STRUCTURE AND FUNCTIONAL CHARACTERIZATION OF THE PHEROMONE BINDING PROTEIN 2 FROM *OSTRINIA FURNACALIS*

By

SALIK RAM DAHAL

Bachelor of Science in Chemistry
Tribhuvan University
Pokhara, Nepal
2009

Master of Science in Chemistry
Tribhuvan University
Kathmandu, Nepal
2013

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2021

STRUCTURE AND FUNCTIONAL

CHARACTERIZATION OF THE PHEROMONE

BINDING PROTEIN 2 FROM *OSTRINIA FURNACALIS*

Dissertation Approved:

Prof Dr. Smita Mohanty

Dissertation Adviser

Prof Dr. K. Darrell Berlin

Prof Dr. Richard Bunce

Prof Dr. Junpeng Deng

ACKNOWLEDGEMENTS

I would like to express my highest gratitude to the following people for their help and advice throughout this project. I thank my advisor, Prof. Dr. Smita Mohanty, for her supervision, continuous support, and attention in my training as a scientist. My committee members Prof. Dr. K. Darrell Berlin, Prof. Dr. Richard Bunce, and Prof. Dr. Junpeng Deng, for their encouragement, guidance, support, and fruitful discussions. I appreciate them more than they may realize. This research was financially supported by National Science Foundation Award CHE-1807722 and DBI-1726397 to Smita Mohanty and National High Magnetic Field Laboratory, which is supported by the National Science Foundation Cooperative Agreement No. DMR-1644779 and the State of Florida.

I want to thank Dr. David Zoetewey for his scientific and technical advice. Furthermore, I again thank Dr. Suman Mazumder, Dr. Uma Katre for guiding and teaching. I am also grateful to the Department of Chemistry and Dr. Asfna Iob for helping with the teaching assistant's job. I would also like to thank my current lab partners, Bharat P. Chaudhary, Omar Al-Danoon, and Jacob Lewellen, for their generous support and help during my graduate work.

I would especially like to thanks my wife, Anjana Bhandari Dahal, for her advice, encouragement, moral support, and tolerance. Without her help, I might not have completed this thesis.

I wish to thank my parents Laxmi Dahal and Narayan Prasad Dahal, and two daughters, Aava and Avni, for their motivation, patience, and constant moral support.

Name: SALIK RAM DAHAL

Date of Degree: MAY, 2021

Title of Study: STRUCTURE AND FUNCTIONAL CHARACTERIZATION OF THE PHEROMONE BINDING PROTEIN 2 FROM *OSTRINIA FURNACALIS*

Major Field: CHEMISTRY

Abstract: Animal olfaction has an immense impact on their survival. The insect olfactory system is the most exquisitely sensitive in the animal kingdom. Moth antennae contain hair-like structures called sensilla, which are involved in detecting chemical signals. A male moth can detect pheromone released by the female from a far distance. The hydrophobic pheromone molecules pass through the pores of the sensillum cuticle and enter into the sensillum lymph. Pheromone-bonding protein (PBP) present in the lymph of the sensilla of the male moth antennae binds and transports the pheromone molecules through the aqueous layer to the receptors that initiate signaling, which leads to mating. PBPs bind to pheromone with high affinity at neutral pH in the open conformation and undergo a conformational switch, and release the ligand at acidic pH. Ligand release and binding occur through the concerted pH-dependent mechanism where two molecular gates (the histidine gate, His70-His95, and the C-terminal tail) play a critical role. *Ostrinia furnacalis* is an agricultural insect pest. The *Ostrinia furnacalis* pheromone binding protein 2 (OfurPBP2) has more than 50%, similarly with the well-studied PBPs including Antheraea Polyphemus pheromone binding protein1 (ApolPBP1) and Bombyx mori pheromone binding protein (BmorPBP). However, there are remarkable differences in both biological gates; a) one of the histidine-gate residues, His70, is substituted by arginine, b) the C-terminal tail has seven charged residues as compared to three. The molecular impact of these substitutions on structure and mechanism of action is unknown. Furthermore, structure and mechanistic studies of several of these proteins are needed to gain the knowledge to design inhibitors through pheromone mimetics, which constitutes a novel mechanism to control these pests.

Our work has dissected the structural details to understand the structural mechanism of pheromone binding and release in this pest. NMR investigations have shown that OfurPBP2 undergoes conformational heterogeneity at acidic pH of 4.5. We have used small-angle X-ray scattering (SAXS) to show the protein is homogeneous, well-folded, and has a compact globular shape. OfurPBP2 consists of seven helices with residues 2-14 (α1a), 16-22(α1b), 27–37 (α2), 46–60 (α3), 70–80 (α4), 84–100 (α5), 107–124(α6), and 131-143 (α7) which are arranged in a globular fold, and contains the three disulfide bridges 19-54, 50-108, and 97-117 enclosing a large hydrophobic binding pocket inside. The structure of the OfurPBP2 contains a C-terminal helix (α7) residues 131-143 extended outside the hydrophobic pocket, which is in contrast with previously studied PBPs, where they have a random coil at pH 6.5. OfurPBP2 binds the pheromones at high pH. The MD simulations were carried to identify the flexible region in the protein structure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

CHAPTER I

LITERATURE REVIEW

1.1 Nuclear Magnetic Resonance (NMR)

1.1.1. The brief history of Nuclear Magnetic Resonance (NMR)

The history of NMR started with the prediction of nuclear spin by Pauli in 1924. Stern–Gerlach experiment in the early 1920s established the concepts of electron spin and magnetic moment of the electron. In this experiment, an applied inhomogeneous magnetic field separates the beams of atoms according to the orientation of the electron magnetic moment[1], which was the experiment to detect a spin magnetic moment. In 1939 I. I Rabi, et al.[2] from Columbia University applied a homogenous magnetic field and radiofrequency electromagnetic energy to the hydrogen molecules simultaneously. Hydrogen atoms absorb the energy and cause deflection of the beam. This deflection depends on the magnetic moments of the atoms, which was a significant breakthrough. In 1944, the first Nobel Prize in physics was awarded to Rabi for the development of the resonance method that enabled the recording of the magnetic properties of atomic nuclei. However, his studies were only limited to small molecules and were observed in a very high vacuum. Felix Bloch from Stanford University[3] and Edward Mills Purcell from Harvard University[4] were able to detect Nuclear Magnetic Resonance successfully in condensed matter.

Russell H. Varian built the first commercial NMR spectrometer under the company called VARIAN in 1951. Albert Overhauser proposed the concept of the Nuclear Overhauser Effect (NOE) in 1953.[5] In 1964, Ernst and Anderson implemented the first Fourier Transform NMR (FT-NMR) on a Varian spectrometer. The development of the FT-NMR method was a revolutionary event in terms of sensitivity enhancements. In 1971, Jean Jeener laid the foundation of the pulse program for two-dimensional experiments. Richard Ernst successfully introduced his idea of two-dimensional nuclear magnetic resonance spectroscopy (2D NMR) to produce the 2D COSY spectra in 1974.[6] In 1991, Richard Ernst was awarded the Nobel Prize in Chemistry for his contributions to the development of the methodology of high-resolution Nuclear Magnetic Resonance (NMR).

Macromolecular structure study by NMR started with introducing high field magnets that separate spectral lines caused by chemically distinct nuclei. Nagayama, Wuthrich, Bachmann, Anil Kumar, and Ernst developed various 2D-NMR techniques to solve protein structures.[7] In 1979, Kurt Wuthrich used Nuclear Overhauser Effect (NOE) measurements for spectral assignments to obtain internuclear distances in three-dimensional (3D) space. In 1982, Gerhard Wagner and Wüthrich published the first sequence-specific assignments on pancreatic trypsin inhibitor.[8] The Wüthrich group developed an algorithm to calculate protein structure from NMR data. In 1985, Michael Williamson, Havel, and Wüthrich reported the first solution-state protein structure of proteinase inhibitor IIA from bull seminal plasma.[9] In 1990, 3D NMR was introduced on unlabeled proteins by the use of triple resonance experiments. In 1997, TROSY (transverse relaxation-optimized spectroscopy) was introduced to calculate the structure's higher molecular weight proteins. Kurt Wüthrich was awarded the Kyoto Prize in 1998 and the Nobel Prize in Chemistry in 2002 for the development of a method to determine the three-dimensional structure of biological macromolecules in solution.[10] Along with Kurt Wuthrich, other distinguished scientists such as Ad Bax, Marius Clore, Dennis Torchia, and Lewis Kay also pioneered in protein structure determination and method development. Ad Bax is a pioneer in developing triple resonance experiments[11], and technology for resonance assignments of isotopically-

enriched proteins, and residual dipolar coupling.[12] Marius Clore is known for the development of three and four-dimensional experimental approaches to study large macromolecules by NMR.[13] Dennis Torchia is well recognized for using isotope labeling and developing techniques for studying protein dynamics. Due to the advancement of isotopic labeling strategies, like deuterium labeling, selective residue labeling, selective methyl labeling, and TROSY-based experiments, the study of higher MW protein is possible.

1.1.2. Basic of Nuclear Magnetic Resonance (NMR)

Some nuclei possess a property called angular momentum or nuclear spin. The nuclei with an ($I \neq 0$) non zero spins are NMR active. These nuclei are charged and spin around on their axis, behaving like a tiny magnet. The nuclear spin is characterized by a spin quantum number 'I', and the magnitude is described by an angular momentum given by equation 1.

$$\boldsymbol{L} = \hbar\sqrt{\boldsymbol{I}(\boldsymbol{I}+\boldsymbol{1})} \tag{1}$$

The L is angular momentum, I is the spin quantum number, and $\hbar$ is Planck's constant divided by 2π. These nuclei possess a nuclear magnetic moment μ when placed in the magnetic field proportional to its spin I (Equation 2).

$$\mu = \frac{\boldsymbol{\gamma I h}}{\boldsymbol{2\,\pi}} \tag{2}$$

The constant γ, is called the gyromagnetic ratio and is a fundamental nuclear constant with a different value for every nucleus, and h is the Planck's constant.

Table: 1: Properties of NMR active nuclei (Adapted from reference 14)[14].

Nuclei	γ (rad s^{-1} T^{-1}) Magnetogyric Ratio	$\gamma/2\pi$ (MHz.T^{-1})	Relative Frequency (ν) MHz	I (Spin)	Natural Abundance(%)
^{1}H	2.6752 * 10^{8}	42.57	100	1/2	99.980
^{2}H	4.106* 10^{7}	6.53	15.35	1	0.016
^{13}C	6.728* 10^{7}	10.70	25.15	1/2	1.108
^{15}N	-2.712* 10^{7}	-4.31	10.14	1/2	0.37
^{19}F	2.5179* 10^{8}	40.05	94.13	1/2	100
^{31}P	1.0841* 10^{8}	17.23	40.52	1/2	100

The nucleus with a spin, I, has 2I+1 possible orientations or states where each state is associated with a different potential energy. In the absence of an external magnetic field, these orientations have equal energy, called zero-field splitting. When the magnetic field is applied, the energy levels split; this is called Zeeman splitting. Each energy level is characterized by the magnetic quantum number and is shown diagrammatically as a function of magnetic field strength in Figure 1.1, using a nucleus of I = ½ as an example.

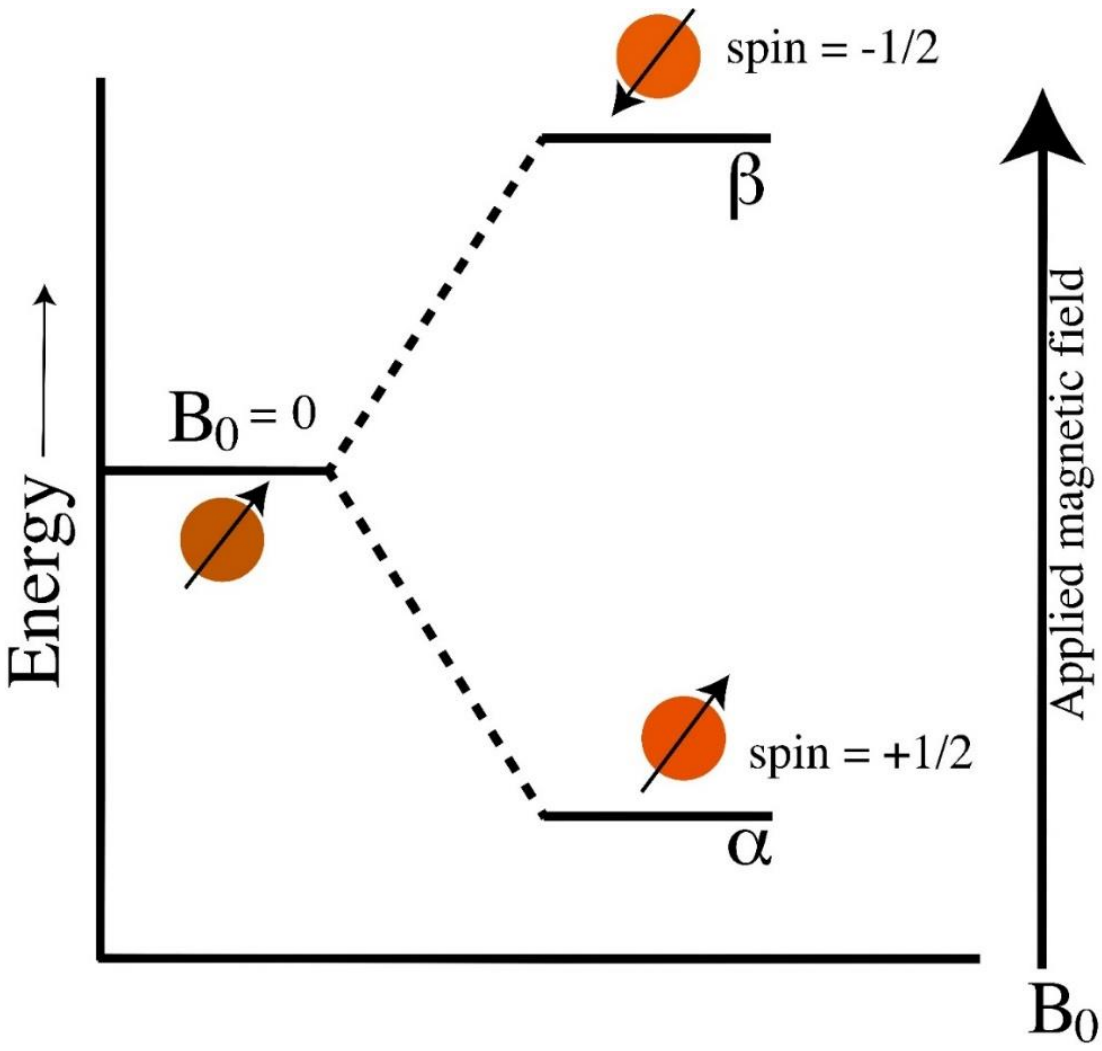


Figure 1.1: Energy splitting as a function of magnetic field strength with spin quantum number 1/2. Figure adapted from reference (14)[14].

The spin population in the two energy levels depends on the energy difference between the two states. The energy difference relies on the magnitude of the external magnetic field and the sensitivity of the nucleus. The ^{1}H resonance frequency expresses the external magnetic field for a NMR instrument. The

nucleus's sensitivity depends upon the gyromagnetic ratio, γ, which is constant for the particular isotope. The energy between the states is shown in equation 3.

$$\boldsymbol{\Delta E = \frac{\gamma h B}{2\pi}} \tag{3}$$

When the nucleus is in a magnetic field, the initial population energy levels are determined by the Boltzmann distribution (Equation 4).

$$\boldsymbol{\frac{N_{\beta}}{N\alpha} = e^{-\Delta E/kT}} \tag{4}$$

The lower energy level contains slightly more nuclei, approximately 0.001% of the proton's total number for a 600 MHz NMR compared to the higher energy level. The nuclei can be excited from the lower energy level into the higher level by applying electromagnetic radiation. The exact frequency of radiation needed to excite the nucleus is determined by the difference in energy between the energy levels given by equation 5.

$$\boldsymbol{\Delta E = h\nu = \frac{\mu B}{I}} \tag{5}$$

Where B, is the strength of the magnetic field at the nucleus.

$$\boldsymbol{\nu = \gamma Bo/2\pi} \tag{6}$$

The ν refers to the absorption or resonance frequency of the shielded nucleus, i.e., observed resonance frequency, and γ, is the gyromagnetic ratio. Each nucleus has its characteristic gyromagnetic ratio (Table1). The gyromagnetic ratio is listed in Table 1. The value of the gyromagnetic ratio for ^{1}H is approximately ten times greater than ^{15}N and four times greater than ^{13}C.[14] Equation 6 is known as the Larmor equation. It states that the absorption frequency of transition is equal to γ multiplied by the nucleus's magnetic field strength. The nucleus is spinning on its axis (Figure 1.2). In the presence of a magnetic field, this axis of rotation will precess around the magnetic field ($\Delta E = h\nu$) at a given frequency, ν, where there is a transition of spin between the states. This frequency of precession is

termed as the Larmor frequency related to the strength of the magnetic field, B_o, which is identical to the transition frequency or resonance frequency.

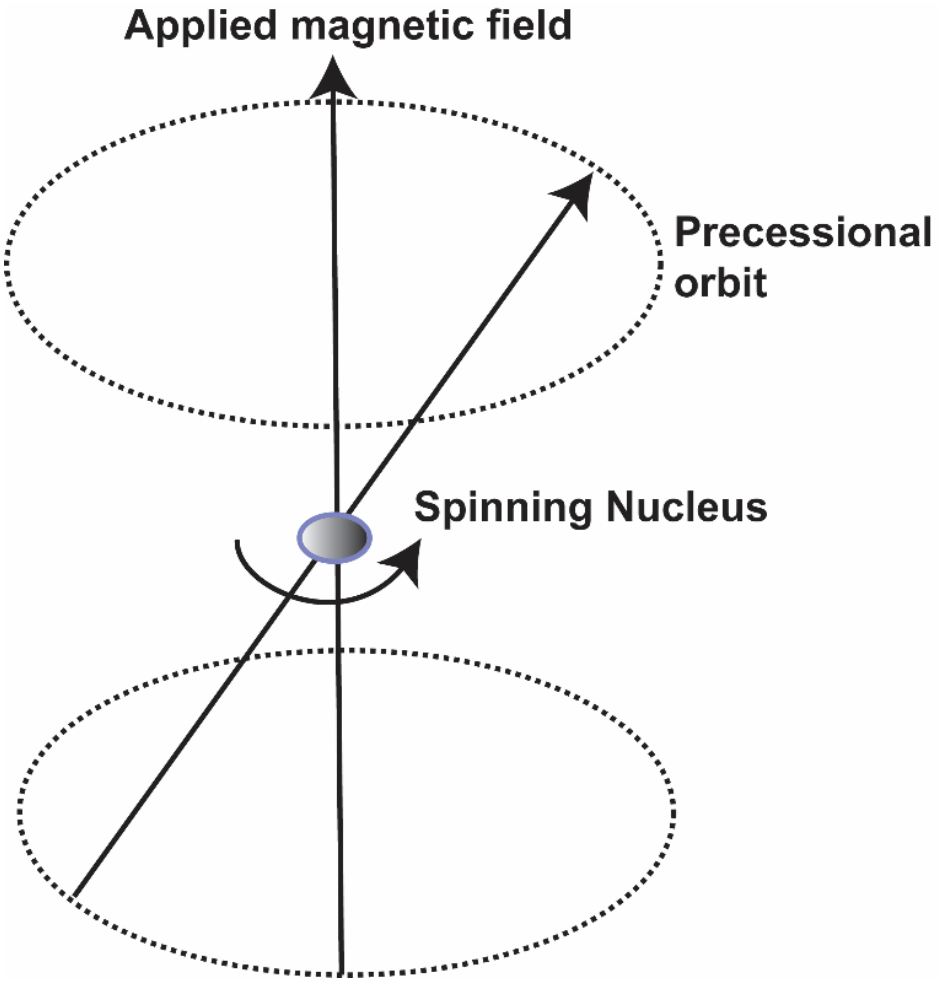


Figure 1.2: Nucleus spinning on its axis with precession frequency in the presence of the applied magnetic field. Figure adapted from reference (14)[14].

The NMR spectroscopy is a less sensitive technique because of the tiny energy differences between α and β states, resulting in the small excess population of nuclei in the ground state versus the excited states (1 in 10^6). The signal is thus proportional to the population difference between the states. Because of low sensitivity, NMR experiments require high protein concentrations (~ 1 mM). The excited state's lifetime is on a millisecond order to second, which is usually beneficial for getting a narrow resonance signal and enough time to manipulate the excited state in a multidimensional experiment. Although NMR lines are very sharp, the lifetime of a given energy state and the relaxation rate play a dominant role in the spectra due to Heisenberg's uncertainty. When the equilibrium is perturbed, the system will take a specific time to return to its original equilibrium state. This process of returning excited spins into a lower energy state is called relaxation. The successful collection of multidimensional spectra depends on the proper consideration of relaxation times.

There are two major relaxation processes:

a. Spin - Lattice relaxation (longitudinal relaxation, T1)
b. Spin - Spin relaxation (transverse relaxation, T2)

The population of the spins at an energy state is given by the Boltzmann distribution. When a molecule is subjected to an external energy source, some of the spins at a lower energy level are excited to the higher energy level. As the spins lose energy and go to the lower energy state, they dissipate energy to the surroundings or the lattice by a process called T1 relaxation. Due to the T1 relaxation, the normal Gaussian population distribution of α and β spin states is established. If T1 is very large, the time to reach the thermal equilibrium is very long, which increases the data collection time. A short T1 is

responsible for acquiring the spectrum in less time. T1, spin-lattice relaxation, does not involve a change in entropy; it is an enthalpy-driven process. T1 is strongly dependent on the magnetic field, and a higher magnetic field generally leads to a slower T1. In T2 relaxations, the spins lose their phase coherence among other nuclei then return to the equilibrium. The time required to reach equilibrium is known as spin-spin relaxation. Both relaxation processes are correlated because when the magnetization is returned to the z-direction, it causes the loss of magnetization in the x-y plane. The T2 is less than or equal to T1 relaxation.[15]

Chemical Shift

The Larmor frequency depends on the strength of the magnetic field. For a single isolated nucleus, the field strength is equal to the external field. But for macromolecules, like proteins, different nuclei are connected with chemical bonds. The magnetic field at a given nucleus depends not only on an applied magnetic field but also on the nucleus local chemical environment, i.e., atoms and electrons around the nucleus either oppose or enhance the applied magnetic field. The difference between the applied magnetic field and the actual field at the nucleus due to local electron density is called nuclear shielding. Due to the local chemical environmental effect, different nuclei in a molecule resonate at slightly different frequencies. The frequency shift of a particular nucleus is called its chemical shift and is given by equation 7.

$$\boldsymbol{\nu = \frac{\gamma(1-\sigma)B}{2\pi}} \tag{7}$$

where σ is the average isotropic shielding constant of the nucleus, and B is the magnetic field.

The strength of the magnetic field, B, varies according to the strength of NMR, and it results in the resonance frequency of the same nucleus varying in different machines. Instead of measuring absolute frequency, the unitless normalized parameter called the chemical shift, δ, of the nucleus is used, which is independent of the instrument's magnetic field.[16] The chemical shift is expressed in parts per million (ppm).

$$\boldsymbol{\delta} = \frac{\boldsymbol{\nu^{sample} - \nu^{ref}}}{\boldsymbol{\nu^{ref}}} * 10^6\ ppm \qquad (8)$$

Coupling

For a multi-dimensional spectrum, magnetization is correlated among the spins, and this takes place mainly in two ways: (a) through chemical bonds (scalar or J coupling or indirect dipole-dipole) and (b) through space (dipolar coupling).

a) Spin-Spin Coupling or Scalar Coupling

The isotropic magnetic interaction between nuclei through chemical bonds is called spin-spin coupling. It is also called scalar coupling or J-coupling. It does not depend on the field strength. The magnitude of scalar coupling depends only on the interaction of the nuclear magnetic dipoles. The main cause of scalar coupling is the indirect magnetic interaction of electrons involved in the chemical bonds. It is negligible when more than three bonds separate two nuclei. It is a mutual interaction that is observed by a splitting of the NMR signal. The frequency difference between the splitting signals lines is called the J-coupling constant. A scalar coupling pattern is useful to obtain the connectivity of atoms in a molecule. The J-couplings provide information on the dihedral angle and bond distances.

b) Dipolar Coupling

Dipolar coupling is the interaction between two spins through space. It depends on both the distance and orientation of the two spins. Dipolar couplings are not directly observable in isotropic solutions due to the fast isotropic tumbling of the molecule. The splitting due to dipolar coupling is not noticeable

since the average dipolar coupling effect will be zero with time. However, they are responsible for the Nuclear Overhauser Effect (NOE) phenomenon. There is a change in intensity of one nucleus signal when the nearby nucleus signal (to which the first is dipolar coupled) is perturbed. Dipolar coupling helps to determine molecular structures and is used to investigate many other phenomena involving interactions.

Nuclear Overhauser Effect (NOE)

In NOE, nuclear spin polarization transfers from one nuclear spin population to another via cross-relaxation. The Nuclear Overhauser Enhancement or Effect (NOE) is the cross-relaxation (i.e., both spins simultaneously relax to their lower or higher energy state and undergo simultaneous spin flips) polarization from one spin to another induced by dipole-dipole interaction. The amplitude depends on the separation of two spins and is used to measure the distance between them. The physical process for NOE is nuclear relaxation. There is a change of one resonance when the transition of another is perturbed. The cross-relaxation rates depend on the distance by d^{-6}.

