle is present in all conformers. One of the oxygen atoms in the SAA residue also forms a hydrogen bond with the NH group of the valine residue giving rise to the C6 cycle. Due to the number of conformers and the hydrogen bonds present the NBO analysis of all conformers is not shown. The participation of the electronegative oxygen atoms of the sugar amino acid residue in the hydrogen bonding can be seen. The NBO analysis of the lowest energy conformer of both groups is shown in the figure.

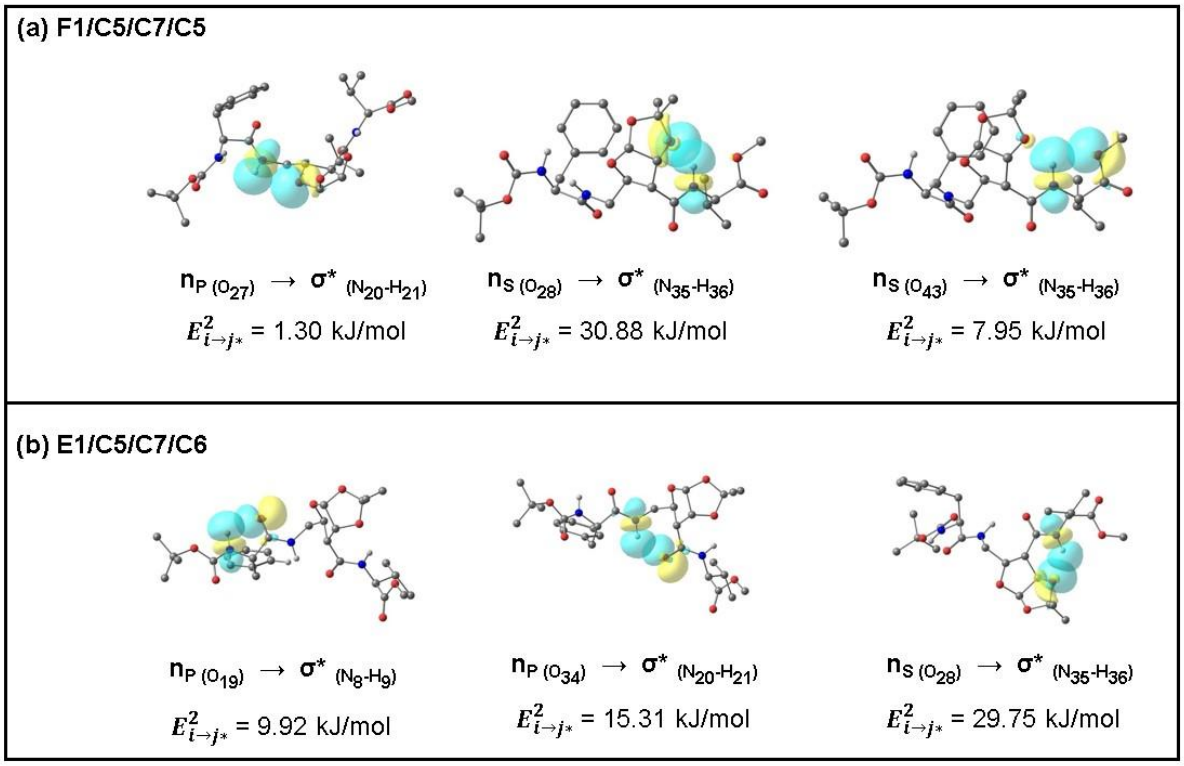

Figure 17: *NBO analysis of the lowest energy conformers of the folded and the extended type and the second-order perturbation energy of the hydrogen-bonds*

Understanding the conformational preferences from the experiment was attempted. Since it is a tripeptide, the sample could not be heated to get into the vapor phase, as heating will disrupt all the non-covalent interactions[50]. When the sample was subjected to laser desorption, the sample readily fragmented getting low-intensity ion signal at the desired mass channel (34 µs, molecular mass of the tripeptide is 577.3 amu). The setup was not able to handle (burn marks were observed on the glass window of the system) when tried to increase the signal intensity by increasing the desorption energy.