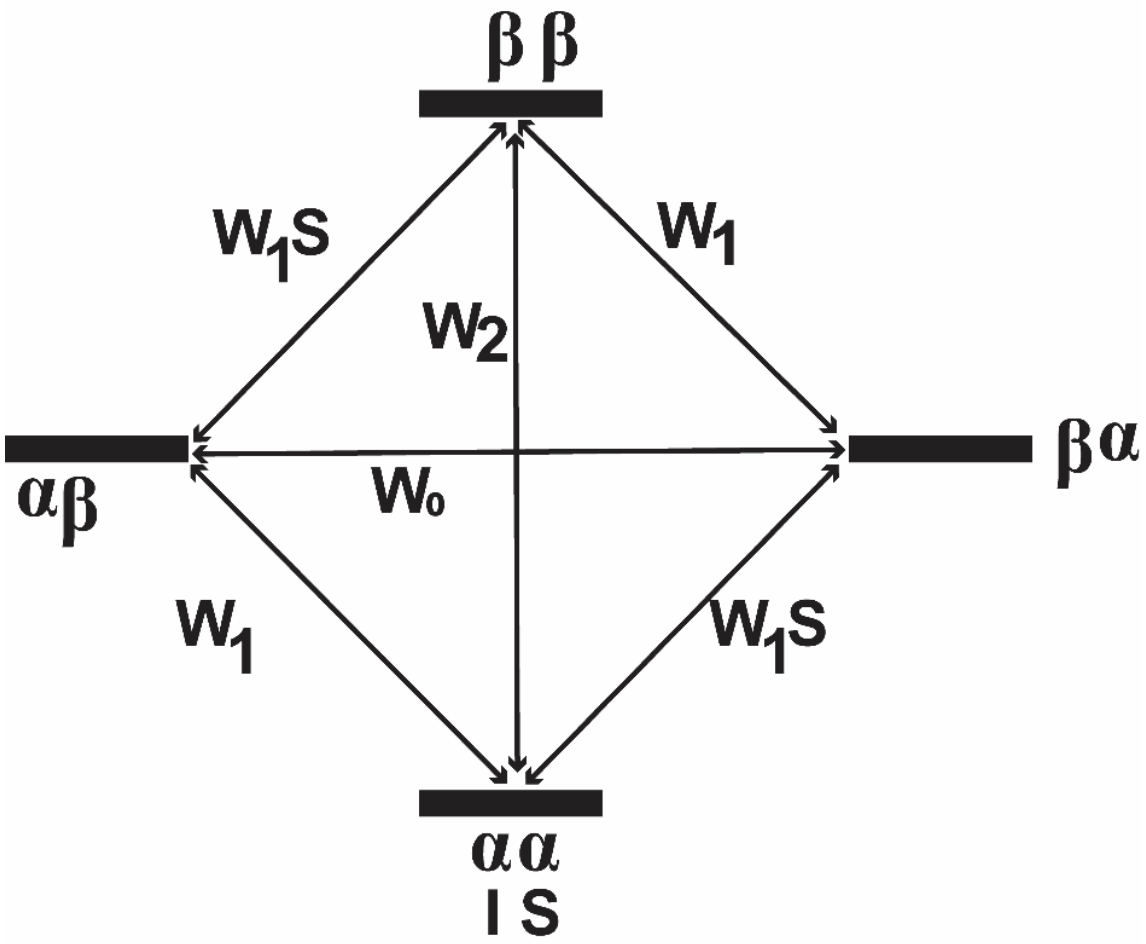


Figure 1.3: Energy diagram for a dipolar-coupled two-spin system. The four states are αα, αβ, βα, and ββ; the zero- single- and double-quantum transitions are represented by w0, w1, and w2, respectively, drawn according to reference.[17]

The two spin, S and I have two spin states (Figure 1.3), α and β, respectively, coupled with the dipolar coupling. These two nuclei after coupling contain four energy states: αα (representing both spins at lower energy states), αβ (S at lower and I at higher energy states), βα (S at higher and I at lower energy states), and ββ (both spins at higher energy states). When the RF pulse is removed, the source spins (S) relax back to a lower energy state via two pathways, which are the primary source of NOE; (1) double quantum transition (W_2 pathways). Both source spin (S) relax back, with the increase in the population

difference giving rise to an increase in signal intensity. (2) Zero-order quantum transition (Wo pathways) is when two anti-parallel spins simultaneously flip, which leads to a decrease in signal intensity. The W_2 exists for small molecules, whereas the Wo transition favors the large molecules, like proteins, where the molecule tumbles slowly, leading to an intensity reduction for spin S, causing a negative NOE.

The molecules which have molecular weights of ~1000-3000 Daltons have very weak NOE because the two relaxation pathways explained above compete for each other in the system. The NOE intensity depends on the distances between the two spin separated by distance r, and given by equation 9.

$$nOe = (1/r)^6 \tag{9}$$

where, r is the distance between two nuclei, where $r \leq 5$ Å.

NOE is only detected between two spins when they are within 5 Å in the space. The intensity is proportional to r^{-6}. When the distance between two spins increases then the intensity of NOE decreases.[18] The intensities measure the internuclear distance. The cross-peak intensities are converted into distances and classified into three groups: strong, medium, and weak, based on their intensities and assigning their distance ranges of 1.8–2.7 Å, 1.8–3.3 Å, and 1.8–5.0 Å, respectively.

1.1.3. NMR in the Structural Biology

Proteins are essential biomolecules in all living organisms. The determination of a three-dimensional structure at the atomic level is the first step to understand its functions. There are currently three different experimental methods available to determine three-dimensional structures of proteins at atomic levels. They are X-ray crystallography, cryogenic electron microscopy (cryo-EM), and NMR spectroscopy. X-ray crystallography is used to obtain the three-dimensional structure of a protein by X-ray diffraction. In cryo-EM, the images are collected to get high-resolution information. In NMR

spectroscopy, information on the distance between atoms can be obtained. NMR spectroscopy is routinely used to study structure, dynamics, and interaction studies for small to medium-sized biological molecules in solution. NMR can provide information on protein dynamics from a picosecond to a second-time scale.

However, NMR has many challenges, such as molecular size limitation, the need for highly concentrated protein samples, low sensitivity, and isotopic enrichment with ^{15}N, ^{13}C, and ^{2}H nuclei. The data acquisition process is also time-consuming (it takes several days). The recent advancement in technology, such as using new superconducting material to increase the magnetic field, using cryoprobes to reduce the noise, and the design of long-term stable electronics, have been advantageous. The major limitation of NMR is the low inherent sensitivity and resolution. This problem is partially alleviated by spectrometer technology's progress, i.e., the development of 1.2 GHz ultra-high field NMR magnets. The development of biochemical methods (recombinant protein expression) allows for fast and straightforward protein sample preparation. Uniform or selective isotopic labeling of hetero-nuclei like ^{15}N, ^{13}C, and ^{2}H simplifies the spectra to a great extent. NMR has major advantages since it allows a study of protein close to the physiological conditions in solution state, temperature, and pH. NMR can also reveal study protein-protein or protein-nucleic acid or protein-peptides/drug interactions. NMR studies of protein dynamics at different time scales can be conducted.[19] The NMR relaxation measurements (spin-lattice and spin-spin relaxation) provide information on the internal motion, conformational exchange, protein folding, ligand binding, reaction kinetics, and molecular recognition.

a. 1D NMR Spectroscopy

One dimensional NMR spectrum is useful to identify the protein folding. The 1D spectra (Figure 1.4) are too complex for interpretation because signals are heavily overlapped. For the protein NMR, the

range is 0 to 12 ppm. The carrier frequency is centered on the water signal at 4.7 ppm. 1D NMR is routinely collected to optimize NMR parameters like spectral width, the number of scans, and 90° pulse before proceeding to 3D NMR.

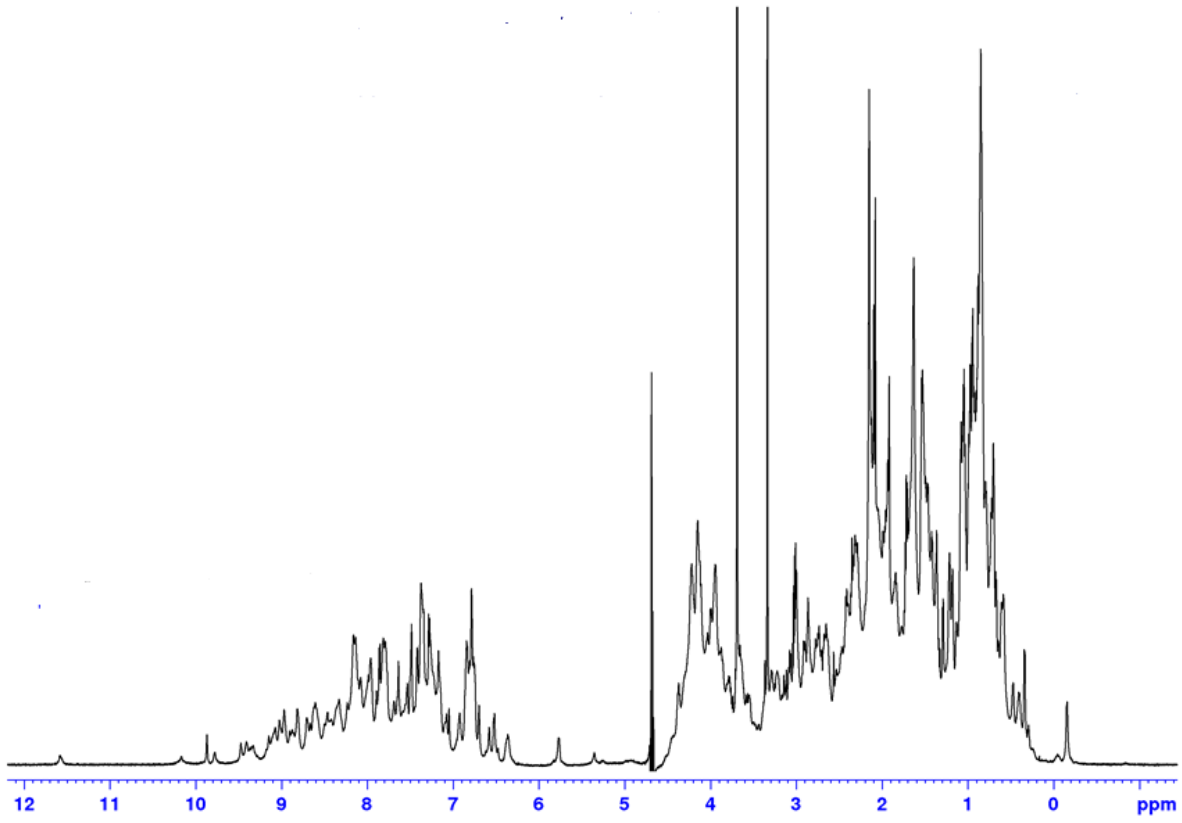


Figure 1.4: Refolded OfurPBP2 collected in AV NEO 700 MHz, Probe 5 mm TXO Z-GRD, 50 mM phosphate buffer, 400 uM, 10 % D_2O.

b. Multidimensional NMR spectroscopy

1. **2D NMR**

The 2D NMR experiments are the collection of series of one-dimensional experiments with a different evolution time. The one-dimensional spectra can be simplified by the introduction of an additional

dimension. In 2D NMR spectroscopy, the combination of different pulse sequences is employed to obtain additional information. The general scheme for two-dimensional spectroscopy is shown in Figure 1.5. In the 2D experiment, the sample is excited by one or more pulses during preparation time. The resulting magnetization is allowed to evolve on time, t1, called evolution. During the mixing time, magnetization is transferred from the first nucleus to a second one, involving further pulses. After the mixing period, the signal is recorded as a function of the second time variable called t2. This sequence of events is called the pulse sequence. The general scheme for two-dimensional spectroscopy is shown in figure 1.5. In the 2D experiment, the sample is excited by one or more pulses during preparation time. The resulting magnetization is allowed to evolve on time, t1 called evolution. During the mixing time, magnetization is then transferred from the first nucleus to a second one, which consists of further pulses. After the mixing period, the signal is recorded as a function of the second time variable called t2. This sequence of events is called the pulse sequence.

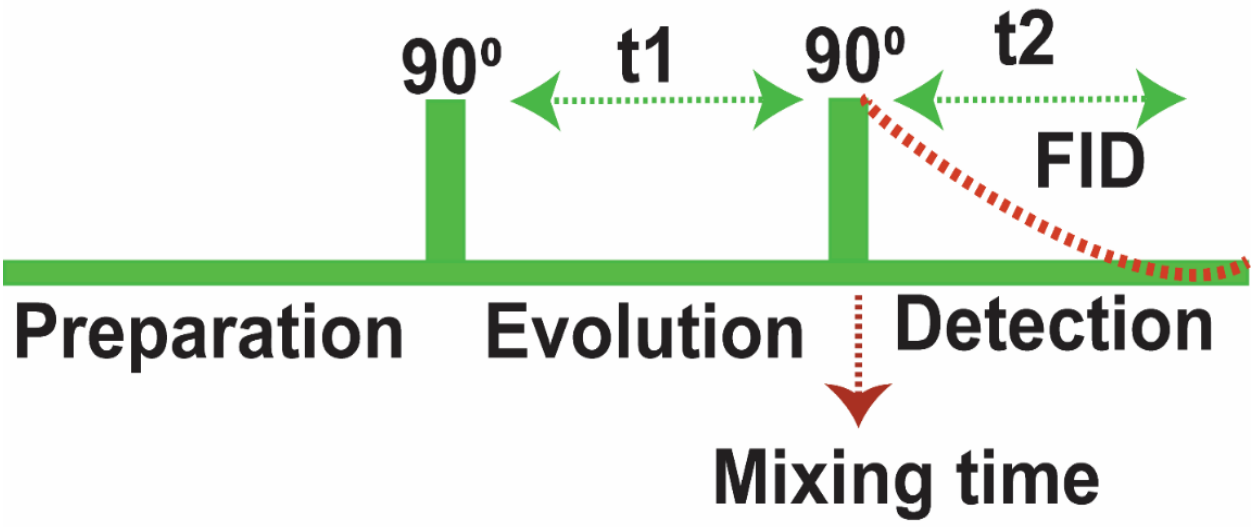


Figure 1.5: General scheme for a 2D experiment. Figure adapted from reference.[20]

There are two types of 2D NMR: homonuclear and heteronuclear NMR.

Homonuclear 2D Spectroscopy

i. Correlated Spectroscopy (COSY)

Correlated Spectroscopy is a homonuclear NMR. In this experiment, magnetization transfer by scalar coupling between the protons that are connected to three chemical bonds.[20] Although it is the simplest experiment, it has low resolution and relatively low sensitivity. This experiment is used routinely for the assignment of low molecular weight protein, there is little resonance overlap. There are many modified versions of the basic COSY experiment, such as DQF-COSY (double-quantum filtered), COSY45, LRCOSY, and ECOSY.

ii. Total Correlation Spectroscopy (TOCSY)

In TOCSY experiments, cross-peaks are observed through bond correlations via spin-spin coupling The TOCSY experiment uses an isotropic mixing time to obtain a cross peak between all hydrogen nuclei of a coupled spin by applying a particular RF frequency. During this mixing time, the coherence transfer takes place through the scalar coupling. As a result, resonances of all the protons attached to a particular spin system are observed. Figure 1.6 provides a general idea of the TOCSY experiment.

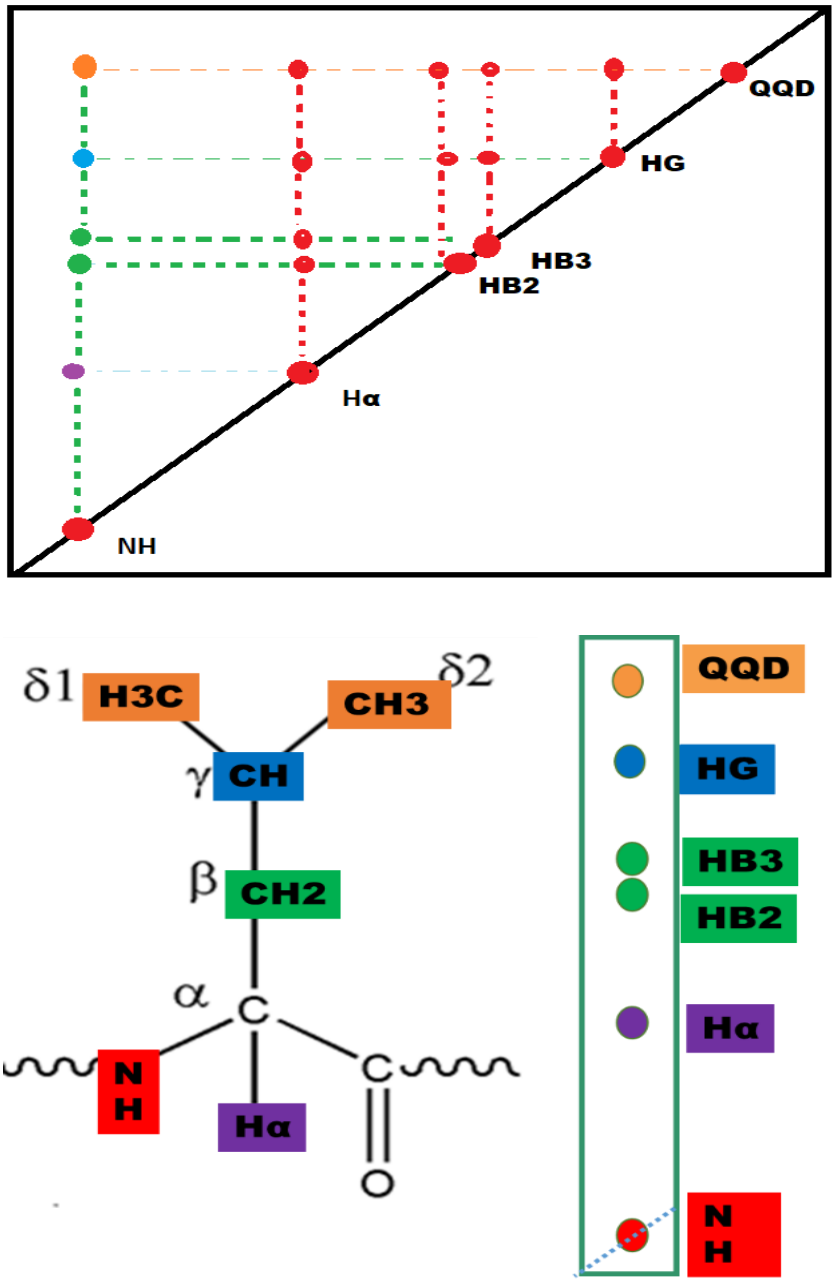


Figure 1.6: Schematic diagram showing TOCSY peak**s.**

iii. Nuclear Overhauser Effect Spectroscopy (NOESY)

NOESY uses the dipole-dipole interaction to correlate pairs of nuclei close in space. The NOE intensity is a first approximation proportional to r^{-6}, where r is the distance between the protons. The correlation between two protons depends on the distance between them, but a signal is usually observed if their distance is smaller than 5 Å. NOESY is the most powerful NMR technique for determining the 3-dimensional structure of molecules. The intensity of cross-peaks provides the distance between two nuclei in NOESY.

Heteronuclear 2D Spectroscopy

In heteronuclear spectroscopy, the signals are obtained by the coupling between the nuclei of two different types. The common heteronuclear 2D experiments are HSQC, HMQC, and HMBC.

i. HSQC (Heteronuclear Single Quantum Coherence)

HSQC spectrum has two axes, one is for the proton chemical shift, and another is for the heteronuclear ($^{13}C/^{15}N$) chemical shift. The HSQC experiment utilizes the INEPT (Insensitive Nuclei Enhancement by Polarization Transfer)[21] sequence to transfer the magnetization of the proton to its bonded hetero nuclei (^{13}C or ^{15}N). The magnetization is then transferred back to the proton by a second INEPT sequence for detection. In HSQC, magnetization transfer from a sensitive proton-nucleus to an insensitive ^{13}C or ^{15}N nucleus occurs via direct J coupling. The 2D spectrum detects HN chemical shifts for each amino acid residue. HSQC spectrum is considered as a protein "fingerprint" or protein signature. It also gives side-chain correlations of H^{ε}–N^{ε} from arginine, H^{ε}–N^{ε} from tryptophan, $H^{\delta 21,22}$ –$N^{\delta 2}$ from asparagine and $H^{\varepsilon 21,\ 22}$–$N^{\varepsilon 2}$ from glutamine residues. Furthermore, Arg Nη-Hη and Lys Nζ-

Hζ side chains can also be visible. However, most of the time, they are also folded, and their chemical shifts are within the 85-90 ppm range. The side-chain peak displays distinct appearances; the smaller peak may appear on top of each peak due to deuterium exchange from D_2O added to an NMR sample. The tryptophan side chains are shifted downfield. There are no NH-signals from proline in the HSQC spectrum as this amino acid lacks an amide proton. Sometimes other backbone correlations might also be missing due to solvent exchange and conformational exchange.

c. Three-Dimensional NMR (3D NMR)

The interpretation of signals for the large proteins was challenging because of signal overlapping due to small chemical shift dispersion for proteins larger than 10 kDa. The limitation of 2D NMR can overcome by adding an extra dimension to the spectra. These experiments are called triple resonance experiments because three different nuclei (^{1}H, ^{13}C, ^{15}N) are correlated. These experiments need double and triple labeled (^{13}C, ^{15}N) proteins. The heteronuclear 3D NMR experiments overcome the overlapping problem (Figure 1.7). The chemical shift dispersion of ^{15}N (100-135 ppm), ^{13}C aliphatic (10-70ppm), ^{13}C aromatic (110-140ppm), and carbonyl (165-185ppm) are larger than those of ^{1}H dispersion. The enhancement of sensitivity is the other advantage of triple resonance spectra. The magnetization is transferred via ^{1}J or ^{2}J couplings (i.e., directly via the covalent chemical bonds). Therefore, the transfer times are shorter, and the signal losses due to the relaxation are smaller than in homonuclear experiments.

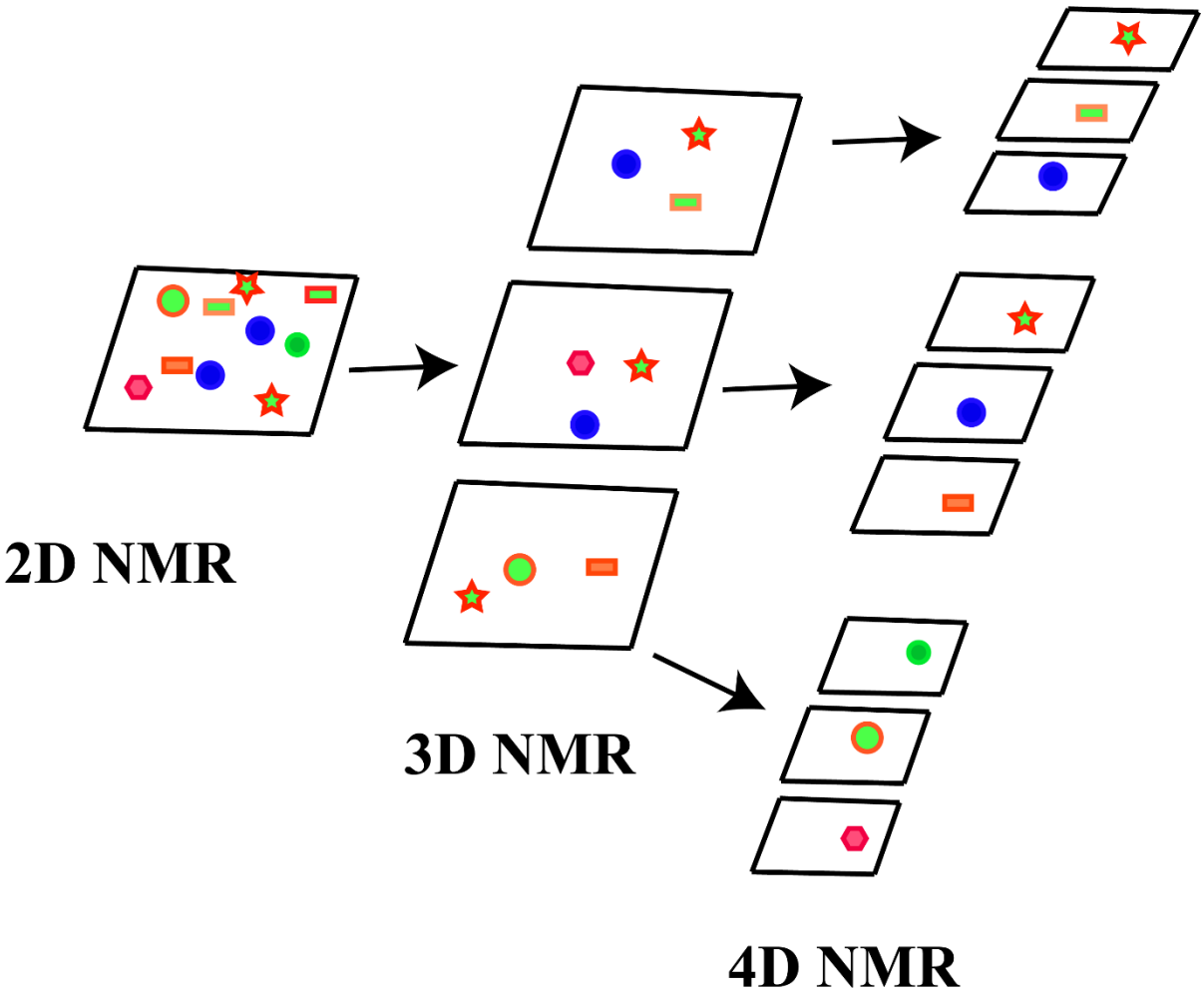


Figure 1.7: Schematic diagram showing NMR data set where peak overlapping is resolving along with the increase in the dimensions.

1.1.4. Protein NMR

a. Isotope Labeling of Proteins

In the absence of ^{15}N and ^{13}C isotopes, only ^{1}H is NMR active. Due to the presence of many H atoms in a large protein, there is severed overlapping of peaks. The structure determination by solution NMR spectroscopy is based on uniformly labeled active isotope (^{13}C and ^{15}N) to alleviate

resonances overlap and to resolve the degeneracy of the chemical shift. Stable isotope labeling has been well established by the manipulation of culturing conditions in the bacterial system: the minimal media containing ^{13}C glucose and ^{15}N ammonium chloride as a carbon and nitrogen source, respectively.

b. Data Acquisitions and Resonance Assignments

The collection of multi-dimensional NMR decreases the overlapping of signals. After collecting NMR data, the spectra are processed, and resonances need to be picked manually. Multiple spectra require cross-referencing, and chemical shifts need to be assigned.

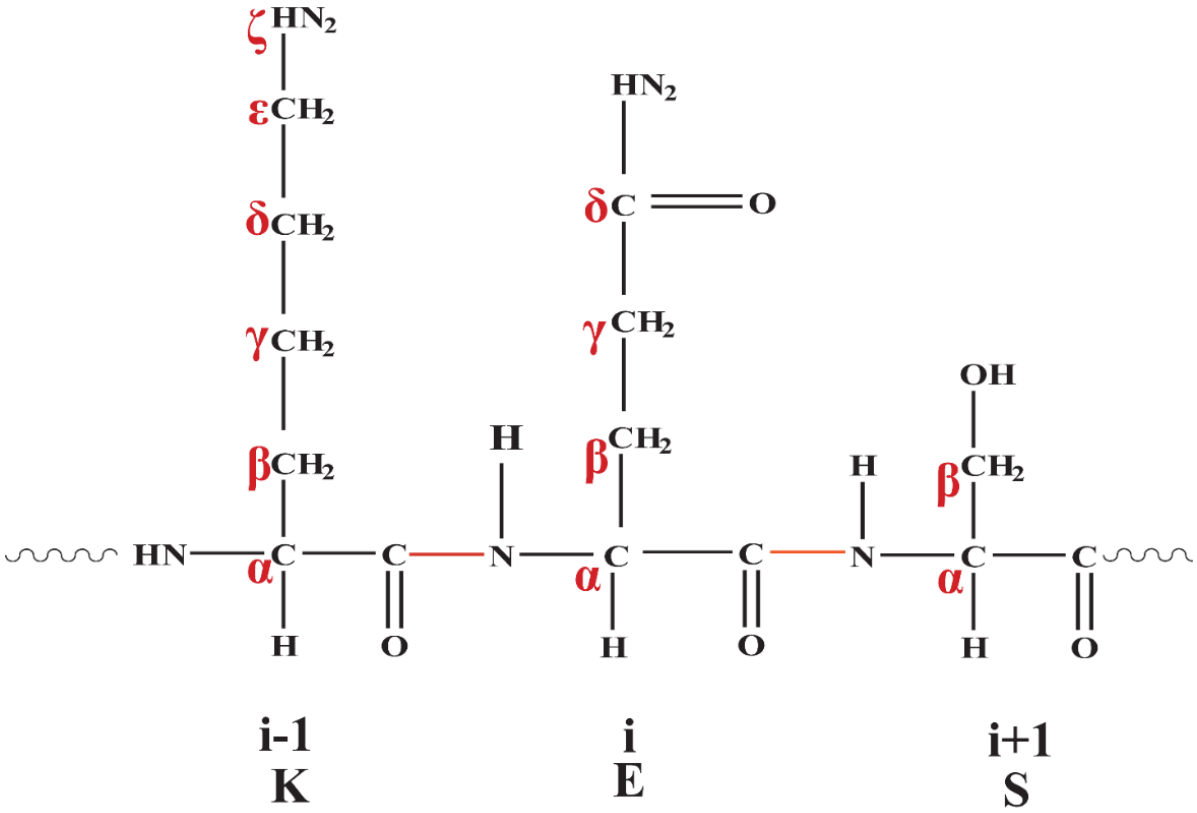


Figure 1.8: Showing i, i-1, i+1, amino acid sequence.

c. Backbone resonance assignment

(i) HNCACB: It is a triple resonance experiment where the chemical shift is correlated to the amide of a residue with the C_α and C_β of the same residue (i th residue) and those of the preceding residue (i-1 residue). In this experiment, the magnetization is transferred from H_α and H_β to $13C_\alpha$ and $13C_\beta$, respectively, and then from $13C_\beta$ to $13C_\alpha$, from here it is transferred to 15N and then to finally to HN for the detection as shown in Figure 1.9. The sensitivity of this experiment is less. Each strip plot contains four signals per residue two from the intra-residue and two from the inter-residue. The C_α and C_β peaks are easily distinguished by color since they have opposite phase signals.

(ii) CBCA(CO)NH: It is a sensitive triple resonance experiment that correlates the resonances of the amide of a residue with the Cα and Cβ of the preceding residue only (i−1 residue). In this experiment, the magnetization is transferred from H_α and H_β to $13C_\alpha$ and $13C_\beta$, respectively, and then from $13C_\beta$ to $13C_\alpha$ of preceding residue, from here it is transferred to ^{15}N then to finally to HN for the detection as shown in Figure 1.9. There are only two cross-peaks observed in this experiment per residue. Sequential assignments can be accomplished with the help of chemical shifts of HNCACB and CBCACONH experiments.

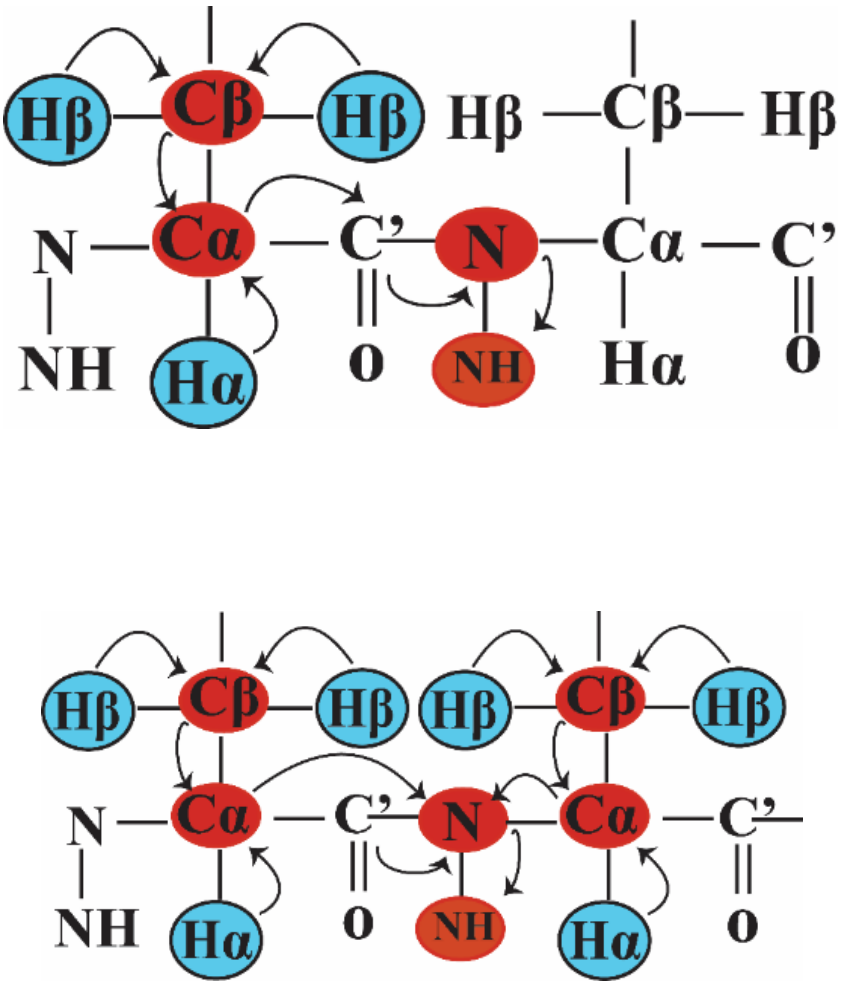


Figure 1.9: Magnetization transfer in CBCACONH and HNCACB experiments respectively**.** Figure adapted from reference.[22]

(iii) HNCA: The amide nitrogen is coupled only with the Cα atom of the same residue (i th residue) and those of the preceding residue (i-1 residue). The coupling to the directly bonded Cα is stronger with greater intensity than Cα i−1 peaks from the preceding residue.

(iv) HN(CO)CA: A triple resonance experiment that correlates the resonances of the amide in a residue with the Cα of the preceding residue (i−1 residue). The HNCA and HN(CO)CA are less sensitive than HNCO but more sensitive than HN(CA)CO.

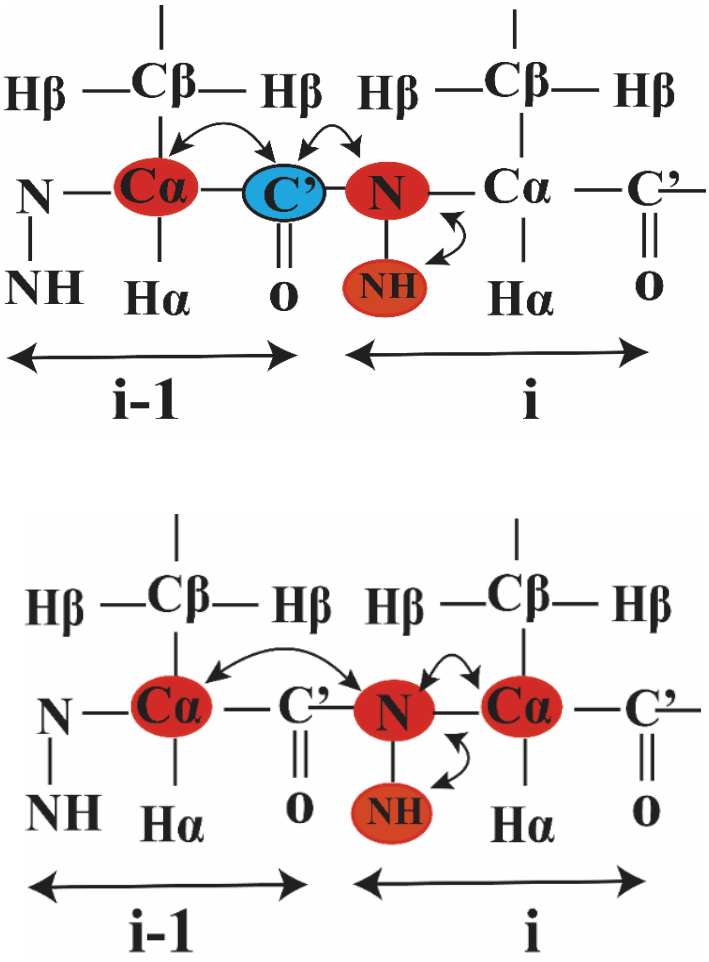


Figure 1.10: Magnetization transfer in HNCA and NH(CO)CA experiments. Figure adapted from reference.[23]

(v) HNCO: This triple resonance experiment provides the connectivities between i th residue amide with the preceding (i-1) residue carbonyl carbon only. It is the most sensitive experiment. One peak per residue is observed in the spectrum and has minimal signal overlap.

(vi) HN(CA)CO: In this experiment, in which the amide resonance of ith residue is correlated with the carbonyl carbon of the same residue, as well as the preceding residue. The coupling between N_i and CO_i is stronger and has a more intense peak than N_i and CO_{i-1}.

This experiment is the least sensitive because preceding CO signals are often missing. It is difficult to link the spin system with others. Low sensitivity is due to the fast relaxation of transverse Cα magnetization, which is further increased when the protein size increases.24 Sometimes increasing the number of scans can improve the signal strength to some extent. The chemical shift for the carbonyl is much less amino acid-specific and assigning the correlation to specific amino acid is very difficult.

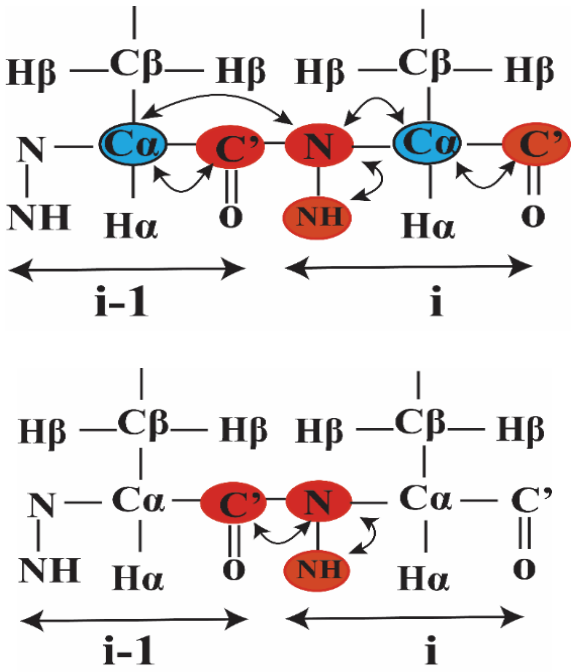


Figure 1.11: Magnetization transfer in NH(CA)CO and HNCO experiments. Figure adapted from reference.[23]

d. Side-chain Resonance Assignment

Assigning side-chain resonances depends on a set of NMR experiments that record through-bond interactions with side-chain carbons and protons for each residue. The complete side-chain resonance assignments can be obtained from 3D HCCH-TOCSY (correlating the proton side-chain and carbon side chain resonances of own residue). This is one of the most useful experiments which is needed for

the complete side chain assignment, where all side-chain carbons and protons are observed simultaneously. The 3D HCCCONH-TOCSY (correlating Hi–Ni from the amide group with protons Hsc i−1 from the side-chain of the preceding residue), and 3D ^{15}N-TOCSY-HSQC (The magnetization is transferred between all the aliphatic protons of the side chain of the same residue). Similarly, the carbon side chains are obtained from CC(CO)NH-TOCSY(correlating Hi–Ni from the amide group with the carbon side chain from the preceding residue, i-1). In this experiment, the signal strengths are reduced due to short T_2 times and cannot be improved only by increasing the number of scans. Furthermore, the time needed for a complete magnetization transfer increases with increasing side chain length.[24]

Table 1.2: List of TOCSY experiments and the information obtained

S.N	TOCSY Experiments	Information obtained
1	3D ^{15}N-TOCSY-HSQC	All the side chain protons from the i^{th} residue (Hα, Hβ2, Hβ3, and Hγ, and Hδ)
2	3D HCCH-TOCSY	All the side chain protons and carbons from the i^{th} residue (Hα, Hβ2, Hβ3, Hγ, and Hδ and Cα, Cβ, Cγ, and Cδ)
3	3D HCCCONH-TOCSY	All the side chain protons from the i-1^{th} residue (Hα, Hβ2, Hβ3, Hγ, and Hδ)
4	3D CC(CO)NH -TOCSY	All the side chain carbons from the i-1^{th} residue (Cα, Cβ, Cγ, and Cδ)

e. NOE Peak Resonance Assignment

In the nuclear Overhauser effect (NOE), nuclear spin polarization transfers from active nuclear spin another by the process called cross-relaxation.[25] The NOE intensity depends on the distances between the two spin separated by distance r, and given by equation 9.

$$nOe = (1/r)6 \quad (9)$$

where, r is the distance between two nuclei, where $r \leq 5$ Å. NOE is only detected between two spins when they are within 5 Å in the space. The intensity is proportional to r^{-6}. NOESY experiments depend on dipolar couplings. The correlations between the two spins are observed when they are approximately 5Å apart. Although two amino acids are far apart, if they are spatially close in the tertiary structure, they can be detected in NOESY. The peak intensities depend on the distance and are indirectly proportional to r^6 between two coupled spins.[18] The mixing time is the most important parameter for NOE experiments. Furthermore, the collection of NOE data over a range of mixing times allows NOE buildup curves to be constructed and analyzed to obtain a more accurate determination of inter-proton distances.

The NOEs are important in the assignment of NMR resonances, to calculate the structures of biological molecules. The ^{15}N HSQC-NOESY-HSQC experiment correlates all protons within a range of approximately 5 Å to the backbone amides. The 3D ^{13}C-HSQC-NOESY provides the correlation of protons, and the magnetization is transferred to the directly coupled ^{13}C, which is then detected in the third frequency domain. There are few drawbacks, including less sensitivity than ^{15}N HSQC-NOESY, and there are no NOE correlated amide signals. Therefore, ^{13}C NOESY is a supplementary experiment but does not replace the ^{15}N NOESY-HSQC experiment.

1.1.6. Advantage and Drawback of Protein NMR

1). Advantages

1. The development of the heteronuclear 3D experiments reduces chemical shift overlapping drastically, and chemical shift dispersion of ^{15}N (range 100-135ppm) and ^{13}C (10-75 ppm) for aliphatic and (110-140 ppm) for aromatic and (165-185 ppm) for carbonyl are much larger than of ^{1}H.
2. NMR has diverse applications for example chemical shift gives secondary structural elements; long-range NOEs provide tertiary structure.
3. HSQC spectrum is used to assess relaxation studies, denaturation studies, diffusion studies, measuring the HN exchange, and monitoring chemical shift change upon addition of ligands. These studies help in understanding dynamics and function.
4. Other structural determination techniques cannot be used for a variety of applications.
5. NMR protein spectroscopy is performed in a solution that resembles the physiological conditions which are more biologically relevant.

2). Disadvantages

1. Low inherent sensitivity
2. Need high protein concentrations
3. Long acquisition time
4. Natural abundance is not suitable for analysis. It is necessary to label ^{15}N and ^{13}C; it is more time-intensive and expensive.
5. Protein NMR spectra are difficult to interpret due to large signals.
6. Signal broadening (for large proteins) leads to peak overlapping.

h. Sensitivity & Resolution of Protein NMR experiments

Resolution and signal sensitivity are critical for protein structure determinations by solution NMR spectroscopy. A sufficient number of data points is necessary to obtain enough peak resolution. The sensitivity of NMR can be enhanced by increasing the number of scans to achieve a better signal-noise ratio. Increasing the concentration of the protein also increases sensitivity. NMR sensitivity signal to noise ratio (S/N) is directly proportional to the signal peak height. The resolution of the spectra depends on several factors: (a) field strength of the NMR spectrometer,[26] (b) the number of acquired data points,[27] and (c) line width of the signals, which are indirectly proportional to T_2 relaxations (T_2 decreases along with the increase in molecular masses leading to a broad signal).[28]

1.1.5. Study of Protein-ligand Interaction by NMR

The use of NMR to detect ligands with an affinity for targets can be performed in several ways, including chemical shift perturbation, differential line-broadening, transferred NOE, NOE filtering, and diffusion-based method.[29,30]

a. Chemical Exchange

Chemical exchange refers to a dynamic process of the protein. There are a variety of dynamic processes: (a) exchange between ligand-free and ligand-bound, (b) monomer to dimer, (c) protonated to deprotonated, and (d) conformational change from A to B. To interpret the experimental data two-state model is used. In NMR, two-states have resonances frequencies v_A and v_B and chemical shift difference $\Delta v = |v_A - v_B|$. The appearance of the NMR spectrum depends on (1) the population of each state and (2) the relative values of exchange rates ($k_{ex}=k_A+k_B$), and 3) the chemical shift difference Δv.[31] The K_{ex} quantifies the average number of stochastic exchange events per unit of time and is expressed in sec^{-1}.

There are three distinct chemical exchange regimes:

1. **Slow ($k_{ex} << |\Delta\nu|$):** In this regime, signals from both free and bound states are observed, reflecting their distinct chemical shifts, intensities, and linewidths. There is no significant interconversion during the NMR experiments frequency detection period. In this limit, the intensity of each peak directly reports on the population of that species.
2. **Intermediate ($k_{ex} \approx |\Delta\nu|$):** In intermediate exchange, one signal is observed at a chemical shift between δ_A and δ_B. Notably, the peak's linewidth is broadened due to interference from the interconversion during the detection period. Anomalous peak broadening (i.e., $R_{ex} > 0$) is a hallmark of the intermediate exchange regime's dynamics.
3. **Fast ($k_{ex} >> |\Delta\nu|$):** One signal is observed because there is rapid interconversion and hence signal averaging during the detection period of the NMR experiment (Figure 1.12).

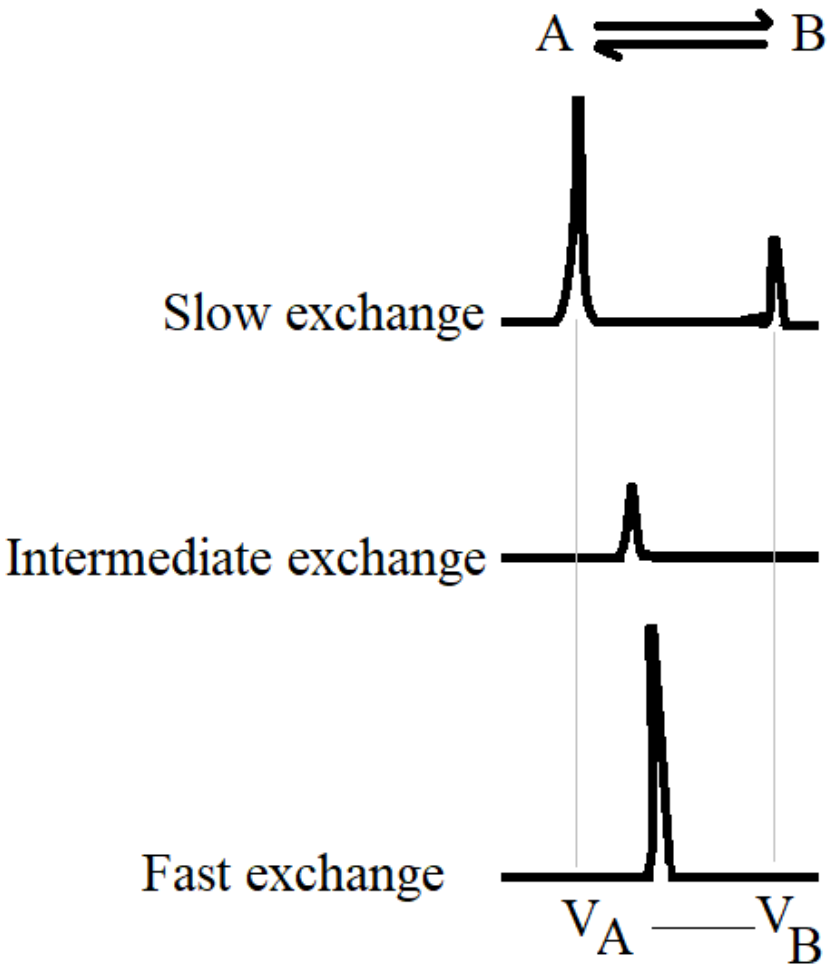


Figure 1.12: Chemical exchange processes (including protein dynamics). Figure adapted from reference(32).[32]

b. NMR Line Shape Analysis for Protein-ligand Interactions

The spectra (i.e., the exchange regime: slow, intermediate, or fast) are strongly affected by the ligand-binding affinity. The tighter-binding complex yields a longer-lived bound state and slower exchange between the free and bound states during signal detection. Intermediate exchange results from intermediate binding due to significant interconversion between the free and bound states during the NMR experiment's detection period. Fast exchange results from weak binding because there is rapid interconversion (hence averaging the signals) during the experiment's detection period (Figure 1.13).