4. Conclusion

The R2PI spectrum of the dipeptide Boc-Gly-L-Pro-NH-Bn-OMe was recorded, and the energy corresponding to the transition from S_0 to S_1 for the most stable

conformer is obtained. More sharp peaks in the R2PI spectrum make it possible to have more than one conformer in the ground state. Recording the RIDIRS and IR-UV hole-burning spectrum will confirm the number of conformers present in the ground state. The conformational preference of the dipeptide is also studied through quantum chemistry calculations and is observed that the folded conformation is more stable at temperatures lower than 350K from the Gibbs free energy calculation. The extended structure getting stabilized at higher temperatures shows that the extended structures are entropically favored. The frequency calculation could be carried out in the different levels of theory and different basis sets to see the change in stability of the conformers and also to get more accurate energies and vibrational stretching frequencies.

From the conformational analysis of the tripeptide, the turn inducing property of the SAA residue is predicted to be due to the presence of the electronegative oxygen atoms that tend to form hydrogen bonds resulting in turns in the secondary structure. Such residues can be added in making peptides having turns at desirable positions. However, it is observed that the puckering effect of the residue plays a role in stabilizing the conformation. If the SAA residue preferentially forms one type of puckering, then the residue can be used to make desirable peptides that are rigid, multifunctional, and turn inducing. The gas-phase spectroscopic study of the tri-peptides and other higher peptides are available to be explored in detail.

The conformational study of the di- and tri- peptide shows the peptide prefers to be in folded conformation at lower temperatures and is the global maxima in the case of the Gly-L-Pro and Phe-SAA-Val peptides. At absolute zero (0 K), the population of the molecule in vibrational and rotational ground state will be maximum having deficient vibrational energy and rotational energy, freezing the position of all atoms in the molecule. The electronegative atoms in the molecule try to form non-covalent bonds to stabilize the structure; thus, the whole conformation is folded to enable such interactions stabilizing the overall structure at low temperature. At high temperature (above 300 K), the energy from the external environment increases the vibrational and rotational energy of the molecule, causing an increase in the intensity of vibrational stretching frequency. The electronegative atoms in the molecule try to form bonds to lower the

overall energy by forming inter-molecular bonds, which is enabled when the molecule adapts into an extended conformation.

5. Future Directions

The electronic absorption spectrum of the dipeptide shows multiple bands, which may correspond to a single or different conformer. The double-resonance spectroscopic experiments UV-UV and IR-UV hole-burning experiments have to be done to find all the stable conformers of the dipeptide. The Resonant Ion-Dip Infrared Spectroscopy (RIDIRS) has to be done to get the IR spectrum of the dipeptide in the NH stretching frequency region.

6. References

1. Animal Physiology by Richard W. Hill, Gordon A. Wyse, and Margaret Anderson. *The Quarterly Review of Biology* **2017,** *92* (3), 330-330.
2. Rolland-Cachera, M. F.; Deheeger, M.; Bellisle, F., Nutrient balance and body composition. *Reproduction, nutrition, development* **1997,** *37* (6), 727-34.
3. Kate A Marsh, E. A. M. a. S. K. B., Protein and vegetarian diets. *The Medical Journal of Australia* **2013,** *199* (4).
4. Brosnan, J. T., Rooyackers, Olav Amino acids: the most versatile nutrients. https://journals.lww.com/co-clinicalnutrition/FullText/2013/01000/Amino_acids___the_most_versatile_nutrients.10.aspx.
5. Marriq, C.; Rolland, M.; Lissitzky, S., Structure-function relationship in thyroglobulin: amino acid sequence of two different thyroxine-containing peptides from porcine thyroglobulin. *The EMBO Journal* **1982,** *1* (4), 397-401.
6. Hvidsten, T. R.; Lægreid, A.; Kryshtafovych, A.; Andersson, G.; Fidelis, K.; Komorowski, J., A Comprehensive Analysis of the Structure-Function Relationship in Proteins Based on Local Structure Similarity. *PLOS ONE* **2009,** *4* (7), e6266.
7. Tyndall, J. D. A.; Pfeiffer, B.; Abbenante, G.; Fairlie, D. P., Over one hundred peptide-activated G protein-coupled receptors recognize ligands with turn structure. *Chem Rev* **2005,** *105* (3), 793-826.
8. Botelho., I. R. S., Biochemistry, Secondary Protein Structure. **2019**.
9. Haimov, B.; Srebnik, S., A closer look into the α-helix basin. *Sci Rep* **2016,** *6*, 38341-38341.
10. Zhang, L.; Hermans, J., 310 Helix Versus .alpha.-Helix: A Molecular Dynamics Study of Conformational Preferences of Aib and Alanine. *Journal of the American Chemical Society* **1994,** *116* (26), 11915-11921.
11. Armen, R.; Alonso, D. O. V.; Daggett, V., The role of alpha-, 3(10)-, and pi-helix in helix-->coil transitions. *Protein Sci* **2003,** *12* (6), 1145-1157.