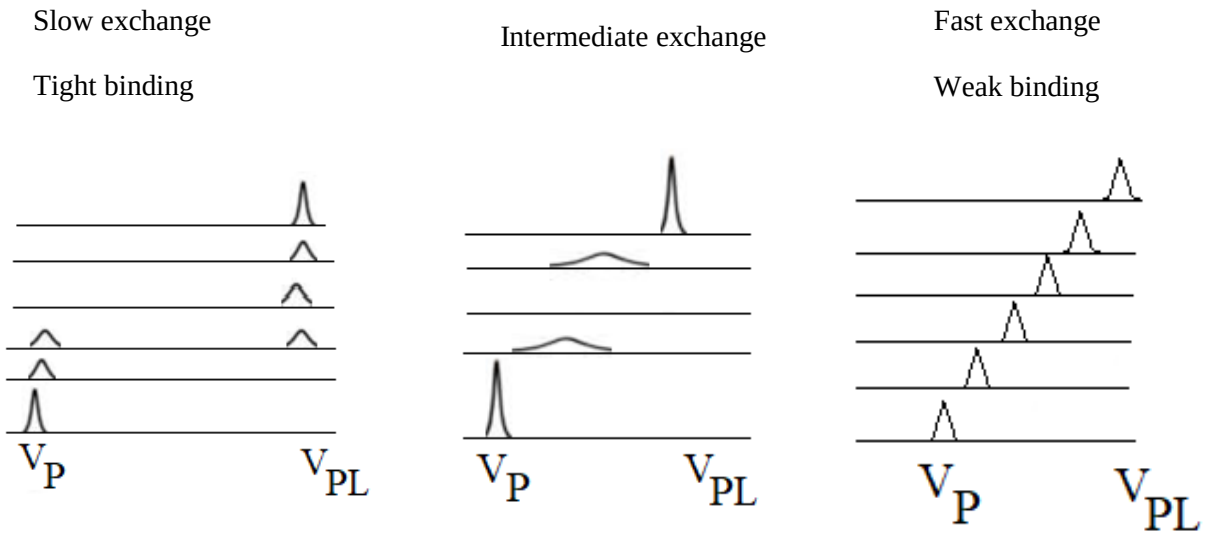
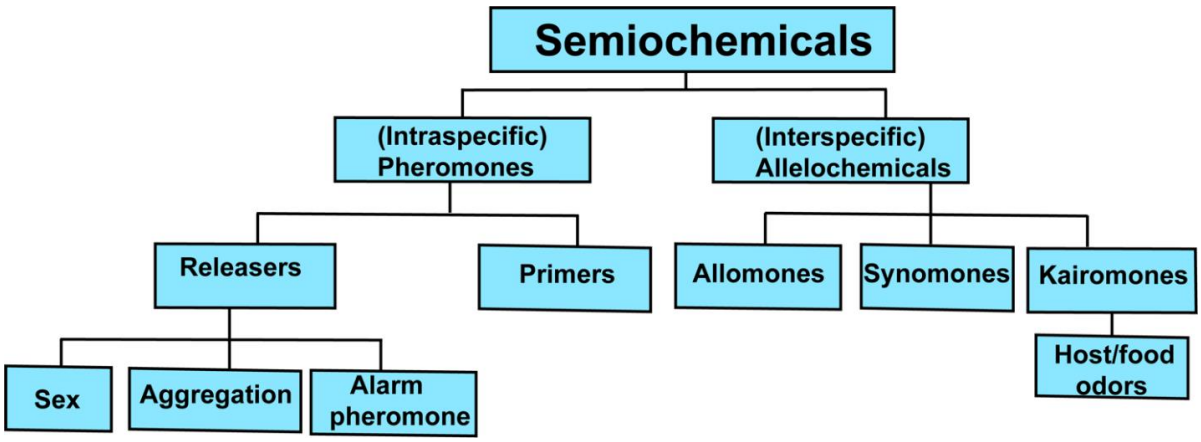


Figure 1.13: Schematic NMR line shape analysis can be used to study protein-ligand interactions, P + L ↔ PL, by acquiring multiple NMR spectra along with a {P}/{L} titration coordinate. The exchange regime (slow, intermediate, or fast) is strongly affected by ligand binding affinity. Figure 1.13 is adapted from ref 31.[31]

1.2. INSECT OLFACTION

Animal olfaction has an immense impact on their survival. Chemoreception controls fundamental behaviors in animals such as searching for food, finding mates, and avoiding enemies. The olfactory system in the insect is an excellent model in neuroscience. Insects use a chemical compound called a pheromone that enables members of the same species to communicate with each other.[33] Male moth antennae are housed by hair-like structures called sensilla which are involved in the detection of chemical signals. Pheromones are one of the chemicals released by female moths which attract and trigger male moths for mating. The term pheromone is derived from the Greek "pherein" (to carry or transfer) and "hormōn" (to excite, stimulate) and was proposed by German biochemist Peter Karlson and Swiss biologist Martin Lüscher in 1959.[33] Bombykol was the first pheromone identified in the silkmoth *Bombyx mori*. Insects use hydrophobic molecules, which initiate the signaling on intra- and inter-species. These volatile molecules are known as semiochemicals. They trigger a natural response in members within the same species. This subclass can be further divided into sex, aggregation, and alarm pheromones shown in Figure 1.14.



.

Figure 1.14: Classification of the semiochemicals. The figure adapted from ref (34).[34]

The sex pheromones are chemical cues which trigger a behavioral response for mating in members of the same species. The male moth has a unique ability to detect and respond to female sex pheromone over a long distance, for example, up to 11 km which was reported in an emperor moth, *Pavonia pavonia*.[35] Female moths can produce small quantities of sex pheromones in the nanogram to picogram range.[36] There are two types of pheromones. One type is called a releaser pheromone, which initiates stimulus immediately, causing the behavioral response upon reception. The second type is called a primer pheromone, which causes long-lasting physiological changes resulting in the behavioral respone[37]. A chemical pheromone is released by female moths into the air and attracts males over long distances. Males perceive female sex substances through their antennae.[35] A chemical substance called a pheromone is released by female moths into the air and attracts males over long distances.[35] The pheromone molecules are generally 12-18 carbons (Figure 1.15) and are unsaturated with diverse functional groups. These airborne odorants molecules are hydrophobic. The aqueous layer is impenetrable for hydrophobic compounds. Thus, transport through this barrier is assisted by odorant-binding proteins/pheromone binding proteins (PBPs) in the sensilliar lymph.

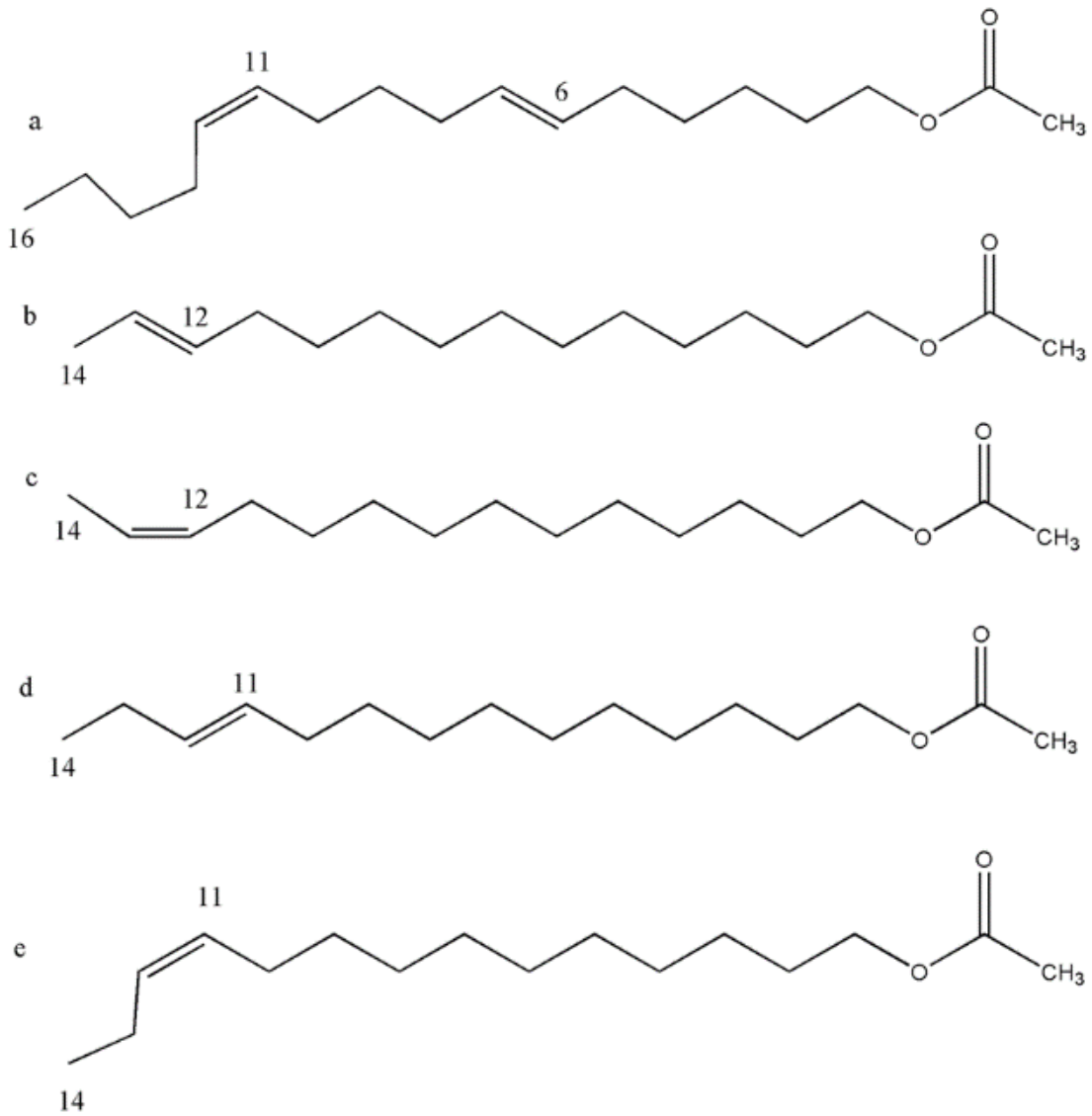


Figure 1.15: Chemical structures of the pheromone molecules:(a) (6*E*,11*Z*)-hexadecenyl-6,11-dienyl-1- acetate of *Antheraea polyphemus* (b) and (c) (*E*) -12 tetradecenyl acetate and (*Z*)-12-tetradecenyl acetate of *Ostrinia furnacalis* (d) and (e) (*E*)-11 tetradecenyl acetate and (*Z*)-11-tetradecenyl of *Ostrinia nubilalis.*

The insect olfaction system serves as an excellent model for understanding olfactory behavior in animals. A pheromone binding protein (PBP) was first identified on the antenna of the male silkmoth *Antheraea Polyphemus* at an estimated concentration of 10 mM.[38] The PBP is mostly found in male

antennae, but a lower concentration is found in the female antennae, for some insects like *Manduca sexta*[39,40], *Spodoptera exigua*[41], and *Cydia pomonella*.[42] The antenna of the moth has a specialized hair-like unit called sensillae. These olfactory sensilla consist of one or several olfactory receptor neurons (ORNs). The major role of ORN is to modulate ion potentials across their plasma membrane and thus participate in the transduction of chemical signals into electrical signals. The outer cuticle layer consists of numerous pore tubules which are approximately 10 to 20 nm in diameter.[43] There is also a cavity filled with the sensilliar lymph (Figure 1.16 and 1.17). The aqueous layer protects the dendrite of the sensory neuron. A sensory neuron is directly connected to the central nervous system.[44] When volatile pheromone molecules reach the cuticle pore, these pheromones are transported by PBPs across the sensillum lymph to the ORs. The pheromone interacts with the pheromone-binding protein. These pheromone molecules are carried and released to the ORs, which initiate signal transduction.

Figure 1. 16: Male moth antennae. Figure adapted from reference 45.[45]

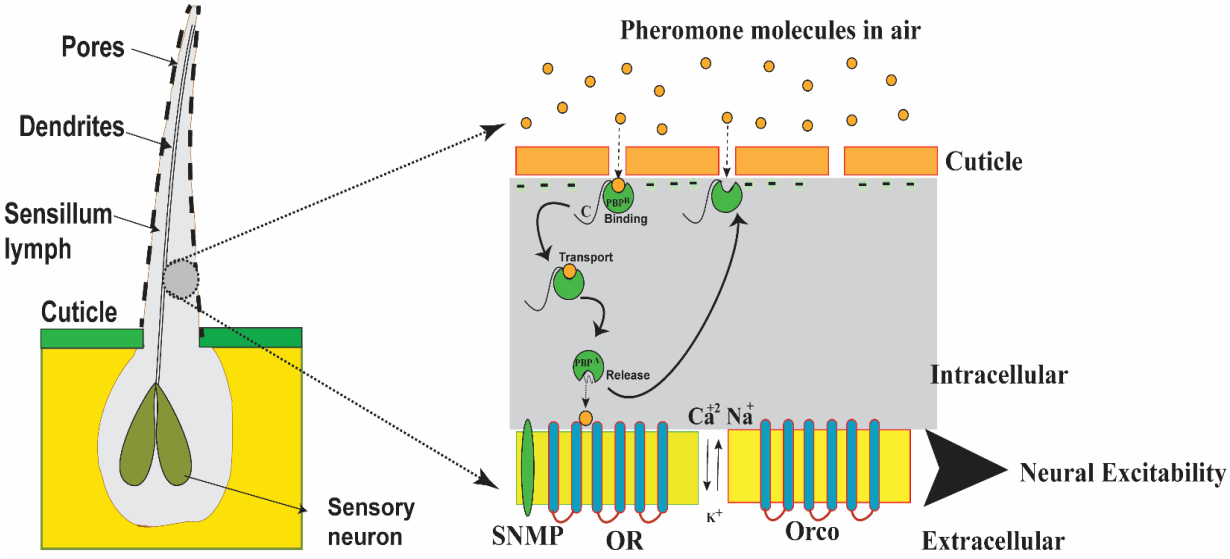


Figure 1. 17: Schematic representation of insect olfaction.

1.2.1. Odorant binding proteins (OBPs)

Odorant binding proteins (OBPs) are the class of olfactory proteins that are responsible for binding and transporting odorants molecules to the receptors. OBPs are small, water-soluble, and extracellular proteins located in the fluid surrounding the sensory dendrite and bind the odorants.[46] OBPs of vertebrates belong to the large superfamily of carrier protein called lipocalins.[47] These proteins contain around 150 amino acids and have mainly a beta-sheet with a large cup-shaped cavity within the beta-barrel structure (Figure 1.18). They usually exist in a homodimer form with the binding pocket inside the barrel that can accommodate ligands. In insects, OBPs are divided into three subfamilies: pheromone binding protein(PBPs), general odorant-binding proteins (GOBPs), and antennal-binding specific protein (ASPx).

Figure 1.18: Odorant binding protein from vertebrates having a beta-barrel structure with a large cup-shaped cavity. Figure adapted from ref 48.[48]

Odorant binding proteins (OBPs) are one class of olfactory proteins responsible for capturing and transporting odorants molecules to the receptors. General odorant-binding proteins bind to a broad range of odorant molecules, while PBPs bind pheromones in a species-specific manner. Pheromone-binding proteins are highly water-soluble, small, extracellular proteins of around 130–150 amino acids, with molecular weights of 10–20 kDa. These acidic proteins contain six highly conserved cysteine residues that form three disulfide bonds.[49] PBPs generally contain six or seven alpha helices that form a hydrophobic binding pocket. The PBPs have various crucial physiological functions: (a) dynamic roles in ligand selectivity, (b) ligand scavenging, (c) responsible for the protection of olfactory receptors(ORs) from saturation[50], and d) ligand desorption from the cuticular wax layer of olfactory pores to the lymph[51] and (e) specific recognition of hydrophobic pheromone molecules and their

transport across the aqueous sensilliar lymph to the olfactory receptors neurons (ORN)[51]. OBPs also protect odorant molecules from degradation by odorant-degrading enzymes during their transport and transfer to the ORs.[52]

Classification of the Odorant-binding Protein

OBPs are classified as Classic OBPs, Plus-C OBPs, Minus-C OBPs, and Atypical OBPs, as shown in Figure 1.19.[53,54] The significant differences within them are the number of cysteine residues and their positions.

i. Classic OBPs are the same as PBPs, general odorant-binding proteins (GOPs) and antennal-binding specific protein (ASPx) consist of 6 conserved cysteine residues. Although classic OBPs have six conserved cysteine residues, they are further classified according to chain length and C-terminus length.[55] They are:

   (1) Long-chain OBPs with 140 amino acids (OBPs of *Bombyx mori* and *Antheraea. Polyphemus*)

   (2) Medium-chain OBPs with around 120 amino acids (OBPs from *Anopheles Gambiae* and *Apis Mellifera*), and

   (3) Short-chain OBPs around 100 amino acids (for example OBPs from *Cockroach Leucophaea maderae*)[55]

ii. Plus-C OBPs have two additional cysteine residues along with a conserved proline.[53,54]

iii. Minus-C OBPs have less than six cysteine residues.[53]

iv. A typical OBP has six cysteine residues like PBPs and an additional cysteine in the C-terminus.[56]

A new subclass of OBPs, C8 OBP, has been proposed which is based on the presence of eight cysteine residues that form 4 disulfide bonds, and a C-terminus that is longer than the classic OBPs.[57] An example of this class is *Anopheles gambiae* OBP7 (AgamOBP7).

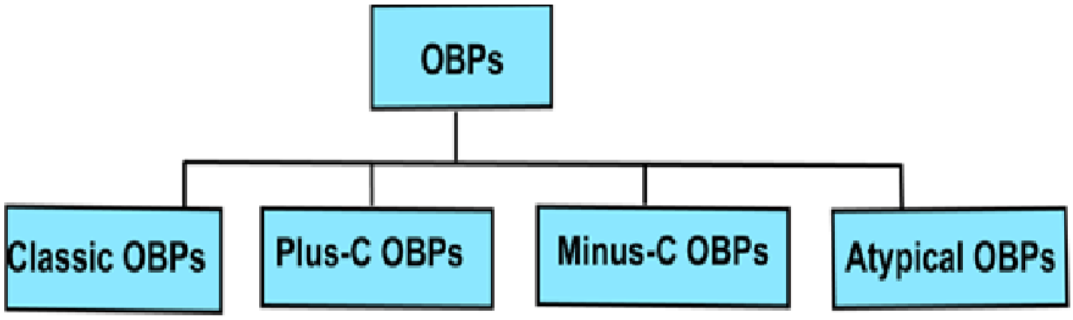


Figure 1.19: Classification of the odorant-binding protein. It is adapted from reference 55.[55]

1.2.3. Insect Pheromone Binding Proteins

1. *Bombyx mori* Pheromone-binding Protein (BmorPBP)

The PBP from the *Bombyx mori* is one of the well-characterized proteins. Two conformations of the protein were observed at pH 6.2 during the DEAE purification.[58] The BmorPBP undergoes the pH-dependent conformational change between pH 6.5 and 4.9. In between pH 5 and 6, a mixture of conformations is in slow exchange on the NMR time scale with the transition taking place at pH 5.4, which was verified by the presence of a double set of peaks in the $\{^{1}H\ ^{15}N\}$ COSY spectra.[59] The BmorPBP exists in the dimer form at pH 6.0. When the pH is lowered at 4.5, the protein appears in the

monomeric form.[60] The BmorPBP was co-crystallized with the species-specific pheromone bombykol. A three-dimensional structure (PDB code: 1DQE) consists of a dimer with the C-terminus region unstructured and extended outside binding pocket[61] shown in Figure 1.20. The pheromone was found to occupy the hydrophobic pocket. The hydroxyl group of bombykol forms a hydrogen bond with the side chain of Ser56.[61]

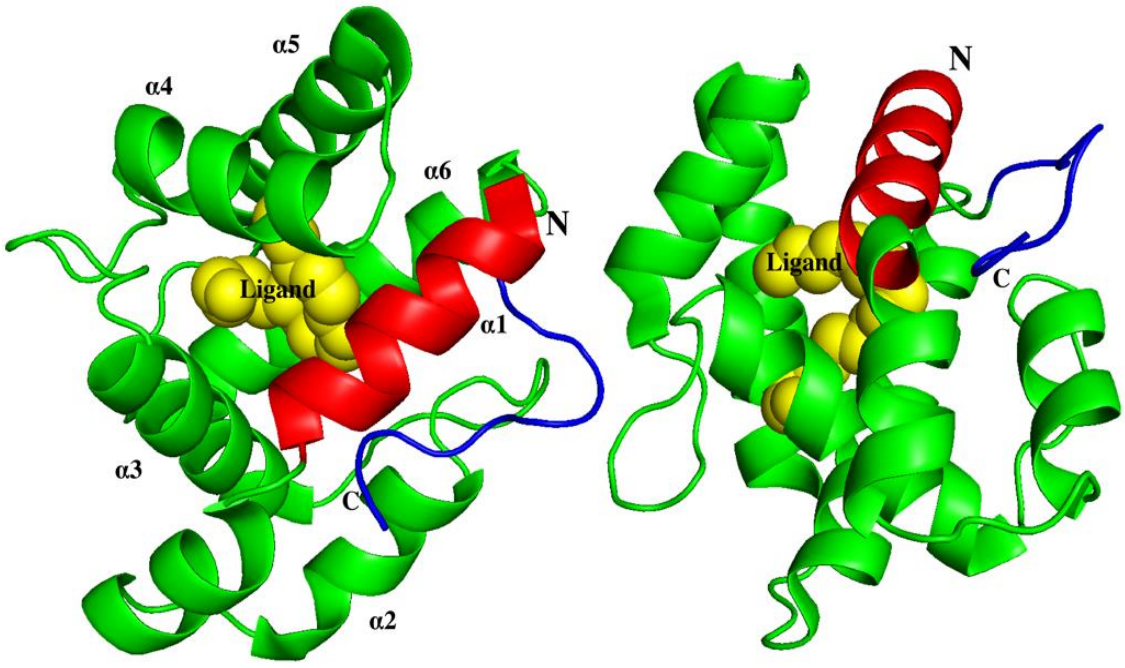


Figure 1.20: X-ray crystal structure of homodimer BmorPBP bound with bombykol (PDB code: 1DQE).[61] The N-terminus is shown in red color, the C-terminus is shown in blue color, and the ligand is shown in yellow.

The NMR structure of BmorPBP at acidic pH (pH 4.5) (PDB ID: 1GMO) consists of seven alpha-helices[62] (Figure 1.21). Interestingly, the C-terminus forms a regular 7th helix and occupies a binding

pocket. The N-terminal region is unstructured.[62] BmorPBP does not bind bombykol at pH 4.5 because the helix occupies the binding cavity.

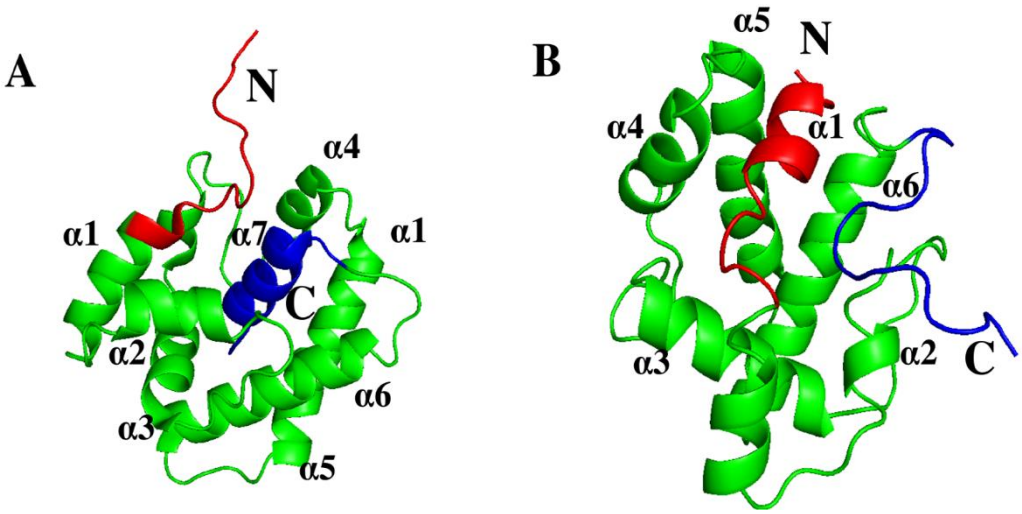


Figure 1.21: NMR structure of BmorPBP : A) pH 4.5 (PDB code : 1GMO)[62], B) Unliganded BmorPBP at pH 6.5 (PDB: 1LS8).[63] The N-termini are colored red, and the C-termini are colored blue. At pH 4.5, the C-terminal peptide forms the 7th α-helix and inserts into the hydrophobic pocket while the N-terminal peptide unwinds. At pH 6.5, the C-terminal peptide is unstructured and remains outside the pocket and exposed to the solvent.

Ligand binding and releasing mechanisms of BmorPBP were investigated by Lee et al. in 2002. They determined the NMR structure of unliganded BmorPBP (PDB ID:1LS8)[63] at pH 6.5 (Figure 1.21 B). The NMR structure of the unliganded form is identical with one of the monomers of the BmorPBP-bombykol complex obtained from crystal structure[61], with an average backbone root mean square deviation (rmsd) value of 1 Å. The NMR structure shows two flexible regions, the loop between α2 and α3, and the unstructured C-terminal segment extended into the solvent.[63]

The crystal structure of apo-BmorPBP at pH 7.5 was solved by Lautenschlager et al. in 2005.[64] The structure was similar to the NMR structure obtained at acidic pH. The C-terminal tail of apo-BmorPBP forms an ordered helix occupying the binding pocket (Figure 1.22).[64] Based on all of these structures, it reveals that BmorPBP exists in PBP^B or open conformation at pH above 6.0, however, in close or PBP^A or ligand-free conformation at pH below 5.0. Both the protein either at apo form (ligand-free) or pH 4.5, the C-terminus helix occupies the pocket which was shown by Katre et al. 2009.[65] Furthermore, the structures of BmorPBP indicate that the C-terminus of the BmorPBP undergoes a pH-dependent coil to helix transition as shown by Mazumder et al. 2019.[66] The coil-helix transition is important to drive the ligand-bound (PBP^B) conformation to ligand-free (PBP^A) conformation, associated with the ligand release. Three charged residues at the C-terminus play a vital role in the coil-helix transition.[66]

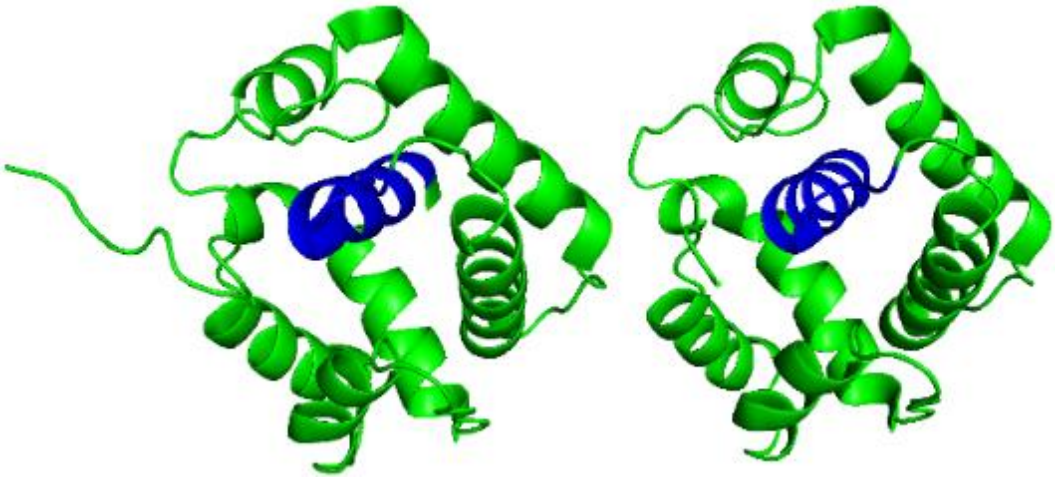


Figure 1.22: X-ray structure (no ligand) Apo BmorPBP at pH 7.5, (PDB code : 2FJY).[64] The C-terminus is showing in blue color, the C-terminus is a helix and inserted in the pocket.

Kinetic studies show that BmorPBP a has high binding affinity (Kd = 105 nM) at pH 7.0, and low affinity (Kd = 1,600 nM) at pH 5.0.[67] The C-terminus segment was truncated (BmorPBP ΔP129-V142) to understand the role of the C-terminus in ligand binding. It was shown that, at pH 5.0, the pheromone binds with the same affinity as the native protein at pH 7.0. This suggests that at low pH, the C-terminus plays an essential role in preventing ligand binding. Furthermore, mutating the tryptophan residue at

positions 37 and 127 did not affect the binding affinity to bombykol.[67] The selectivity of males towards their respective sex pheromones is extraordinarily sensitive. The biological activity of bombykol decreases by a billion-fold even if there is a change in the stereochemistry of one double bond.[68] The PBP-receptor at the dendritic membrane discriminates between ligands even if there is a subtle difference in the ligands.[69] To find the selectivity of BmorPBP, Lautenschlager et al. 2007, determined the X-ray crystal structure complexed with a non-pheromone ligand like iodohexadecance (PDB code: 2P71) and with bell pepper odorant (PDB code: 2P70)[69] (Figure 1.23). It was found that ligands with very different geometries, from straight-chain carbon to aromatic molecules, can also fit into the cavity of the BmorPBP.[69] Their study suggests that BmorPBP can bind a non-pheromone ligand, but the affinity is minimal. Ligands with such low affinities may be dropped from the complex and are inactivated by aggressive odorant-degrading enzymes.[70]

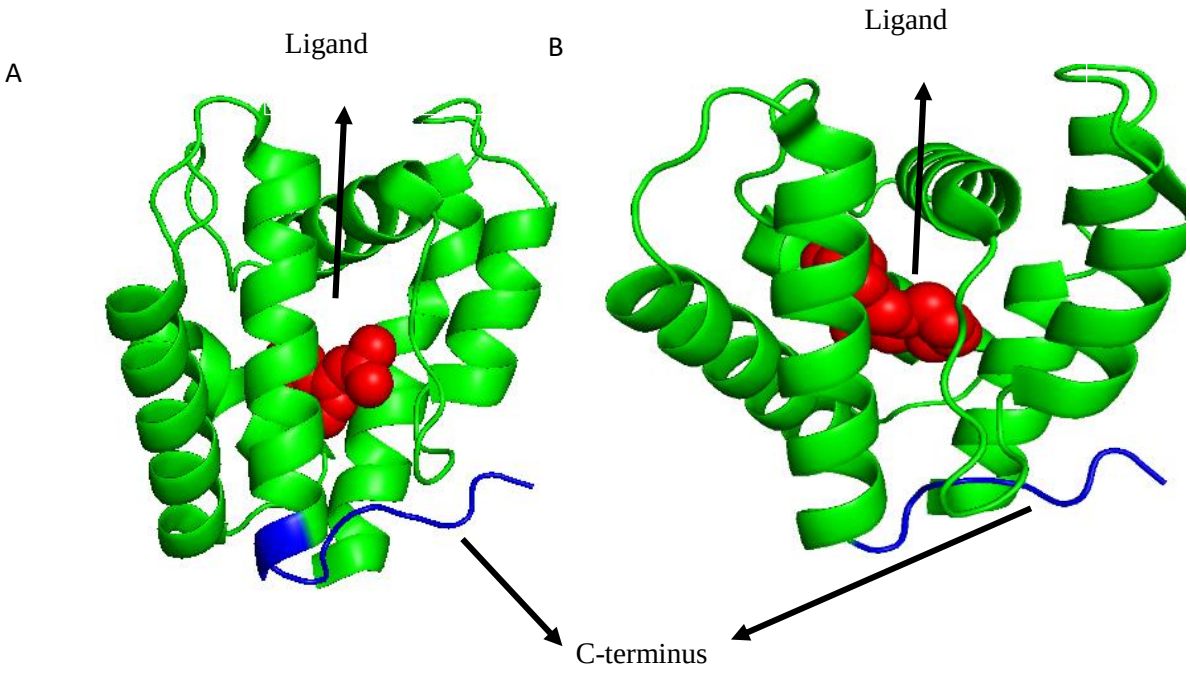


Figure 1.23: X-ray crystal BmorPBP complexed with A) Iodohexadecance (PDB code: 2P71) and with B) bell pepper odorant (PDB code: 2P70).[69]

The NMR structure of the truncated BmorPBP(1-128) at pH 6.5 resembles closely to BmorPBP^{B} form[71], as shown in Figure 1.24. At pH 4.5, protein still exists in a bound "B" form. These results firmly explain that BmorPBP(1–128) is unable to eject ligands at low pH, clearly suggesting the role of the C-terminus in ligand releasing function.

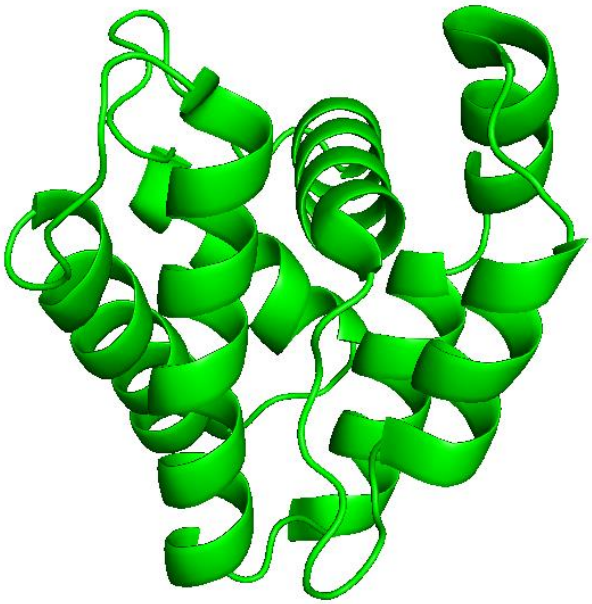


Figure 1.24: NMR structure of BmorPBP(1-128) at 6.5 (PDB code: 1XFR).[71]

Based on the structure and function studies on BmorPBP, both pH- and ligand-dependent mechanism has been proposed.[61,63,64,72,73] At physiological pH, bombykol binds with a higher affinity to BmorPBP in comparison to the other ligands that were studied[74]. During transport, the pheromone forms a complex with $BmorPBP^{B}$ and is protected from degradation by degrading enzymes present in the sensillum lymph. When BmorPBP–pheromone complex diffuses toward the membrane-receptor, their stability is reduced by the acidic environment near the membrane.[58,75] It has been reported that the local pH is reduced by up to 2 pH units due to the negatively charged dendritic membrane[76,77] or lipid head groups.[74,75] Bombykol is released when the protein undergoes a pH-induced conformational change at the receptor. The ligands that have a lower affinity to the protein are released sooner before reaching the membrane and are subjected to degradation.[74]

2. *Antheraea polyphemus* Pheromone-binding Protein 1 (ApolPBP1)

Antheraea polyphemus pheromone binding protein1 (ApolPBP1) was the first PBP identified on male antennae of the wild silkmoth *Antheraea polyphemus*.[38] Three different pheromone-binding proteins were identified in this species. ApolPBP1 has over 50% sequence identity with PBPs from other moth species but differs in substrate specificity. ApolPBP1 binds (6*E*,11*Z*)-hexadeca-6,11-dienyl-1-acetate pheromone only at pH above 6.0. Similar to BmorPBP, it undergoes a pH-dependent conformational transition at an acidic pH.[78,79] NMR structure of ApolPBP1 at pH 6.3 has 9 helices packed in a globular structure (residues 1–125), enclosing a large hydrophobic cavity. The C-terminus of this protein composed of residues Pro126–Val142 is unstructured and extends to the solvent.[80] Residue Asn53 plays a key role in the recognition of acetate pheromones. The overall structure of ApolPBP1 is similar to BmorPBP.[63]

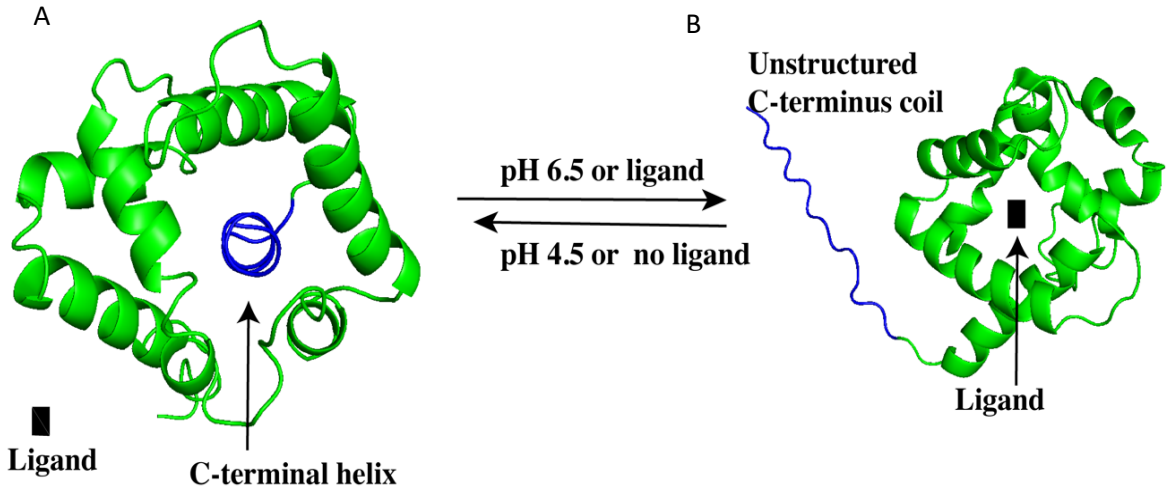


Figure 1.25 : NMR structure of ApolPBP1: A) pH 4.5, PDB code: 2JPO[81], B) pH 6.5 PDB code: 1QWV 6.5.[80] There is a conformation transition due to pH change and the presence and absence of ligand.

The NMR structure of ApolPBP1 at pH 5.2[82] shows the flexibility of the N-terminus and loop (Leu33-Asp39) segments resulting in conformational exchange in intermediate time scale, leading to the peak broadening (Figure 1.26). Two sets of resonances were observed due to the presence of two conformations, PBP^{B} (ligand-bound) and PBP^{A} (ligand-free).[82] It is clear from this report that ApolPBP1 exists as a mixture of both the conformations at pH below 6.0 but above 5.0.[82]

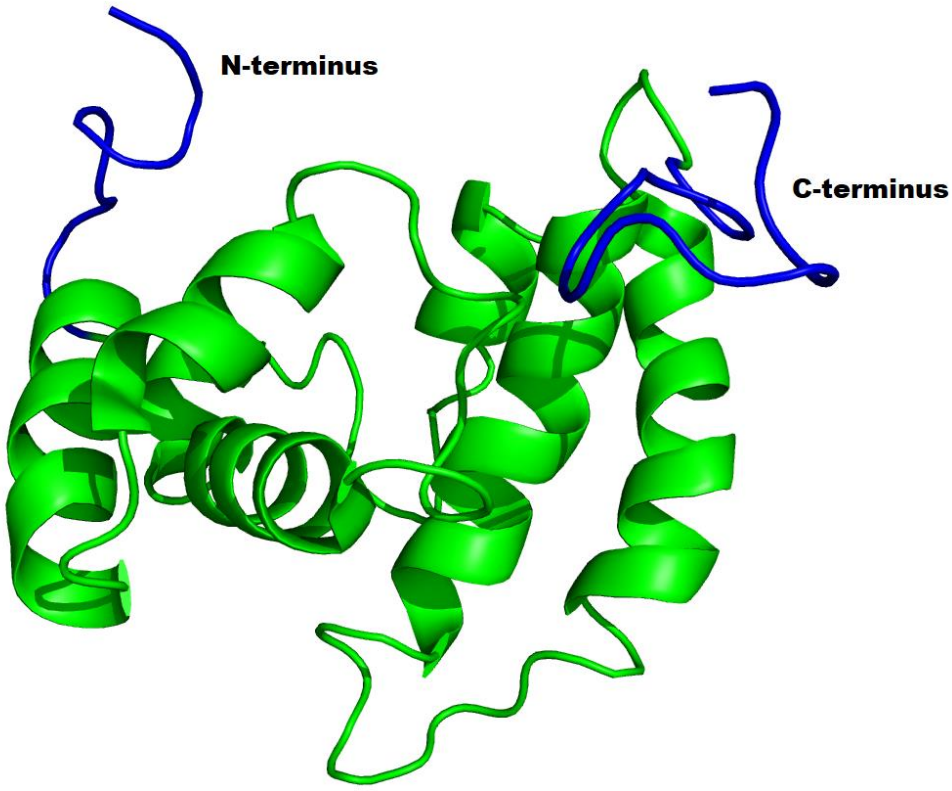


Figure 1.26: NMR structure of ApolPBP1 at pH 5.2 (PDB code: 1TWO). The figure is taken from reference 82.[82]

The NMR structure of the ApolPBP1 at pH 4.5 consists of six α-helices, arranged in a globular fold that encapsulates a central helix α7 formed by the C-terminal polypeptide segment Met131 to Val142[81].

The structure is similar (rmsd = 1.7 Å) to that of BmorPBP at 4.5 pH. Similar to BmorPBP, ApolPBP1 also undergoes a pH-dependent conformational transition between PBP^B and PBP^A. Katre et al. 2009[65], were the first to show that the recombinant ApolPBP1 picks up a hydrophobic ligand endogenous to the *E. coli* cells during protein expression. The purified protein exists in ligand-bound or PBP^B conformation unless this ligand is removed through an additional step of purification called delipidation. This lipid-bound ApolPBP1 has the open (bound) conformation at pH 6.5, a mixture of both bound (PBP^B) and free (PBP^A) conformations at pH below 6.0, and closed (ligand-free) conformation at low pH below 5.[65,82] This work explained the conformational heterogeneity observed for undelipidated ApolPBP1 at pH 5.2[82] and the undelipidated BmorPBP at pH 5.5.[59] However, the delipidated ApolPBP1 does not undergo a pH-dependent conformational switch. In other words, the delipidated protein (ligand-free form) primarily exists in closed conformation at all pH levels.[65] The pH-induced conformational switch only holds good for the ligand-bound (open) conformation.[65] The ligand-free conformation readily converts to the ligand-bound conformation upon the addition of a ligand at pH 6.0 or higher.[65]

In both ApolPBP1 and BmorPBP, the C-terminus plays a critical role in the conformational change between bound and free forms based on pH and the presence or absence of a ligand. The role of the C-terminal tail of ApolPBP1 in ligand binding and release mechanism was reported by Katre et al. in 2013.[76] The mechanism of ligand binding and release in ApolPBP1 is carried by two important molecular switches: (i) His70 and His95 situated at one end of the binding pocket and (ii) the C-terminus at the other end of the binding pocket.[65,76] To understand the role of the three C-terminal charged residues (Asp132, Glu137, and Glu 141) in the reversible coil ⇔ helix transition, further investigation was carried out through mutagenesis and biophysical characterization of the mutant proteins.[66] It was shown that Glu137 and Glu141 are critical for the reversible coil ⇔ helix transition.[66]

The ApolPBP1E137QE141Q double mutant remains in the open (PBP^B) conformation at all pH levels. Mutation of these two-terminal acidic residues together knocks out the protein switch and adversely affects both ligand binding and release functions.[66]

3. *Amyelois transitella* Pheromone-binding Protein 1 (AtraPBP1)

The NMR structure of *Amyelois transitella* pheromone-binding protein1 (AtraPBP1) at pH 4.5[83] is similar to that of ApolPBP1[81] and BmorPBP[62], where the C-terminal helix is internalized in the binding pocket[83] (Figure 1.27). The repulsion between charged His69 and His70 at pH 4.5 opens up one end of the hydrophobic cavity while the newly formed C-terminal helix enters the pocket through the other end pushing the ligand out. Thus, at pH 4.5, the C-terminal helix occupies the hydrophobic cavity in AtraPBP1 similar to ApolPBP1[81] and BmorPBP.[62] Most importantly, there is no delipidation effect at pH 4.5[83], because the protein already releases the ligand and adopts a ligand-free conformation at pH 4.5.[65] In fact, the conformation of delipidated protein at any pH level is identical to the undelipidated protein at pH 4.5.[65] The pheromone binding affinity at neutral pH is always higher than that at acidic pH for all the wild-type protein.[84] The affinity for C-terminus truncated AtraPBP1 is reported to be 100-fold more at pH 5.0 and 1.5-fold more at pH 7.0.[84]

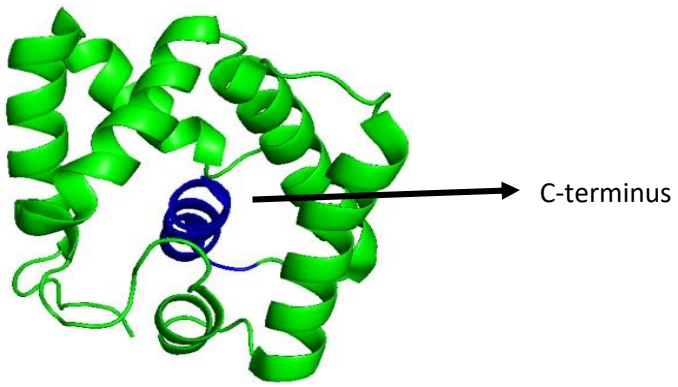


Figure 1.27: NMR structure of *Amyelois transitella* pheromone binding protein 1 (AtraPBP1) at pH 4.5 (PDB code: 2KPH).[83] The C-terminus is shown in blue color, and it is helical and inserted in the pocket.

The crystal structures of AtraPBP1, complexed with the pheromone[85] (Figure 1.28), show a similar structure with BmorPBP[72] with a 1.1 Å rmsd. The C-terminus remains in the binding pocket at acidic pH stabilized by salt bridges (His80-Glu132, His95-Glu141).[83,85] At pH 7.0, these histidine residues get deprotonated, which triggers the C-terminus leaving the binding cavity, contributing to the opening of the hydrophobic cavity.[85]

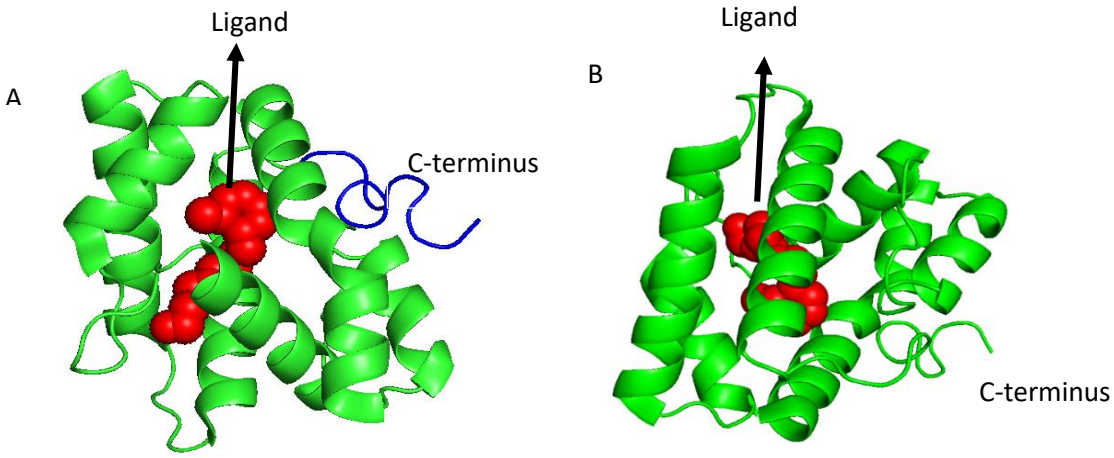


Figure 1.28: X-ray structure AtraPBP 1 at pH 6.5 A) PDB ID: 4INW, X-ray structure complex with ligand (11Z,13Z)-hexadecadienal at 6.5, C-terminus showing the blue color unstructured, and the ligand is shown in the red color. B) PDB ID: 4INX, X-ray AtraPBP1 complex with (Z, Z)-11,13 hexadecadienol ligand 6.5.[85] The C-terminus is unstructured, and the ligand is shown in red.

4. *Lymantria dispar* PBP from Gypsy Moth (LdisPBP)

Lymantria dispar, a serious pest, was introduced to North America in 1869. The major pheromone component is (7*R*, 8*S*)-7,8-epoxy 2-methyloctadecane, commonly termed (+)-disparlure.[86] There are two PBPs found in this species. Both proteins bind to the sex pheromone and are named LdisPBP1 and LdisPBP2. These two proteins are found to present at the concentration of 13.4 mM in the sensilla

lumen of the male moth antenna.[87] The pheromone binding process in both LdisPBP (PBP1 and PBP2) is reversible. LdisPBP1 has a higher affinity toward the (-) enantiomer, while LdisPBP2 has a higher affinity toward the (+) enantiomer.[88] The NMR structure of LdisPBP1 at pH 4.5, has seven helices.[89] The C-terminal residues form a helix and occupy a binding pocket and the N-terminus is disordered (PDB ID: 6UM9)[89] . At a neutral pH, it exists as a mixture of two conformations.[89] There are also both pH and ligand-induced conformational changes. The conformational transition takes place at a pH between 5.6 and 6.0. At neutral pH, it exists as a mixture of two conformations.

5. Cockroach *Leucophaea maderae* Pheromone-binding Protein (LmaPBP)

The cockroach (*Leucophaea maderae)*, pheromone binding protein (LmaPBP) is found in the female moth antennae. Males release the pheromones, and females get attracted to the male. The crystal structure of LmaPBP had shown that the internal cavity is more hydrophilic, which is the opposite of the Lepidopteran moth PBPs describes above.[63,80,85] The LmaPBP structure ends just after the sixth helix[90], which is surprisingly different from Lepidopteran moth PBPs.[63,80,85] The pheromones of this species consist of a blend of 4 compounds: (a) 3-hydroxybutan-2-one (H3B2) (b) butane-2,3-diol (c) senecioic acid (3-methylbutenoic acid), and (d) (*E*)-2-octenoic acid (Figure 1.29).[90] The absence of the amino acid stretch corresponding to the seventh helix, and the evidence of a hydrophilic binding cavity, suggests a new ligand releasing mechanism in the sense the ligand release is direct.[90]

The pH-dependent mechanism of pheromone expulsion by the 7th helix proposed for BmorPBP and ApolPBP1 is likely valid for long PBPs that bind hydrophobic (C14–C16) pheromones. The current pH-dependent mechanism is not directly applicable to PBPs of other insect classes like LmaPBP, DmelOBP, and AmelPBP as they have shorter C-terminal regions.

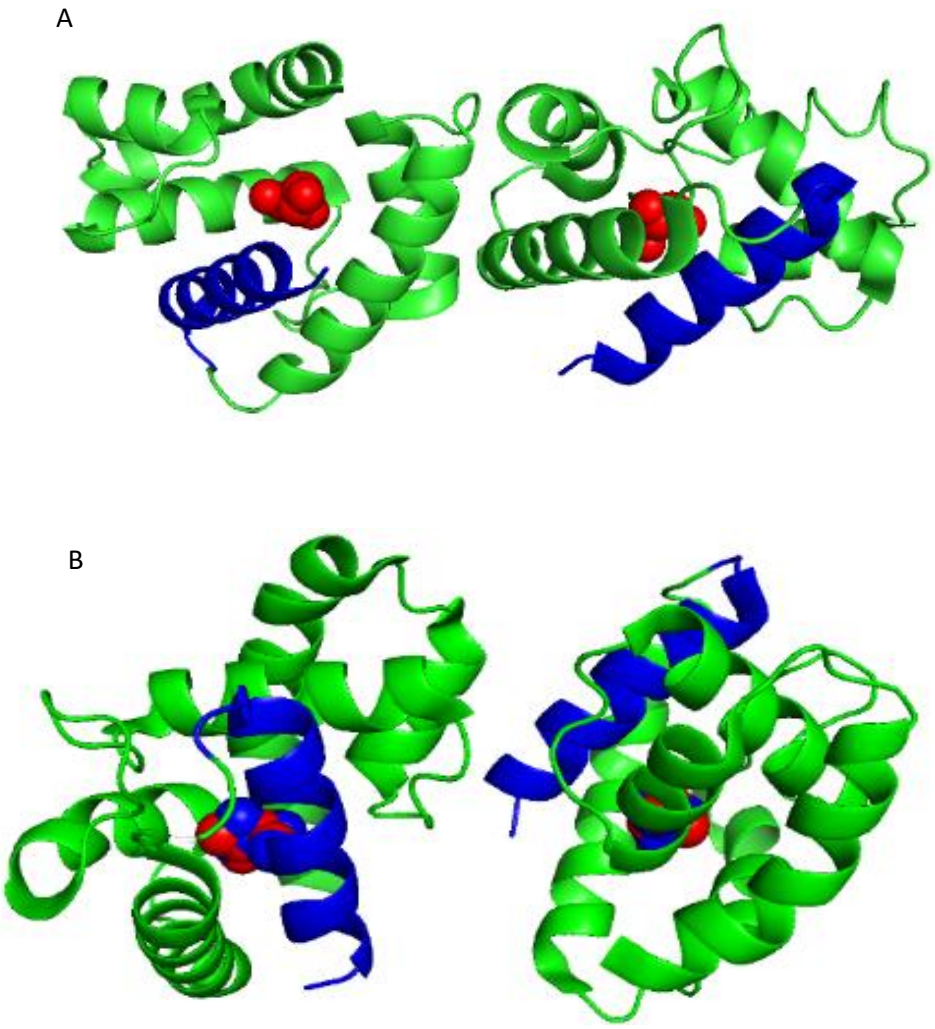


Figure 1.29: A) LmaPBP apo (no ligand) 25% glycerol, PDB code: 1ORG, B) LmaPBP+ ANS ligand (with glycerol also) two monomer, PDB code: 1OW4, C) LmaPBP+H3B2 (S/R, 3-hydroxybutan-2-one) blend 1P28.[90] The C-terminus is shown in the blue color, which is helical, and the ligand is shown in red.