12. Maurya, S. R.; Mahalakshmi, R., N-helix and Cysteines Inter-regulate Human Mitochondrial VDAC-2 Function and Biochemistry. *J Biol Chem* **2015,** *290* (51), 30240-30252.
13. Perczel, A.; Gáspári, Z.; Csizmadia, I. G., Structure and stability of β-pleated sheets*. *Journal of Computational Chemistry* **2005,** *26* (11), 1155-1168.
14. Wu, X.; Brooks, B. R., Beta-hairpin folding mechanism of a nine-residue peptide revealed from molecular dynamics simulations in explicit water. *Biophys J* **2004,** *86* (4), 1946-1958.
15. Al-Karadaghi, S., Protein Secondary Structure: α-Helices and β-Sheets. **2019**.
16. Chou, K. C., Prediction of tight turns and their types in proteins. *Anal Biochem* **2000,** *286* (1), 1-16.
17. Milner-White, E. J., Situations of gamma-turns in proteins. Their relation to alpha-helices, beta-sheets and ligand binding sites. *Journal of molecular biology* **1990,** *216* (2), 386-97.
18. Milner-White, E. J.; Ross, B. M.; Ismail, R.; Belhadj-Mostefa, K.; Poet, R., One type of gamma-turn, rather than the other gives rise to chain-reversal in proteins. *Journal of molecular biology* **1988,** *204* (3), 777-782.
19. Mielke, S. P.; Krishnan, V. V., Characterization of protein secondary structure from NMR chemical shifts. *Prog Nucl Magn Reson Spectrosc* **2009,** *54* (3-4), 141-165.
20. Alberts B, J. A., Lewis J, et al., *Molecular Biology of the Cell.* 4th edition. ed.; 2002.
21. Liu, L.; Mayo, D. J.; Sahu, I. D.; Zhou, A.; Zhang, R.; McCarrick, R. M.; Lorigan, G. A., Determining the Secondary Structure of Membrane Proteins and Peptides Via Electron Spin Echo Envelope Modulation (ESEEM) Spectroscopy. *Methods Enzymol* **2015,** *564*, 289-313.
22. Marinelli, L.; Fornasari, E.; Di Stefano, A.; Turkez, H.; Arslan, M. E.; Eusepi, P.; Ciulla, M.; Cacciatore, I., (R)-alpha-Lipoyl-Gly-l-Pro-l-Glu dimethyl ester as dual acting agent for the treatment of Alzheimer's disease. *Neuropeptides* **2017,** *66*, 52-58.
23. Kumar, S.; Mishra, K. K.; Singh, S. K.; Borish, K.; Dey, S.; Sarkar, B.; Das, A., Observation of a weak intra-residue C5 hydrogen-bond in a dipeptide containing Gly-Pro sequence. *The Journal of Chemical Physics* **2019,** *151* (10), 104309.
24. Burade, S. S.; Saha, T.; Bhuma, N.; Kumbhar, N.; Kotmale, A.; Rajamohanan, P. R.; Gonnade, R. G.; Talukdar, P.; Dhavale, D. D., Self-Assembly of Fluorinated Sugar Amino Acid Derived α,γ-Cyclic Peptides into Transmembrane Anion Transport. *Organic Letters* **2017,** *19* (21), 5948-5951.
25. Burade, S. S.; Shinde, S. V.; Bhuma, N.; Kumbhar, N.; Kotmale, A.; Rajamohanan, P. R.; Gonnade, R. G.; Talukdar, P.; Dhavale, D. D., Acyclic alpha gamma alpha-Tripeptides with Fluorinated- and Nonfluorinated-Furanoid Sugar Framework: Importance of Fluoro Substituent in Reverse-Turn Induced Self-Assembly and Transmembrane Ion-Transport Activity. *J Org Chem* **2017,** *82* (11), 5826-5834.
26. Xia, Y.; McLuckey, S. A., Evolution of instrumentation for the study of gas-phase ion/ion chemistry via mass spectrometry. *J Am Soc Mass Spectrom* **2008,** *19* (2), 173-189.
27. Prentice, B. M.; McLuckey, S. A., Gas-phase ion/ion reactions of peptides and proteins: acid/base, redox, and covalent chemistries. *Chem Commun (Camb)* **2013,** *49* (10), 947-965.
28. Kumar, S.; Biswas, P.; Kaul, I.; Das, A., Competition between Hydrogen Bonding and Dispersion Interactions in the Indole···Pyridine Dimer and (Indole)2···Pyridine Trimer Studied in a Supersonic Jet. *The Journal of Physical Chemistry A* **2011,** *115* (26), 7461-7472.
29. Mandal, A.; Singha, M.; Addy, P. S.; Basak, A., Laser desorption ionization mass spectrometry: Recent progress in matrix-free and label-assisted techniques. *Mass Spectrometry Reviews* **2019,** *38* (1), 3-21.
30. Pauly, H.; Toennies, J. P.; Amdur, I., 3. Neutral-Neutral Interactions. In *Methods in Experimental Physics*, Bederson, B.; Fite, W. L., Eds. Academic Press: 1968; Vol. 7, pp 227-360.
31. Smalley, R. E.; Wharton, L.; Levy, D. H., Molecular optical spectroscopy with supersonic beams and jets. *Accounts of Chemical Research* **1977,** *10* (4), 139-145.