6. Odorant Binding Protein from Honey Bee *Apis mellifera*

Odorant binding proteins are divergent not only across insect species but also within the same species. The honey bee, *Apis mellifera* (Amel) has two OBPs: AmelASP1, and ASP2. Although both ASP1 and ASP2 share 13-20 % sequence identity with the moth PBPs mentioned above, the general structural characteristics of OBPs of *Apis mellifera* are conserved with other PBPs. However, they also show a broad specificity for ligands. ASP2 has a unique ability to bind both hydrophilic and hydrophobic ligands.[91] There is conformational heterogeneity in the absence of ligands, and the protein is stabilized when the ligand binds.[91] *Apis mellifera* PBP1 (AmelASP1) has a distinct conformational switch than that of BmorPBP1 and ApolPBP1. AmelASP1 binds the ligand at pH 4.0 and releases it at the neutral pH of 7.0. Interestingly, AmelASP1 forms a dimer at neutral pH that is stabilized by the N-terminus[92,93], suggesting a different pheromone binding and release mechanism[91-94]. The holo-state has less conformational flexibility than the apo-state (Figure 1.30). The presence of ligands in the hydrophobic cavity stabilizes the overall structure of ASP1. There are about 20 crystal structures of ASP1 in either apo or complexes with various ligands. These results show that the medium-chain PBPs from insects may exhibit different ligand release or receptor recognition mechanisms.[92]

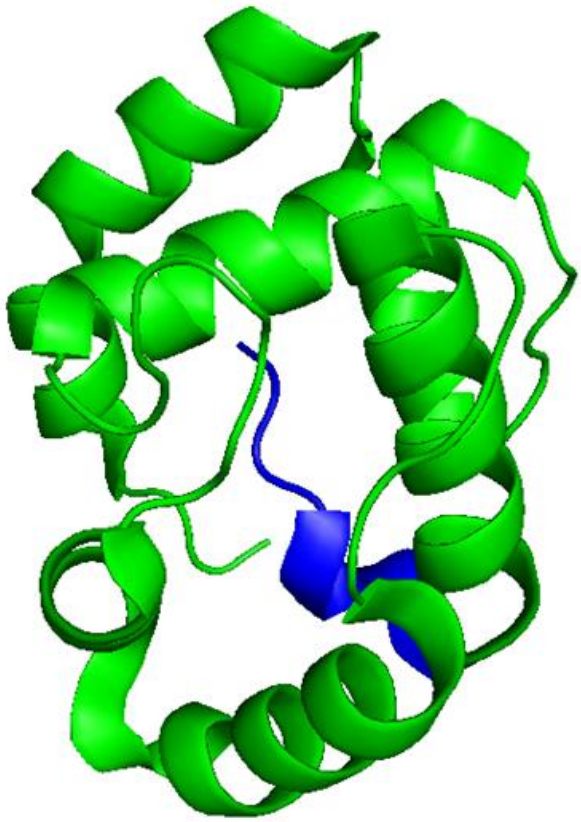


Figure 1.30: X-ray crystal structure at pH 7.0 Apo (No ligand) ASP1 (PDB code: 3CAB)[92], blue color representing C-terminus region.

7. *Bombyx Mori* General Odorant-binding Protein 2 (BmorGOBP2)

The X-ray crystal structure of the general odorant-binding protein of *Bombyx mori* (BmorGOBP2) shows a significant deviation in the C-terminus from ApolPBP1, BmorPBP1, and AtraPBP1. Although BmorGOBP2 has 12 residues on the C-terminal peptide segment (Figure 1.31). It forms an amphipathic α-helix that packs on top of the N-terminal helix.[95] It is reported that the C-terminus does not participate in ligand binding and is localized as a helix outside the binding pocket both in free and bound forms[95] (Figure 1.32). There is no conformational switch on the BmorGOBP2 structure upon ligand binding.

```
OfurPBP2    WAPDHELLLEEMMAEMKQ    144
ApolPBP1    WVPNMDLVIGEVLAEV--    142
BmorPBP     WAPSMDVAVGEILAEV--    142
AtraPBP1    WAPNMEVVVGEVLAEV--    142
BmorGOBP2   IAPEVAMVEAVIEKY---    142
```

Figure 1.31: Primary sequences comparison of the C-terminus residues of the PBPs of the moths: *Ostrinia furnacalis (*OfurPBP2*)*, *Antheraea polyphemus* (ApolPBP1), *Bombyx mori* (BmorPBP), *Amyelois transitella* (AtraPBP1), *Bombyx mori* General odorant-binding protein BmorGOBP2 (BmorGOBP2).

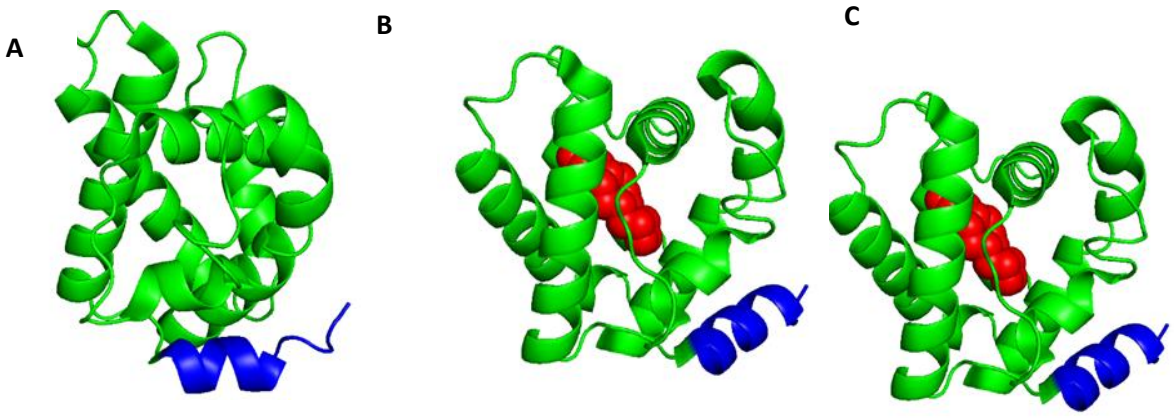


Figure 1.32: X-ray crystal structure of BmorGOBP2 A) No ligand, PDB code: 2wc5 B) bombykol, PDB code: 2wc6 C) 10E)-hexadecen-12-yn-1-ol, PDB code: 2wcm.[95] The C-terminus alpha-helical in all the structures is shown in the blue color. The ligands are shown in red color.

8. Mosquito Odorant-binding Protein

Anopheles gambiae (A. gam) mosquito is a vector for causing malaria parasites. The odorant-binding proteins (OBPs) of this mosquito participate in odorant recognition. The *Anopheles gambiae* odorant-binding protein (AgamOBP07) has 8 cysteines forming 4 disulfide bridges. The 7th helix is located at the surface of the protein locked by the fourth disulfide bridge. As mentioned previously, ApolPBP1, BmorPBP, and AtraPBP1 with a long C-terminal tail are classified as long classical OBPs. However, OBPs from *Drosophila melanogaster*, *Apis mellifera,* and *Anopheles gambiae* has a shorter C-terminal tail and belong to medium C-terminal subclasses.[57]

The structures of OBPs from *Anopheles gambiae* (AgamOBP1)[96], Aedes aegypti (AaegOBP1)[97] (Figure 1.33), and Culex quinquefasciatus (CquiOBP1·MOP)[98] from dipterans show the C-terminus is a loop, and the residues at the C-terminus (Val125 in AgamOBP1)[96], and (Ile125 in AaegOBP1)[97] forms a hydrogen bond with a Tyr 54. At lower pH, a disruption of this hydrogen bond destabilizes the C-terminal loop releasing it from the binding pocket. This results in the exposure of the binding pocket to the solvent. This decreases the binding affinity of the protein to the ligand at acidic pH. They undergo a pH-dependent conformational change.

Similarly, wang et al., in 2020, determined the crystal structure of odorant-binding protein 22 from *Aedes aegypti* (AeOBP22). They reported that in the bound state, the C-terminal tail forms a seventh α-helix and is situated to one end of the ligand-binding pocket. However, in the apo-state, the C-terminus is disordered.[99] This observation is opposite to what was observed for moth PBPs studied earlier. Similarly, the crystal structure of apo *Apis mellifera* odorant-binding protein (AmelOBP14) was solved, where the C-terminus forms the alpha helix, which is exposed outside the pocket protein.[100]

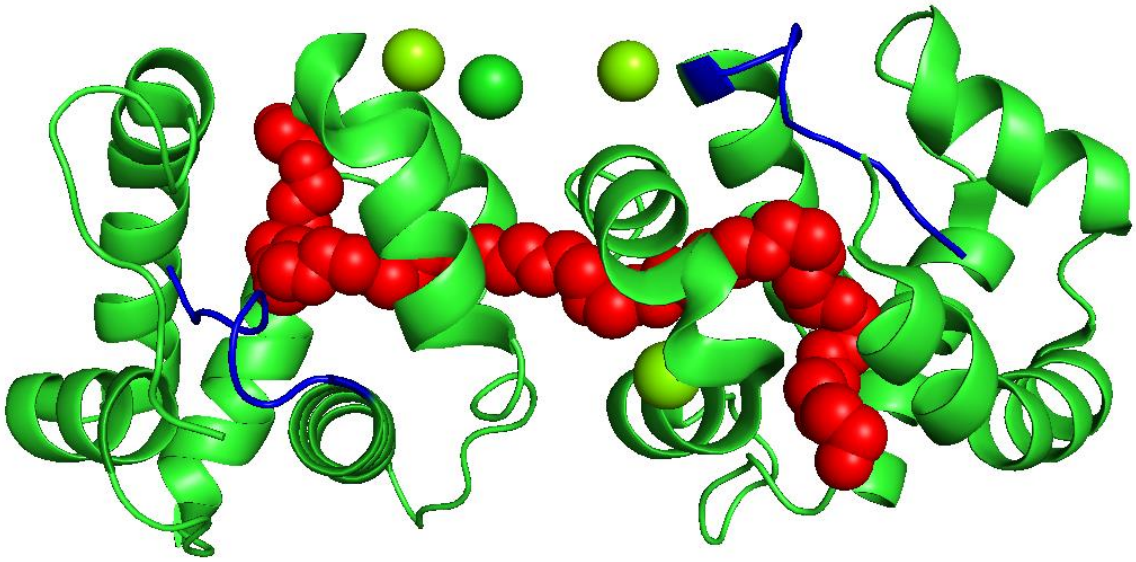


Figure 1.33: Crystal structure of odorant-binding protein 1 (AaegOBP1) from Aedes aegypti (PDB code: 3K1E)[97], red color showing ligand binding across the dimer and blue color representing the c-terminus region.

Likewise, the odorant-binding protein of the fruit fly *Drosophila melanogaster*, LUSH exists in a partially molten globule state/unstructured state in the absence of a ligand[101]. The binding of a ligand causes a conformational change to a folded and active form, increasing the overall protein stability[101]. The crystal structure of LUSH at pH 4.6 and pH 6.5 is the same. The ligand binding is independent of change in pH.[102]

Recently, structural studies were carried on pheromone binding protein on the European grapevine moth, *L. botrana* (LbotPBP1). It has been shown through the molecular dynamic simulation that the sex pheromone (14 carbon atoms) formed hydrogen bonds with Ser9, Ser56, and Trp114 to participate

in the specific recognition and stabilization of sex pheromones.[103] The chain length, functional group, and percentage of the pheromone blend likely play a crucial role in binding.[103]

The pheromone binding mechanism proposed for BmorPBP, AtraPBP1, and ApolPBP1 is not the universally accepted mechanism for the entire Lepidoptera order. For example, ApolPBP3 was also found to bind in both high and low pH levels.[104] The odorant-binding protein from different insect orders like *Dyctioptera* (cockroach *Leucophea maderae*; LmaPBP), *Hymenoptera* (honeybee, *Apis mellifera* ASP2) and *Diptera* (fruit fly *Drosophila melanogaster,* LUSH), and General odorant-binding protein from *Bombyx mori* (BmorGOBP2), have similar folding and architecture. However, there are differences in the relative positions of helices and key interacting residues. The variation in the mechanism depends on the species-specific interaction, potentially defined by pheromone chemical nature.

Broadly there are three different modes of conformational change in PBPs and GOBPs

1. The PBPs (BmorPBP, ApolPBP1, and AtraPBP1) exist in two major conformations:PBP^A or PBP^B based on the pH. At a pH lower than 5.0, the C-terminus forms an α-helix and occupies the binding pocket. At a pH higher than 6.0, the C-terminus region is in an extended conformation hanging outside the binding pocket. Thus, the binding pocket is available for pheromone binding at pH above 6.0. These proteins undergo both pH-induced and ligand-induced conformational changes.

2. In LmaPBP[90], D. melanogaster OBP, LUSH[102], Amel-ASP1[92], BmorGOBP[95], and chemosensory protein from *Mamestra brassicae*[105], the C-terminus is short to form a helix. There is no conformational change associated with pH. However, these proteins only undergo conformational change induced by ligand binding. For example, the cockroach PBP (LmaPBP)

a short-chain PBP with118 residues. Unlike medium and long-chain OBPs, this protein lacks the hydrophobic C-terminal peptide segment as seen for ApolPBP1, BmorPBP, and AtraPBP1. The binding pocket contains polar and charged residues, forming a hydrophilic pocket that binds a hydrophilic ligand.

3. Odorant binding proteins (OBPs) from dipterans *Anopheles gambiae* (AgamOBP1)[106], *Aedes aegypti* (AaegOBP1)[97], and *Culex quinquefasciatus* (CquiOBP1·MOP)[98] show a different mechanism of pheromone binding. These proteins lack a C-terminal segment unlike PBPs of *Antheraea. polyphemus, Bombyx mori,* and *Amyelois transitella.* Thus, there is no pH-dependent C-terminal coil to helix transition or competition of the C-terminus with the ligand for the binding pocket. But, they do undergo pH-dependent conformational change like the above moth PBPs without exhibiting ligand-dependent conformational change. These findings suggest a different mechanism of ligand binding and release.[97,98,106] Most likely their shorter C-terminus region cannot fold and from an α-helix. These mosquitoes OBPs exist as a dimer with a central cavity interconnected through a hydrophobic tunnel running through the dimer interface. The C-terminus of these OBPs is sheltered between helices 1 and 3 and form hydrogen bond interactions involving the C-terminal carboxylate, "lid" on the central cavity. It has been suggested that the "lid" is pH sensitive, and a drop in pH may result in disruption of the hydrogen bond network, resulting in the release of the ligand from the complex.

1.2.5 *Ostrinia furnacalis* Pheromone Binding Protein 2 (OfurPBP2)

Ostrinia furnacalis pheromone binding protein 2 (OfurPBP2) is the major PBP found in the male antennae of *Ostrinia furnacalis.*[107] Five pheromone-binding proteins from *Ostrinia furnacalis* have

been reported. The PBP2 and PBP3 have been shown to have a male-biased expression in the male antennae, suggesting that these proteins are involved in detecting female-secreted pheromone.[107] The *Ostrinia furnacalis*, also known as the Asian corn borer (ACB) belongs to the family Crambidae and order Lepidoptera. This species is considered as the model system to study pheromone evolution because of the following reason:

Presence of the Δ14-desaturase Gene

Moth pheromone is secreted by the female abdominal gland from the unsaturated fatty acid precursor produced by desaturases that show the range of stereo and regio-specificities. ACB is unique *Ostrinia* species which is known to use the Δ14-desaturase gene that produces its unique mixture of Z/E12-14:OAcs pheromone component.[108,109]

Odorant Receptor Genes (OR): Polymorphism or Mutation on OR Genes

The odorant receptor genes are present and expressed in the male moth antennae. These genes are responsible for detecting and discriminating sex pheromones produced by females. Hence, they are called sex pheromone receptors.[110] A single amino acid polymorphism at position 148 of the third transmembrane domain of the receptor gene controls the selective response toward the *E*-12 and *Z*-12 pheromones produced by ACB females.[109,111]

Position of the Double Bond in Pheromone

The genus *Ostrinia* is considered an excellent model system to elucidate the molecular mechanism underlying moth olfaction.[112,113] The sex pheromones from the Lepidoptera moth are mono- or di-unsaturated C10 to C18 straight-chain aldehydes, alcohols, or acetates with the site of unsaturation at the odd and even position in their carbon chain.[114] The mono-unsaturated compound with unsaturation at the even-numbered position in the carbon chain is extremely rare.[114] The *Ostrinia* female produces a blend of monosaturated tetradecenyl acetate (C14) sex pheromone, where the position of double bonds 9, or 11, or 12) and geometry [cis (*Z*) or trans (*E)*][113] varies in different species. Moreover, pheromone

specificity could be in part due to the difference in pheromone blend ratios, the double bond position, and the presence of enzymes involved in pheromone biosynthesis.

Furthermore, in the *Ostrinia,* most species use a different ratio of *E*-11 and *Z*-11-tetradecenyl acetate as their pheromone. ACB is unique within the *Ostrinia*, having evolved to use the same acetate pheromone but with a shift in the location of the double bond, (*E*)-12- and (*Z*)-12-tetradecenyl acetate (*E*-12 and *Z*-12).[109] This subtle structural change in the pheromone structure imparts the species specificity for the pheromone. The *E*-12 and *Z*-12 are the unique pheromones within the Lepidoptera.[115]

The Blend of Pheromones: Difference in the Blend Ratio

The female moth species produces a 1:1 mixture of *Z*-(cis) and *E (trans)*-12-tetradecen-1-ol acetate.[114] The ACB reductases (the enzyme that catalyzes the specific reduction of the fatty-acyl pheromone precursors into fatty alcohol) slightly prefer *E*-12 over *Z*-12.[116]

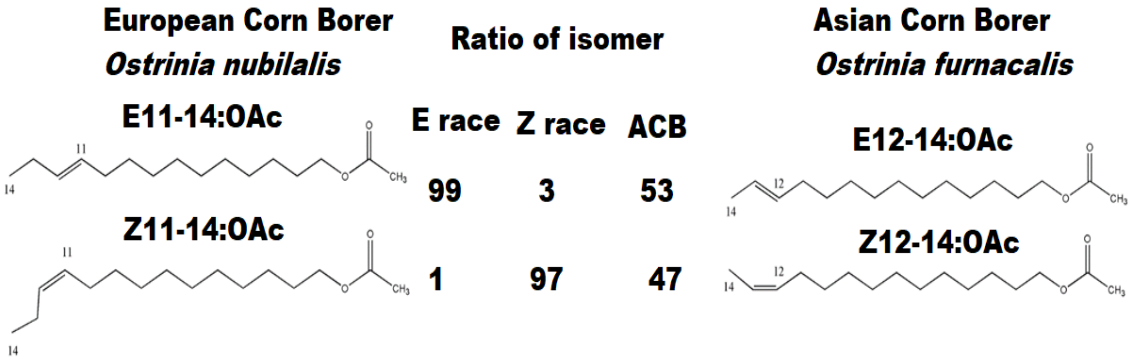


Figure 1.34: Pheromones used by the corn borers, with ratios of each isomer in the pheromone blend. Asian corn borers use *E* and *Z*-12-14:OAc, whereas European corn borers use *E* and *Z*-11-14:OAc.[117]

The sex pheromone ratios of the Asian Corn Borer vary based on geographical locations. The female sex pheromone of the Asian Corn Borer, *Ostrinia furnacalis*, in Taiwan, has been confirmed as (*Z*)-12-tetradecenyl acetate and its geometric isomer (*E*)-12-tetradecenyl acetate in a ratio of 3:1 by gas chromatography and gas chromatography-mass spectrometry in a selected ion monitoring mode.[118] In China, the ratio of sex pheromone components (*E*: *Z*-12-tetradecenyl acetate) of *O. furnacalis* has been identified as 53:47[119], 1:1 (E: Z) in the Philippines[120] and 2 : 3 (E: Z) in Japan.[121-123, 124] But for *Ostrinia nubibalis* (European corn borer), the ratio of *E*-11 to *Z*-11 is 99:1 (called as E-race) and 3:97 (called as Z-race).[125]

1.2.6 Status of Corn Borer in the USA

The European Corn Borer (ECB) was introduced in North America in early 1917. ECB caused crop losses of over 1 billion US dollars annually nationwide and 35 million US dollars in the northeastern United States.[126,127,128] Over 80 million acres of field corn (*Zea mays*) and 600,000 acres of sweet corn, worth about $40 billion and $1 billion, respectively, are grown in the U.S. each year. ECB accounts for over $1.85 billion in control costs and grain losses annually.[129] In 2006, 88% of the fresh market sweet corn acreage was treated with one or more insecticide applications for a total of 605,000 lbs of insecticides. ACB attacks many other crops, such as sorghum, small grains, potatoes, beans, tomatoes, and peppers.[122] The Southwestern Corn Borer, *Diatraea grandiosella,* causes about $1 million in damage in the Western High Plains.[130] Corn or maize is a staple crop of many Native Americans and is grown in dozens of shapes and colors. The yield of crops has been dramatically influenced by pests carried by humans, either intentionally or accidentally, into this region. The European corn borer originated in Eurasia and was accidentally introduced into North America.[131] This insect readily

adopted corn as a host and has since caused hundreds of millions of dollars in crop losses. In Canada and the U. S, losses resulting from ECB damage and control costs exceed $1 billion annually.[113]

1.2.7 Asian Corn Borer *Ostrinia furnacalis (Guenee*)

Identification

Initially, ACB was described as *Botys furnacalis* by the scientist, Guenee in 1854.[122] The Asian Corn Borer (ACB) is also known as the *Ostrinia furnacalis* (Guenee). It was first recorded as a pest of maize in Southeast Asia in *1906.*[132] Before 1966, it was misidentified as a European corn borer, *Ostrinia nubilalis*. After 1970, Mutuura and Monroe revised the genus *Ostrinia* and named it as separated species.[122] In Japan, it is also called oriental corn borer.[121]

Distribution of Species and Host

ACB was found in India, China, Korea, Japan, Australia, and Western pacific islands, including Java, Sulawesi, Philippines, Borneo, New Guinea, Solomons, Guam, Europe, Oceania, and western Micronesia.[122] The primary food source for the Asian Corn Borer is corn. It also feeds on bell peppers, cotton, hope, millet, pearl millet, foxtail millet, sugarcane, sorghum, and ginger. Although this insect feeds on several crops, it is a significant pest of corn, *Zea mays*. It has frequently caused severe damage to corn in many countries, such as Japan, Korea, China, Philippines, Indonesia, Thailand, Malaysia, Australia, Marianas, Papua New Guinea, and Solomon Islands.[133] The caterpillar stages are doing the most damage. They bore holes into the stems and cobs and feed on the silk as well as the kernels. The mature caterpillars commonly feed on the stalks. The leaves may wilt above the entry holes. (*http://www.pestnet.org/fact_sheets/maize_asian_corn_borer_115.htm*)

Yield Loss

ACB causes severe yield loss of corn. The yield loss is around 20-80 % in the Philippines, 100 % in the Marianas, 95 % in Taiwan.[122] In China, the Asian Corn Borer (*Ostrinia furnacalis*, Guenée) is considered one of the most destructive insect pests of maize and causes an estimated loss of 6–9 million tons annually.[134]

The Problem of Using Pesticides

The use of pesticides causes environmental pollution and poses a threat to human health. These pests develop resistance to pesticides, and there is the possibility of a secondary pest outbreak[135], which another challenge. The application of insecticides is logistically challenging because of the height of the crops. There is only a small window of the opportunity to kill larvae as they live inside a stalk. So, the search for alternative control of these pests is the pressing need of today.

Olfactory-based insect control based on pheromone will be a promising strategy. This method has many advantages such as non-toxic, species-specific, low cost, and important for integrated pest management. To achieve this, we need to understand how pheromones and pheromone binding proteins activate the signals for mating and reproduction. Thus, understanding the pheromone recognition, binding, and release by the pheromone-binding protein is very important to design "anti-pheromone" or pheromone mimetics to control the agricultural pest. This study involves the multidimensional approach to investigate the structure and function of *Ostrinia furnacalis* pheromone-binding protein (OfurPBP2).