32. Sin, C. H.; Tembreull, R.; Lubman, D. M., Resonant two-photon ionization spectroscopy in supersonic beams for discrimination of disubstituted benzenes in mass spectrometry. *Analytical Chemistry* **1984,** *56* (14), 2776-2781.
33. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X. H., H. P.; Izmaylov, A. F. B., J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *M. J. Frisch et al., GAUSSIAN 09, Revision B.01, Gaussian, Inc., Wallingford CT (2009)*

34. Glendening, E. D.; Landis, C. R.; Weinhold, F., NBO 6.0: Natural bond orbital analysis program. *Journal of Computational Chemistry* **2013,** *34* (16), 1429-1437.
35. Bajaj, K.; Madhusudhan, M. S.; Adkar, B. V.; Chakrabarti, P.; Ramakrishnan, C.; Sali, A.; Varadarajan, R., Stereochemical Criteria for Prediction of the Effects of Proline Mutations on Protein Stability. *PLOS Computational Biology* **2007,** *3* (12), e241.
36. Milner-White, E. J.; Bell, L. H.; Maccallum, P. H., Pyrrolidine ring puckering in cis and trans-proline residues in proteins and polypeptides: Different puckers are favoured in certain situations. *Journal of molecular biology* **1992,** *228* (3), 725-734.
37. Vitagliano, L.; Berisio, R.; Mastrangelo, A.; Mazzarella, L.; Zagari, A., Preferred proline puckerings in cis and trans peptide groups: Implications for collagen stability. *Protein Science* **2001,** *10* (12), 2627-2632.
38. Li, X.-Y.; Wang, Y.-H.; Yang, J.; Cui, W.-Y.; He, P.-J.; Munir, S.; He, P.-F.; Wu, Y.-X.; He, Y.-Q., Acaricidal Activity of Cyclodipeptides from Bacillus amyloliquefaciens W1 against Tetranychus urticae. *Journal of Agricultural and Food Chemistry* **2018,** *66* (39), 10163-10168.
39. Marinelli, L.; Fornasari, E.; Di Stefano, A.; Turkez, H.; Genovese, S.; Epifano, F.; Di Biase, G.; Costantini, E.; D'Angelo, C.; Reale, M.; Cacciatore, I., Synthesis and biological evaluation of novel analogues of Gly-l-Pro-l-Glu (GPE) as neuroprotective agents. *Bioorganic & Medicinal Chemistry Letters* **2019,** *29* (2), 194-198.
40. Ho, B. K.; Brasseur, R., The Ramachandran plots of glycine and pre-proline. *BMC Structural Biology* **2005,** *5* (1), 14.
41. Halgren, T. A., MMFF VI. MMFF94s option for energy minimization studies. *Journal of Computational Chemistry* **1999,** *20* (7), 720-729.