CHAPTER II

EXPRESSION, PURIFICATION, AND BIOPHYSICAL CHARACTERIZATION OF *OSTRINIA FURNACALIS* PHEROMONE BINDING PROTEIN 2 (OfurPBP2)

2.1 Introduction

Lepidopteran male moths have a highly selective olfactory system capable of detecting female-secreted pheromones. Pheromone-binding proteins (PBPs) in male antennae bind and transport the hydrophobic pheromone across the aqueous lymph to the olfactory receptor. These moths PBPs bind pheromone at physiological pH and release them at acidic pH near the receptor. In previous studies, five different pheromone-binding proteins from *Ostrinia furnacalis* have been reported.[125,136] Among them, PBP2 and PBP3 have been shown to have a male-biased expression in the antennae of male moths. These proteins are actively involved in detecting female-secreted pheromone.[136] The *Ostrinia* moth recognizes pheromones that blend cis-trans isomers of *E*-11 and *Z*-11-tetradecenyl acetate. The *Ostrinia. furnacalis* moth is uniquely evolved to use a mixture of *E*- and *Z*-12-tetradecenyl acetate (Figure 2.1) as their sex pheromone components. Although OfurPBP2 has 50 % sequence identity with well-studied proteins such as ApolPBP1 and BmorPBP, there are two striking differences observed for the two biological gates: the Arg70 replaces His70 in the His70-His95 gate[76], and there are four additional charged residues (Asp130, His131, Glu136, and Lys143 in the C-terminal gate.

The structure and mechanistic detail of the OfurPBP2 are not known. The impact of these substitutions on the structure and function of ligand binding will help to understand pheromone perception on *Ostrinia. furnacalis*. Furthermore, Lepidopteran PBPs that have been studied previously have two conformations: (i) ligand-binding conformation (B form), at high pH (ii) ligand releasing conformation (A form) at acidic pH. These PBPs undergo a conformational switch. The pheromone is released during this conformational change.[62,65,72,73,76,80-82] The pH-driven conformational change is associated with ligand binding at high pH and releases at low pH. [51,62,64,65,76,137-139] To understand the mechanism of ligand binding and release and the effect of pH, the detailed structural characterization of the *Ostrinia furnacalis* PBP2 is necessary.

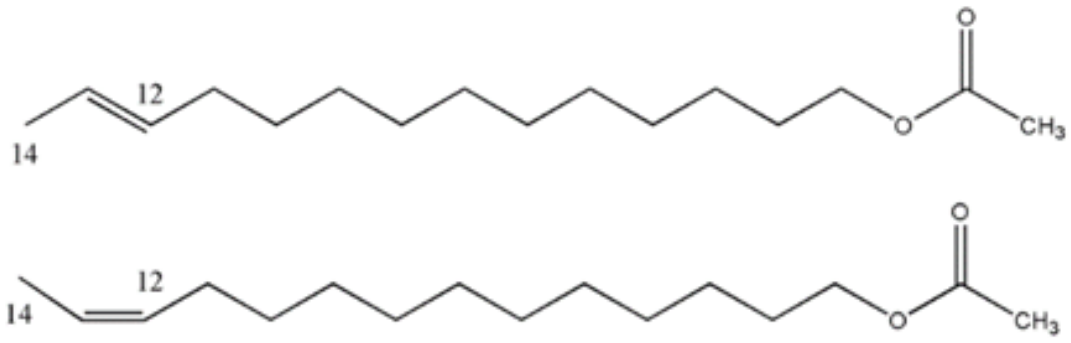


Figure 2.1: The chemical structures of the pheromone molecules of (E)-12 tetradecenyl acetate and (Z)-12-tetradecenyl acetate of *Ostrinia furnacalis***.**

OfurPBP2 is 16 kDa, soluble protein. It has six Cys residues, which are strictly conserved (Figure 2.2). The production of recombinant OfurPBP2 is essential for structural characterization. One of the important goals is to obtain well-folded protein in milligram quantities. For the recombinant

proteins that are expressed as inclusion bodies, refolding of the protein is another critical step. Overexpression of protein with appropriate isotopes is essential for structure determination by solution-state NMR. The common platform for the expression of the isotopically labeled protein is *E-coli* bacterial cells, as they can grow in well-defined minimal media. The screening of suitable host cell lines, optimization of temperatures, and media are necessary. This chapter discusses the refolding process, protein purification, and biophysical characterizations of OfurPBP2.

```
      10         20         30         40         50         60
SQAVMKDMTK NFIKAYEVCA KEYNLPEAAG AEVMNFWKEG YVLTSREAGC AILCLSSKLN

         70         80         90        100        110        120
LLDPEGTLHR GNTVEFAKQH GSDDAMAHQL VDIVHACEKS VPPNEDNCLM ALGISMCFKT

        130        140
EIHKLNWAPD HELLLEEMMA EMKQ
```

Figure 2.2: The amino acid sequence of the OfurPBP2.

Table 2.1: Lists of some properties of OfurPBP2

Molecular weight (Da)	16103.54
Theoretical pI	5.05
Number of amino acids	144
Total number of negatively charged residues (Asp + Glu)	22
Total number of positively charged residues (Arg + Lys)	13
Ext. coefficient (M^{-1} cm^{-1})	15845

2.2 Materials and Methods

a. Cloning and Overexpression of OfurPBP2

Dr. Suman Mazumer did the cloning of the OfurPBP2 gene and optimization of expression in Dr. Mohanty's laboratory[#]. Briefly, after identifying the signal peptide, the gene without the signal peptide was sub-cloned into the pET-21a vector. The gene-specific primers were amplified using forward primers: 5′-GGAATTCCATATGTCACAAGCAGTGATGAAAGAC-3′; and reverse primers 5′-GCGGATCCTCATTGCTTCATTTCGGCCAT-3 through the polymerase chain reaction (PCR). The amplified fragments were purified and were excised using NdeI and BamHI double digestion. It was then ligated between NdeI and BamHI restriction sites of pET21a vector (Novagen/EMD Millipore) by incubation with ligase. The recombinant pET21a containing OfurPBP2 insert was confirmed by sequencing at the core facility at Oklahoma State University. The recombinant pET21a/OfurPBP2 plasmid was introduced into an *Escherichia coli* Origami 2 cell (Novagen/EMD Millipore) by the process called transformation. A single colony from the transformed agar plate was selected and inoculated in 25 mL of Luria-Bertani (LB) media containing 100 µg/L of tetracycline and 100 µg/L of ampicillin. The culture was incubated at 37 ºC overnight. The overnight starter culture was diluted to an OD_{600} of 0.1 in fresh LB media (containing antibiotics: 100 µg/L of tetracycline and 100 µg/L of ampicillin), and grown at 37 ºC to OD_{600} of 0.50-0.60. At that point, the temperature was reduced to 30 ºC, and protein expression was induced with 1 mM of isopropyl-β-Dithiogalactopyranoside (IPTG). The cells were harvested by centrifugation after incubation for 6 hours.

[#] Cloning was done by Dr. Suman Mazumer in Dr. Mohanty laboratory: Mazumder, S., Dahal, S.R., Chaudhary, B.P. et al. Structure and Function Studies of Asian Corn Borer *Ostrinia furnacalis* Pheromone Binding Protein2. Sci Rep 8, 17105 (2018). https://doi.org/10.1038/s41598-018-35509-x

b. Overexpression of the isotope-labeled OfurPBP2

For expressing ^{15}N labeled protein, cells were grown in M9 minimal media culture containing 0.12% $^{15}NH_4Cl$ (Cambridge Isotope Laboratories). For $^{15}N/^{13}C$ double-labeled protein, cells were grown in M9 minimal media culture containing $^{15}NH_4Cl$ and ^{13}C-glucose as carbon sources. The ingredient of M9 media is listed in Table 2.2. The cells were grown for 16 hours after induction with IPTG, and all other procedures were the same as above. The cells were harvested by centrifugation at 9,000 rpm using a Sorvall LYNX 4000 centrifuge for 20 min at 4 °C and kept frozen at −20 °C until needed.

Table 2.2: Components of M9 minimal media for 250 mL culture

M9 Media (250 ml)	**Amount**
K_2HPO_4	3.25g
KH_2PO_4	2.5g
Na_2HPO_4	2.25g
K_2SO4	0.6 g
$^{15}NH_4Cl$	0.3g
1M $MgSO_4$	0.5 ml
Thiamine	1.25 ml
0.1 M $CaCl_2$	125 ul
Yeast Extract	0.5 ml
Trace element	250 uL
^{12}C glucose (for single label) ^{13}C glucose (for double label)	5 ml
Ampicillin 100 mg/ml)	250 ul
Tretracycline (100 mg/ml)	250 ul

c. Cells Freeze-thaw and Lysis

The cells were frozen in liquid nitrogen or a -80 °C freezer and thawed at room temperature for 7-8 cycles. Cells expand as ice crystals form during the freezing process and contract during the thawing process. This process weakens the cell walls. The cells were resuspended in a lysis buffer containing Bacterial Protein Extraction Reagent (B-PER) with 1 mM EDTA, 1 mM PMSF, and a cocktail protease inhibitor. The thawed cell suspensions were lysed using a sonicator. A short pulse of 8 sec was given and kept in the ice bath for 1 min with shaking, and the process was repeated 15 times.

d. Protein Refolding

1. Preparation of Inclusion Body (IB)

After lysis, the cells were centrifuged at 12,000 rpm for 30 min at 4 °C. The pellets were suspended in a 10% B-PER solution and then sonicated. The supernatant was removed after centrifugation. The washing was repeated two more times, with a 10% B-PER solution. The inclusion bodies were stored at -20 ºC until needed.

2. Solubilization of Inclusion Body

The inclusion body pellet (~0.5 gram) was dissolved in a 15 mL buffer containing 50 mM Tris-HCl pH 8.0 and 6 M guanidine hydrochloride (GdnHCl). This high concentration of GdnHCl acts as a denaturant. OfurPBP2 contains 6 cysteine residues, and thus 10 mM dithiothreitol (DDT) was added to reduce unwanted disulfide bond formations. The dissolved IB solution was left at room temperature overnight and then was centrifuged at 12,000 rpm for 30 min at 4 °C.

3. Dilution of the Solubilized Protein in Refolding Buffer

The supernatant (15 mL) was diluted with Dialysis Buffer 1 (DB #1; 50 mM Tris-HCl, pH 8.0 and 2 M Gdn.HCl) in a 1:1 ratio. All the dialysis buffers used for refolding the protein are listed in Table 2.3. The sample was transferred to the dialysis bag and kept in the 4 L of buffer (DB#1) overnight at 4 °C with slow stirring.

4. Refolding by Stepwise Dialysis

The primary purpose of dialysis is to gradually remove the denaturants and introduce disulfide-exchange reagents such as reduced and oxidized glutathione. The dialysis bag was transferred to 4 L of Dialysis Buffer 2 (DB#2), which contained (50 mM Tris-HCl, pH 8.0, 1 M Gnd.HCl, 0.8 M ArgHCl, 3 mM reduced glutathione, and 0.9 mM oxidized glutathione) and was kept overnight at 4 °C with stirring. The dialysis bag was transferred to Dialysis Buffer 3 (DB#3), which contained (50 mM Tris-HCl, pH 8.0, 0.5 M Gnd.HCl, 0.4 M ArgHCl, 1.5 mM reduced glutathione, and 0.45 mM oxidized glutathione). It was kept overnight at 4 °C with stirring. After dialysis on the third buffer overnight, the sample was centrifuged at 12,000 rpm at 4 °C for 30 min. Then the sample was kept on a new dialysis bag and kept in Dialysis Buffer 4 (DB# 4) (50 mM Tris-HCl, pH 8.0, 250 mM NaCl, 0.2 M ArgHCl, 3mM reduced glutathione, and 0.9 mM oxidized glutathione) and left overnight at 4 °C with stirring.

Table 2.3: List of dialysis buffers used during the refolding of OfurPBP2

Protein refolding from Inclusion body (IB)

Dialysis Buffers	Compositions	Volume	Dialysis time
Buffer # 1	50 mM Tris-HCl pH 8.0, 2 M Guanidine hydrochloride (Gdn.HCl)	4L	Overnight (O/N)
Buffer #2	50 mM Tris-HCl pH 8.0, 1M Guanidine hydrochloride (Gdn.HCl), 0.8 M Arginine hydro-chloride (ArgHCl), 3 mM reduced glutathione, and 0.9 mM oxidized glutathione	4L	Overnight (O/N)
Buffer #3	50 mM Tris-HCl pH 8.0, 0.5 M Guanidine hydrochloride (Gdn.HCl), 0.4 M Arginine hydro-chloride (ArgHCl), 1.5 mM reduced glutathione, and 0.45 mM oxidized glutathione (50 % of Buffer #2)	4L	Overnight (O/N)
Buffer #4	50 mM Tris-HCl pH 8.0, 250 mM NaCl, 0.2 M Arginine hydro-chloride (ArgHCl), 3 mM reduced glutathione, and 0.9 mM oxidized glutathione	4L	Overnight (O/N)

e. Purification of OfurPBP2

Final dialysis was done in a buffer (20 mM Tris-HCl, pH 8.0) overnight at 4 °C, and the protein solution was centrifuged at 12,000 rpm and 4°C for 25 min. The final dialysis was done to remove the salts or other small molecules.

The purification was performed by anion exchange DEAE chromatography. The net charge of the OfurPBP2 at pH 8.0 is negative. It binds to the column containing positively charged beads of diethyl aminoethyl (DEAE). A stepwise gradient method was used. The 5 mL dialyze protein sample was used injection on the DEAE column. Multiple injections were carried depending on the protein sample volume (for 30 mL sample, 6 injections were carried). The 20 mM Tris-HCl at pH 8.0 was used as a starting buffer. The elution buffer consists of 20 mM Tris-HCl 50 mM at pH 8.0 including 1 M NaCl. The elution chromatogram consists of multiple peaks. Each peak point sample was taken and analyzed by SDS-PAGE. The peak "A" in the chromatogram is the peak of interest, which contains OfurPBP2. The fraction from 51 to 56 contains protein of our interest. These fractions were collected and concentrated by centrifugation at 3,000 rpm at 4 °C.

Final purification was carried by size exclusion chromatography using a Superdex 75 column fitted to an ÄKTA FPLC (GE healthcare). After concentrating the fraction from anion exchange chromatography, 1.5 ml of concentrated protein sample was injected into size exclusion chromatography. The injection was carried out 3 times. The mobile phase consists of 20 mM phosphate buffer, 150mM NaCl, 1 mM EDTA, and 0.01% (w/v) NaN_3. The protein was eluted isocratically. The fractions containing the pure monomeric protein (E1 to E12) were collected based on SDS PAGE analysis and stored at 4 °C. Protein concentration was calculated from the absorption at 280 nm (A_{280}) as 15845 M^{-1} cm^{-1} extinction coefficient[140], as explained in these references (141 and 142).[141,142]

f. Delipidation of OfurPBP2

The delipidation of OfurPBP2 was performed by Dr. Mohanty by modifying the original protocol mentioned by Bette *et al.*[143] and Katre *et al.*[65] We have optimized the protocol by modifying

temperature, incubation time, time of shaking, and also the volume of the Lipidex resin. Briefly, the protein was buffer-exchanged to 50 mM sodium citrate buffer at pH 4.5 (buffer A) and concentrated to 1.0 mL using a Millipore ultrafiltration concentrator with a molecular weight cut-off of 3,000 Da. The 15 mL of Lipidex ^{TM}-1000 resin was manually packed and washed 15-20 times with water to remove the residual methanol and equilibrated with buffer, and then equilibrated with a citrate buffer. The protein was loaded in the Lipidex column and incubated at 37 ℃ for 30 minutes, and then was eluted manually under gravity until the absorbance at A_{280} was negligible. The eluted protein was concentrated to 2 mL and was buffer exchanged to 15 mM sodium phosphate buffer at pH 6.5 with 1 mM EDTA, 0.01% sodium azide, and 10% D_2O. Protein concentrations were determined spectrophotometrically using the theoretical extinction coefficient, A_{280} of 15845 $M^{-1}cm^{-1}$.

g. Matrix-assisted Laser Desorption Ionization (MALDI)

The MALDI-TOF measurement was done on the protein sample, which was desalted with a CENTRI-SPIN 10 column (Princeton Separations, NJ). For the matrix solution, 2,5-Dihydroxybenzoic acid (DHB) was used. MALDI spectra were collected on a Voyager DE-PRO MALDI-TOF mass spectrometer (Applied Biosystems) at Oklahoma State University's core facility.

h. Circular Dichroism (CD)

The circular dichroism (CD) experiments were performed on a Jasco J-810 spectropolarimeter. The quartz cell cuvette with a 0.05-cm was used. The data were collected at 25°C. The concentration of samples was 30 μM in 20 mM phosphate buffer at pH levels of 6.5, 5.5, and 4.5. Three different samples were prepared with the same concentration and volume at three different pH. The buffer

used was a 20 mM phosphate buffer. The concentration of the protein was 30 µM, and the volume was 130 ul of the protein sample was taken in the cuvette. The data were averaged over 5 scans with a response time of 4 s and with a scan speed of 100 nm/min. The spectra were corrected by subtracting the blank spectra (buffer). CD ellipticity values were converted to normalized values (mean molar ellipticity per residue) by the standard method. The secondary structure contents were quantified through deconvolution of CD spectra by using CDSSTR, CONTINLL, and SELCON3 programs incorporated in the CDPro software package.[144] During deconvolution, the number of amino acids/residues was taken as 144, and a molecular weight of 16.2 kDa was used.

i. Thermal Denaturation by Circular Dichroism

Unfolding of OfurPBP2 with increasing temperature was monitored by measuring the change in mean residual ellipticity at 222 nm using a Jasco J-810 spectropolarimeter. For this study, 30 uM protein sample in a 0.05 cm path-length cuvette either at pH 6.5 or 4.5 was heated from room temperature to 108 °C. The temperature was ramped at 2 °C/min and controlled by a Jasco programmable Peltier element. A scan rate of 1 °C/min was used. Far-UV CD spectra were recorded every 2 °C, and the dichroic activity at 222 nm was continuously monitored every 2 °C with a 4-sec averaging time. All the spectra were corrected using the buffer. The reversibility of thermal denaturation was checked by cooling the thermally denatured protein at room temperature. The reversible thermal denaturation process was analyzed by fitting baseline and transition data to a two-state model.

j. Fluorescence Spectroscopy

The fluorescence binding assay was performed on a PerkinElmer LS 55 Fluorescence Spectrophotometer at room temperature with a 1 cm light-path quartz cuvette. The emission and excitation slit widths were set to 5 nm. The sample was excited at 337 nm, and emission spectra were recorded from 370 to 600 nm. All experiments were repeated thrice for reproducibility. The extrinsic fluorescent probe *N*-phenyl-1-naphthylamine (1-NPN) with the 1mM solution in methanol was used to monitor change in fluorescence intensity at 420 nm. A 1 uM protein solution of OfurPBP2 was prepared in a 20 mM phosphate buffer, pH 6.5, in the presence of 0.3% methanol at room temperature. Phosphate buffer with the appropriate amount of 1-NPN and methanol was prepared for the control experiment. The affinity of OfurPBP2 was determined by adding a final concentration of 0–20 μM of 2 mM 1-NPN stock solution. The fluorescence spectra were recorded after incubation for 10 min at the excitation wavelength of 337 nm and emission of 370–600 nm. The amount of protein, amount of ligand added during the fluorescence experiment is shown in Table 2.4. To calculate the binding constant, relative fluorescence intensity (F_R) of the protein at different NPN concentrations were calculated as $(F_c - F_{min})/(F_{max} - F_{min})$, where F_c has corrected fluorescence intensity at ligand concentration [C], F_{min} is the minimum fluorescence intensity when ligand concentration is 0 μM, and F_{max} is the maximum fluorescence intensity. The data were fitted using OriginPro version 6.1 to a nonlinear curve fit of the plot of $(F_c - F_{min})/(F_{max} - F_{min})$ against [C] with the equation corresponding to a single binding site. The K_d values were calculated using Equation 2.2,

$$y = \frac{B * X}{K + X} \tag{2.2}$$

The B is the maximum relative fluorescence intensity, y is the relative fluorescence intensity, and X concentration of ligand.

Competitive Displacement Assay

The competitive displacement of NPN from the delipidated OfurPBP2 was performed with E and Z pheromone. The 2 μm delipidated OfurPBP2 was equilibrated overnight with 2 μm NPN at 4 °C, which were then titrated with 1.0 mM stock of sex pheromone (E and Z). After each addition, the complex was incubated for 10 min before recording the spectrum. The fluorescence spectra of pheromones with NPN in the absence of the protein served as controls. The spectra were recorded for triplicate. The binding affinity of each of the two sex pheromones was measured by using relative fluorescent intensities were analyzed by Origin. The IC_{50} values were determined at the ligand concentrations where the NPN fluorescence was quenched to half of its maximal intensity. The calculation was done using the equation: $Ki = [IC_{50}]/(1 + [1\text{-NPN}]/K1\text{-NPN})$, where [1-NPN] is the free concentration of 1-NPN and $K_{1\text{-NPN}}$ is the dissociation constant of the complex protein/1-NPN.

Table 2.4: Amount of pheromone added in a competitive displacement assay

Protein:NPN: Pheromone	Volume of pheromone	Total volume added
1:1:0.0	0.0 ul	0.0ul
1:1:0.025	0.15 ul	0.15 ul
1:1:0.05	0.15 ul	0.30 ul
1:1: 0.075	0.15 ul	0.45 ul
1:1: 0.09	0.15 ul	0.60 ul
1:1: 0.115	0.15 ul	0.75 ul
1:1:0.15	0.15 ul	0.90 ul
1:1:0.2	0.3 ul	1.2 ul
1:1:0.25	0.3 ul	1.5 ul
1:1:0.3	0.3 ul	1.8 ul
1:1:0.4	0.6 ul	2.4 ul
1:1:0.5	0.6 ul	3.0 ul
1:1:0.7	1.2 ul	4.2 ul
1:1:1	1.8 ul	6.0 ul
1:1:1.4	2.4 ul	8.4 ul
1:1:2	3.6 ul	12 ul

k. NMR Measurements

The NMR data were collected at 35 °C on a 800 MHz NMR Spectrometer, CP-TCI, 5mm inverse triple resonance high-resolution cryoprobe, with actively shielded single axis Z-gradient at Oklahoma State University. The pH titrations were carried out on NMR samples contained 400 μM uniformly ^{15}N-labeled OfurPBP2 in 50 mM phosphate buffer at pH 6.5, 5% D_2O, 1 mM EDTA, and 0.01% (w/v) NaN_3 in a NMR Shaped tube. The pH titrations of OfurPBP2 were carried out at pH 6.5, 5.5, and 4.5. The protein at pH 4.5 was back titrated to pH 6.5 using 1 M NaOH. The 2D-{^1H, ^{15}N} hetero-nuclear single quantum coherence (HSQC) spectra were collected at each pH. All data were processed using NMRPipe and analyzed by Sparky.

2.3.**Results and Discussion**

2.3.1. **Cloning and Optimization of Expression**

The cloning of the OfurPBP2 gene into the pET21a vector and the optimization of the pET21-OfurPBP2 in various bacterial strains were carried out by Dr. Suman Mazumder in Dr. Mohanty's laboratory#.[145] The complete OfurPBP2 gene sequence result is given in Figure. 2.3. The lac operator controls the transcription of the protein of interest in the pET21a vector, where IPTG (isopropyl-β-D thiogalactopyranoside) was used to induce the protein expression. The lac repressor inhibits the genes to produce the protein, However, when IPTG is present, it binds to the lac repressor and releases the tetrameric repressor from the lac operator, thus allowing the protein to express. The protein expression was optimized using Origami 2 cells (Stratagene). Origami 2 strains have mutations in glutathione reductase (gor) and thioredoxin reductase (trxB), facilitating proper disulfide bond formation.

Mazumder, S., Dahal, S.R., Chaudhary, B.P. et al. Structure and Function Studies of Asian Corn Borer *Ostrinia furnacalis* Pheromone Binding Protein2. Sci Rep 8, 17105 (2018). https://doi.org/10.1038/s41598-018-35509-x

ATG TCA CAA GCA GTG ATG AAA GAC ATG ACG AAG AAC TTT ATA AAA GCC TAT GAA GTG TGT GCA AAG
ATG TCA CAA GCA GTG ATG AAA GAC ATG ACG AAG AAC TTT ATA AAA GCC TAT GAA GTG TGT GCA AAG
M S Q A V M K D M T K N F I K A Y E V C A K
GAG TAC AAT CTG CCT GAG GCC GCA GGA GCA GAG GTG ATG AAC TTT TGG AAG GAA GGC TAC GTG
GAG TAC AAT CTG CCT GAG GCC GCA GGA GCA GAG GTG ATG AAC TTT TGG AAG GAA GGC TAC GTG
E Y N L P E A A G A E V M N F W K E G Y V
TTG ACG AGT CGC GAG GCA GGA TGC GCC ATC CTC TGC CTT TCA TCC AAG CTG AAC CTG CTG GAC CCT
TTG ACG AGT CGC GAG GCA GGA TGC GCC ATC CTC TGC CTT TCA TCC AAG CTG AAC CTG CTG GAC CCT
L T S R E A G C A I L C L S S K L N L L D P
GAG GGG ACT CTG CAC CGT GGA AAT ACT GTC GAG TTC GCC AAG CAA CAT GGC TCT GAC GAC GCT ATG
GAG GGG ACT CTG CAC CGT GGA AAT ACT GTC GAG TTC GCC AAG CAA CAT GGC TCT GAC GAC GCT ATG
E G T L H R G N T V E F A K Q H G S D D A M
GCT CAC CAA CTG GTT GAC ATT GTC CAT GCT TGC GAG AAG TCC GTC CCG CCC AAT GAA GAC AAC TGC
GCT CAC CAA CTG GTT GAC ATT GTC CAT GCT TGC GAG AAG TCC GTC CCG CCC AAT GAA GAC AAC TGC
A H Q L V D I V H A C E K S V P P N E D N C
CTG ATG GCG TTG GGC ATC TCC ATG TGC TTC AAG ACC GAG ATC CAC AAG CTG AAC TGG GCG CCC GAC
CTG ATG GCG TTG GGC ATC TCC ATG TGC TTC AAG ACC GAG ATC CAC AAG CTG AAC TGG GCG CCC GAC
L M A L G I S M C F K T E I H K L N W A P D
CAC GAG CTG TTG CTA GAG GAG ATG ATG GCC GAA ATG AAG CAA TGA
CAC GAG CTG TTG CTA GAG GAG ATG ATG GCC GAA ATG AAG CAA TGA
H E L L L E E M M A E M K Q

Figure: 2.3: Nucleotide (top) and corresponding amino acid (bottom) sequences of OfurPBP2.