42. Plumley, J. A.; Dannenberg, J. J., A comparison of the behavior of functional/basis set combinations for hydrogen-bonding in the water dimer with emphasis on basis set superposition error. *Journal of computational chemistry* **2011,** *32* (8), 1519-1527.
43. Shao, Q.; Gao, Y. Q., Temperature Dependence of Hydrogen-Bond Stability in β-Hairpin Structures. *Journal of Chemical Theory and Computation* **2010,** *6* (12), 3750-3760.
44. Yatsyna, V.; Mallat, R.; Gorn, T.; Schmitt, M.; Feifel, R.; Rijs, A. M.; Zhaunerchyk, V., Conformational Preferences of Isolated Glycylglycine (Gly-Gly) Investigated with IRMPD-VUV Action Spectroscopy and Advanced Computational Approaches. *The Journal of Physical Chemistry A* **2019,** *123* (4), 862-872.
45. Sohn, W. Y.; Kim, J. J.; Jeon, M.; Aoki, T.; Ishiuchi, S.-i.; Fujii, M.; Kang, H., Entropic effects make a more tightly folded conformer of a β-amino acid less stable: UV-UV hole burning

and IR dip spectroscopy of l-β3-homotryptophan using a laser desorption supersonic jet technique. *Physical Chemistry Chemical Physics* **2018,** *20* (30), 19979-19986.
46. Yatsyna, V.; Mallat, R.; Gorn, T.; Schmitt, M.; Feifel, R.; Rijs, A. M.; Zhaunerchyk, V., Competition between folded and extended structures of alanylalanine (Ala-Ala) in a molecular beam. *Physical Chemistry Chemical Physics* **2019,** *21* (26), 14126-14132.
47. Tian, G.-Z.; Wang, X.-L.; Hu, J.; Wang, X.-B.; Guo, X.-Q.; Yin, J., Recent progress of sugar amino acids: Synthetic strategies and applications as glycomimetics and peptidomimetics. *Chinese Chemical Letters* **2015,** *26* (8), 922-930.
48. Chakraborty, T. K.; Srinivasu, P.; Tapadar, S.; Mohan, B. K., Sugar amino acids and related molecules: Some recent developments. *Journal of Chemical Sciences* **2004,** *116* (4), 187-207.
49. van Well, R. M.; Marinelli, L.; Altona, C.; Erkelens, K.; Siegal, G.; van Raaij, M.; Llamas-Saiz, A. L.; Kessler, H.; Novellino, E.; Lavecchia, A.; van Boom, J. H.; Overhand, M., Conformational Analysis of Furanoid ε-Sugar Amino Acid Containing Cyclic Peptides by NMR Spectroscopy, Molecular Dynamics Simulation, and X-ray Crystallography: Evidence for a Novel Turn Structure. *Journal of the American Chemical Society* **2003,** *125* (36), 10822-10829.
50. Lodish H, B. A., Zipursky SL, et al., Molecular Cell Biology, 4th edition. **2000**.