2.3.2. **Protein Overexpression and Protein Refolding**

For the structural characterization, milligram quantities of the pure protein were required. The yield of the soluble protein was negligible. Recombinant OfurPBP2 was primarily expressed as inclusion bodies (IB). During recombinant protein expression, inactive and insoluble materials accumulate as intracellular aggregates, which are called inclusion bodies.[146] The bacterial system may not support the appropriate pairing of disulfide bonds in the recombinant protein, leading to insoluble protein pellets.[147] Denaturing agents are used to dissolve the IBs and the protein is refolded to the native form. The OfurPBP2 was refolded by step-wise dialysis which was explained earlier. The protein refolding process is rigorous, expensive, and challenging. The significant advantage of refolding inclusion bodies is to obtain a large quantity of highly pure protein in the native form.

For example, from 500 mL of culture, we obtained approximately 30 mg/mL quantity of pure OfurPBP2. The SDS PAGE is showing an analysis of protein refolding. (Figure 2.4).

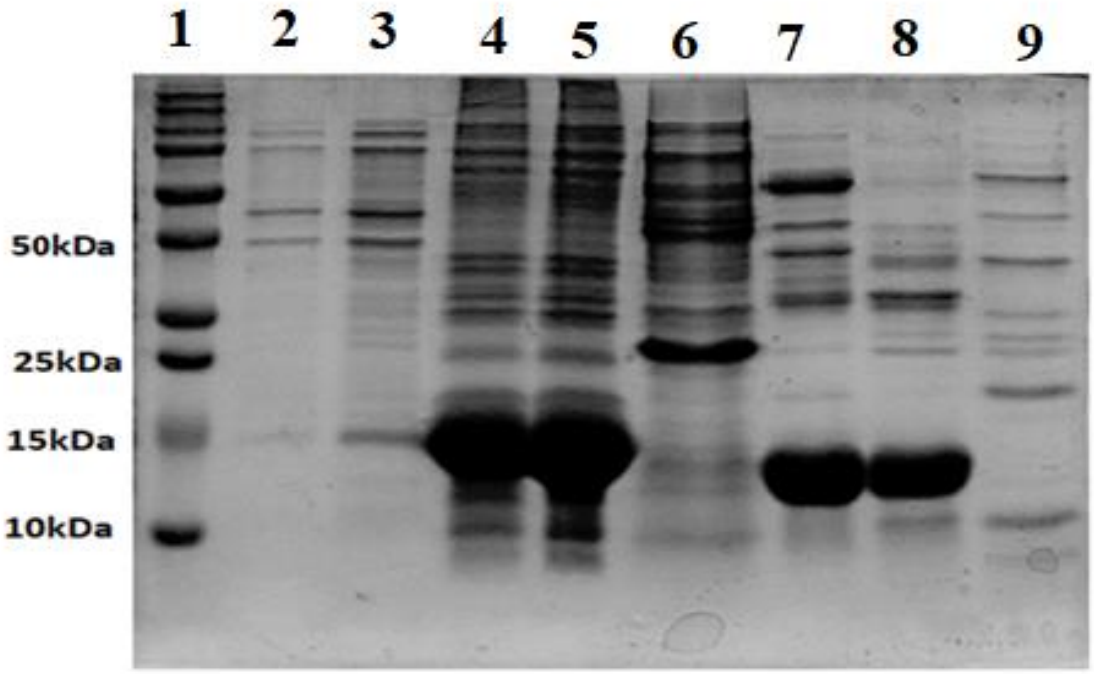


Figure 2.4: SDS-PAGE analysis of OfurPBP2 refolding. Lane 1: protein molecular weight marker, lanes 2, 3, and 9 are supernatants, lanes 4 and 5 inclusion bodies(IBs), lane 6 pellets after centrifugation, lane 7-8 protein after final refolding.

2.3.3. **Protein Purifications**

1. Dialysis

Dialysis is a widely used technique for removing small molecules from the protein through a semipermeable membrane, such as a cellulose membrane, based on the diffusion principle. Small molecules can pass through the membrane, while large biological molecules remain inside the dialysis bag. The primary purpose of dialysis is to remove salt and other small molecules. The

dialysis was done at 4 °C, as the temperature plays an important role. Similarly, the volume of the buffer used for the dialysis also plays an important role. The 25 mL protein sample was taken in the dialysis bag and kept in the 4L of 20 mM Tris-HCl pH 8.0 buffer overnight at 4 °C.

2. Ion Exchange Chromatography

Ion exchange chromatography separates proteins based on their net charge. OfurPBP2 has an isoelectric point of 4.5. For the anion exchange chromatography, Tris buffer at pH 8.0. was used. At this pH, the protein has a net negative charge, and it can bind to the positively charged beads of diethyl aminoethyl (DEAE) cellulose in an anion exchange column. Adsorbed protein molecules are desorbed from the resin competitively and are eluted by the competing chloride ions as the concentration of these ions are increased in the mobile phase. Two different methods can be used to elute the protein from a DEAE column, step-gradient, or linear gradient. During the purification of OfurPBP2, we employed the linear gradient to estimate the percentage of salt necessary for the elution of the protein. Later, a step gradient was used to elute the protein. A stepwise gradient scheme for the DEAE purification is shown in Figure 2.5. SDS-PAGE analysis was conducted for each peak shown in Figure 2.4. Peak “A” in this chromatogram was identified to contain refolded OfurPBP2.

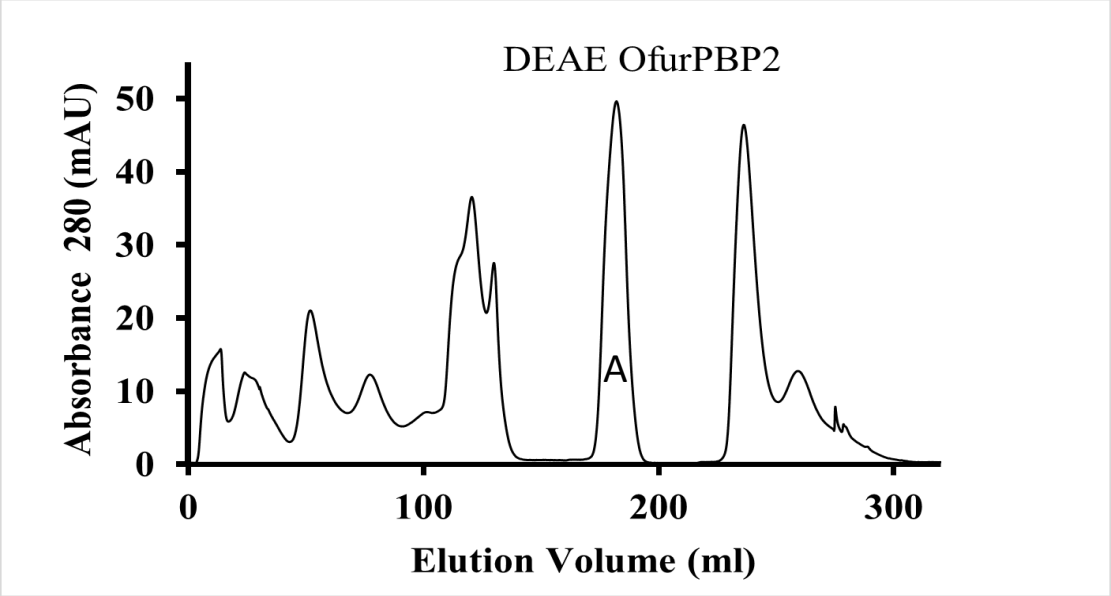


Figure 2.5: Elution profile of the OfurPBP2 from ion exchange with DEAE column.

3. Size Exclusion Chromatography

Finally, the protein was purified using size exclusion chromatography (also known as gel filtration chromatography). In this chromatographic method, the molecules in the solution are separated by their size or molecular weight. A column consists of porous beads as a stationary phase made of an insoluble hydrated polymer such as dextran or agarose or polyacrylamide. Sephadex, sepharose, and bio-gel are commercially available substrates for the stationary phase. The mobile phase consists of a phosphate buffer at pH 8.0. As the solution travels down the column, large molecules pass around the beads. Because of their large size, they cannot enter inside the pore and elute first, while small molecules travel through the pores of the stationary phase and take a longer time to elute. The samples were eluted isocratically, so there was no need to use different buffers during the separation. OfuPBP2 was purified using a Superdex-75 column (Figure 2.6) fitted to ÄKTA

FPLC (GE healthcare). The E1 to E12 consist of pure protein. The SDS-PAGE in Figure 2.7 is showing pure protein.

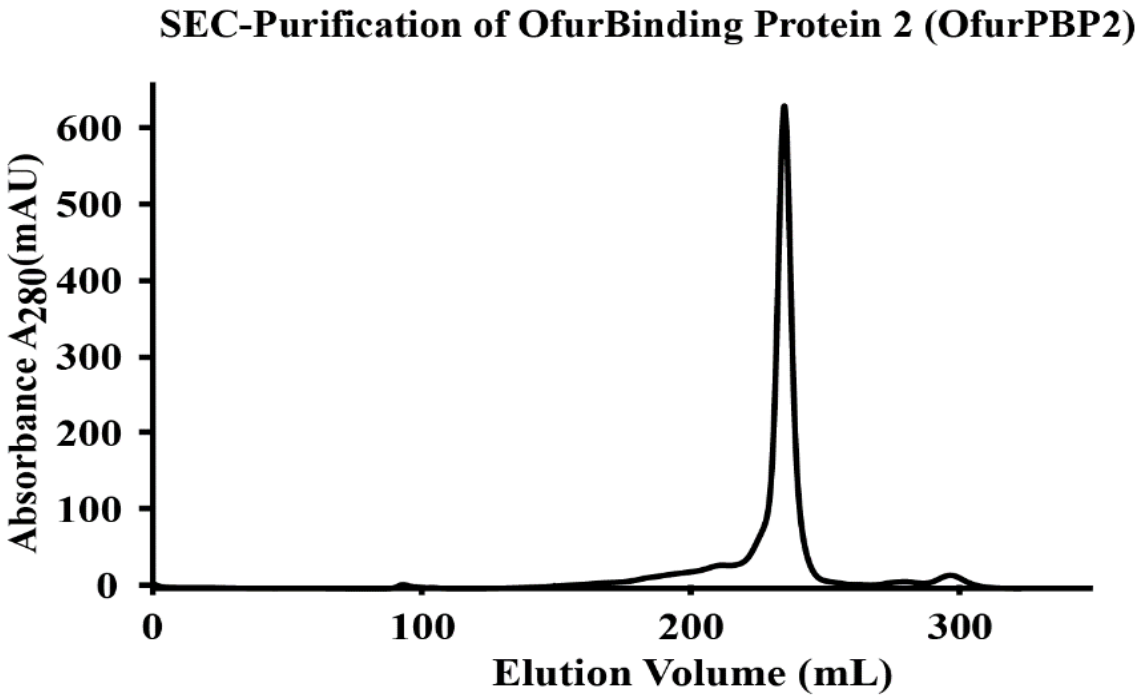


Figure 2.6: Size-exclusion chromatography (SEC) profile of OfurPBP2.

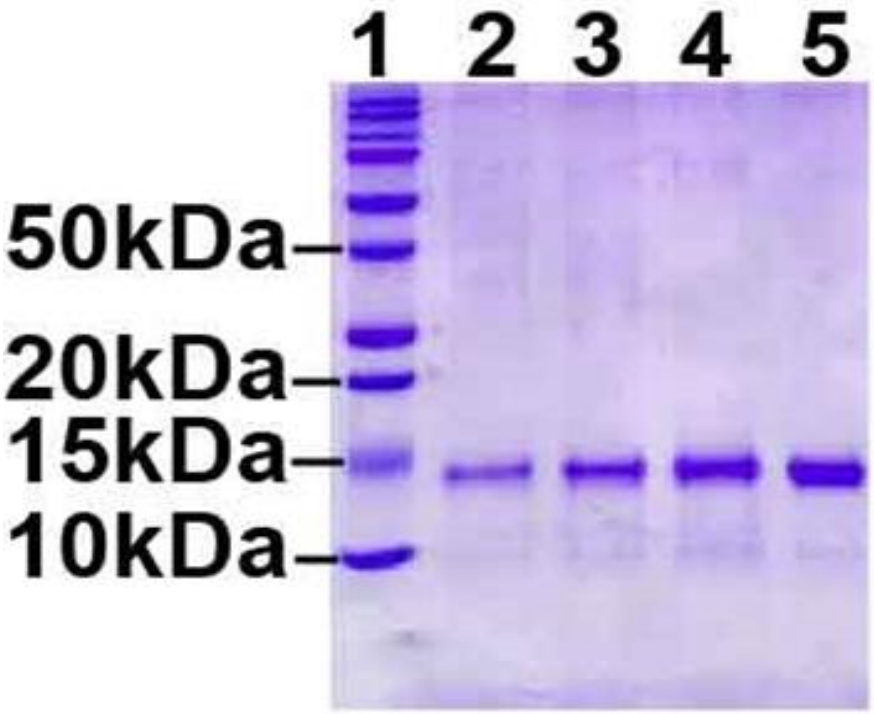


Figure 2.7: SDS-PAGE after purification of OfurPBP2. The mobility of the protein in the SDS-PAGE gel is dependent on its molecular mass. Lane 1: protein molecular weight marker; lanes 2-5: pure protein after size exclusion chromatography. The single protein band after SEC indicates that the protein is very pure.

2.3.4. **Delipidation of OfurPBP2**

During the expression of lipid-binding proteins, the protein picks up a hydrophobic molecule from the host bacterial system.[65] OfurPBP2 was expressed bound to an endogenous hydrophobic ligand from the bacteria. The endogenous lipids were removed from the protein with hydrophobic interaction column chromatography (HIC). This process is called delipidation. These hydrophobic compounds bind to Lipidex-1000 resin as the protein passes through the column.[148] Ligand-binding

assays were carried out with the free OfurPBP2 after the removal of the bacterial ligand through the delipidation process.

2.3.5. **Matrix-assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF)**

SDS-PAGE can determine an approximate molecular mass from a polypeptide chain's relative mobility versus protein marker. Mass spectroscopy was used to determine the accurate molecular weight of the protein. The concentration of the protein sample was 100 uM for MALDI-TOF. Protein was mixed with 10 mg/mL of beta-hydroxybutyric acid in a 1:1 ratio and applied to the metal plate. The matrix transforms the laser energy into excitation energy for the sample, which leads to the sputtering of analyte and matrix ions from the surface of the mixture. When the laser beam pulsates across the plate, the laser energy is absorbed by the matrix, causing ionization. Once ionized, these ions are accelerated into the time-of-flight mass analyzer. In a positive ionization mode, the protonated molecular ions ($M+H^+$) are usually the dominant species. Sometimes doubly charged molecular ions occur at approximately half the m/z value, and dimeric species at about twice the m/z value are also observed. The mass spectrum of purified OfurPBP2 exhibited a molecular ion peak at 16.092 kDa, which matched the theoretically calculated molecular mass of 16.109 kDa shown in Figure 2.8.

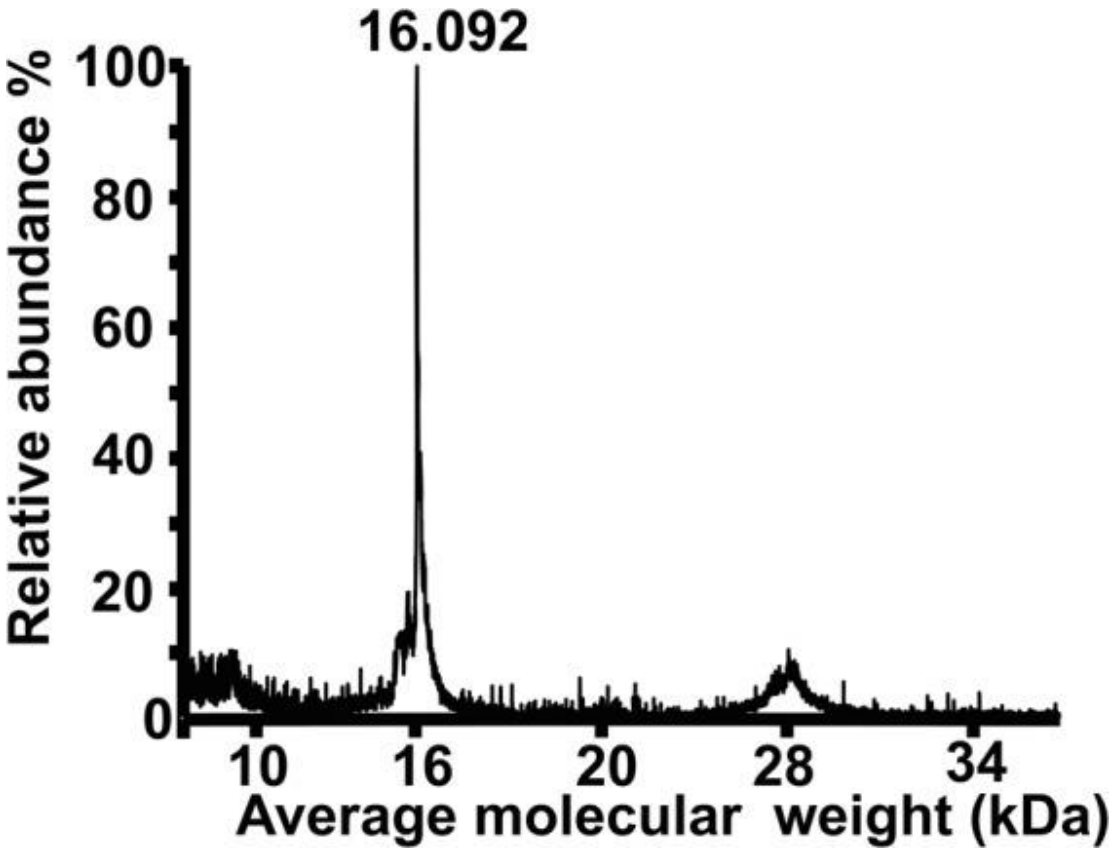


Figure 2.8: MALDI-TOF analysis of the molecular mass of the purified OfurPBP2.

2.3.6. **Circular Dichroism**

Circular dichroism (CD) is an excellent method of determining the secondary structure of proteins. It is based on the differential absorption of left and right circularly polarized light. Circular dichroism spectra are a signature of the protein. The CD spectrum is the sum of all individual residues protein, influenced by the protein 3-dimensional structure. Protein secondary structure can be determined by circular dichroism spectroscopy in the far-UV spectral region of 190 to 250 nm. At these wavelengths, the chromophore is the peptide bond. The peptide bonds in the protein are asymmetric and absorb in the UV region below 250 nm of the spectra. For an α-helix, intense

electronic absorption band centered at 190 nm due to π→π* transition involved the π-electrons of the carbonyl C=O. The 'W' shaped spectra with troughs around 222 and 208 nm indicate the presence of α-helical structures.

a. Effect of pH on the Secondary Structure by Far UV CD

Far-UV CD spectra of OfurPBP2 at pH 6.5, 5.5, and 4.5 (Figure.2.9) had the characteristics of a typical alpha-helical protein with two CD minima, one centering around 208–209 nm and the second around 222–225 nm, and maxima at 195 nm. At pH 6.5 and 5.5, the secondary structure of OfurPBP2 was similar, suggesting that pH does not affect the protein structure much at these pH levels. At pH 6.5, the secondary structure of OfurPBP2 has 46 % α-helix content. The α-helix content decreases substantially to 43 %, and 37 % when pH is decreased to 5.5 and 4.5, respectively (Table 2.5). The CD spectrum recorded at pH 4.5 is quite different compared to pH 6.5 and 5.5. The decrease in helical content when lowering the pH indicates a significant change in protein structure at acidic pH. At low pH, hydrophobic interaction is less favorable, making the protein likely to unfold.

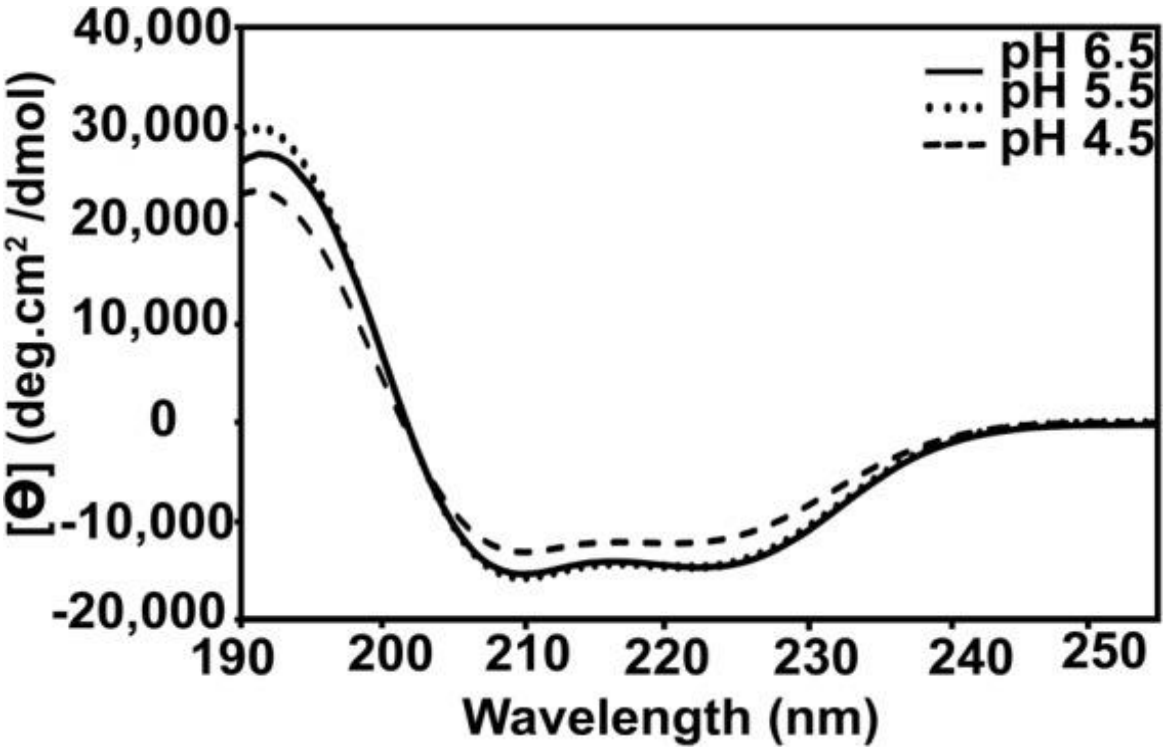


Figure 2.9: A) Circular dichroism (CD) spectroscopic analysis of the OfurPBP2 at room temperature. Far UV-CD spectroscopic analysis of OfurPBP2 in 20 mM sodium phosphate buffer at pH 6.5, pH 5.5, and pH 4.5. The protein concentrations were 30 μM. Characteristic minima at 208 nm and 222 nm at all pH levels are indicative of a highly helical protein**.**

Table 2.5: Percentage of helical content

Protein	**pH 6.5**	**pH 5.5**	**pH 4.5**
Helix (%)	46	43	37

b. **Effect of Temperature on Circular Dichroism Spectra of OfurPBP2**

Circular dichroism (CD) is used to monitor thermal stability by recording the spectrum as a function of temperature. The two negative peaks at 222 nm and 208 nm and one positive peak with a maximum at 195 nm were observed at pH 6.5, which are characteristic of a helical protein. Increasing temperature resulted in a decrease in the magnitude of the peak, and a slight shift toward the higher wavelength in the positive peak (Figure 2.10), indicating a loss of secondary structure loss. At pH 6.5, θ_{222} and θ_{208} decreased by 38 % and 26 %, respectively, at the melting temperature of 90 °C, while the values decreased by 70 % and 52% at 106 °C. The decrease in the percentage of ellipticity at pH 4.5 is comparable with pH 6.5; θ_{222} and θ_{208} reductions by 31% and 25% at the melting point. There is a more substantial reduction in the positive peak at 195 nm at pH 4.5 than at pH 6.5, suggesting OfurPBP2 unfolds more rapidly at the lower pH.

With the increase in temperature, there is a gradual decrease in two negative absorption bands. The protein slowly changes to predominantly random coil form, indicating a transition from ordered secondary structures to the unfolded state. An isodichroic point observed at 203 nm supports the two-state nature of the unfolding.[149,150] The heat-induced denaturation of OfurPBP2 follows a two-stage mechanism. The observation of an isodichroic point, suggesting a cooperative two-state helix-to-coil transition model. This happens when the structure unfolds and becomes disordered. The peak will depress or even disappear if the protein aggregates.

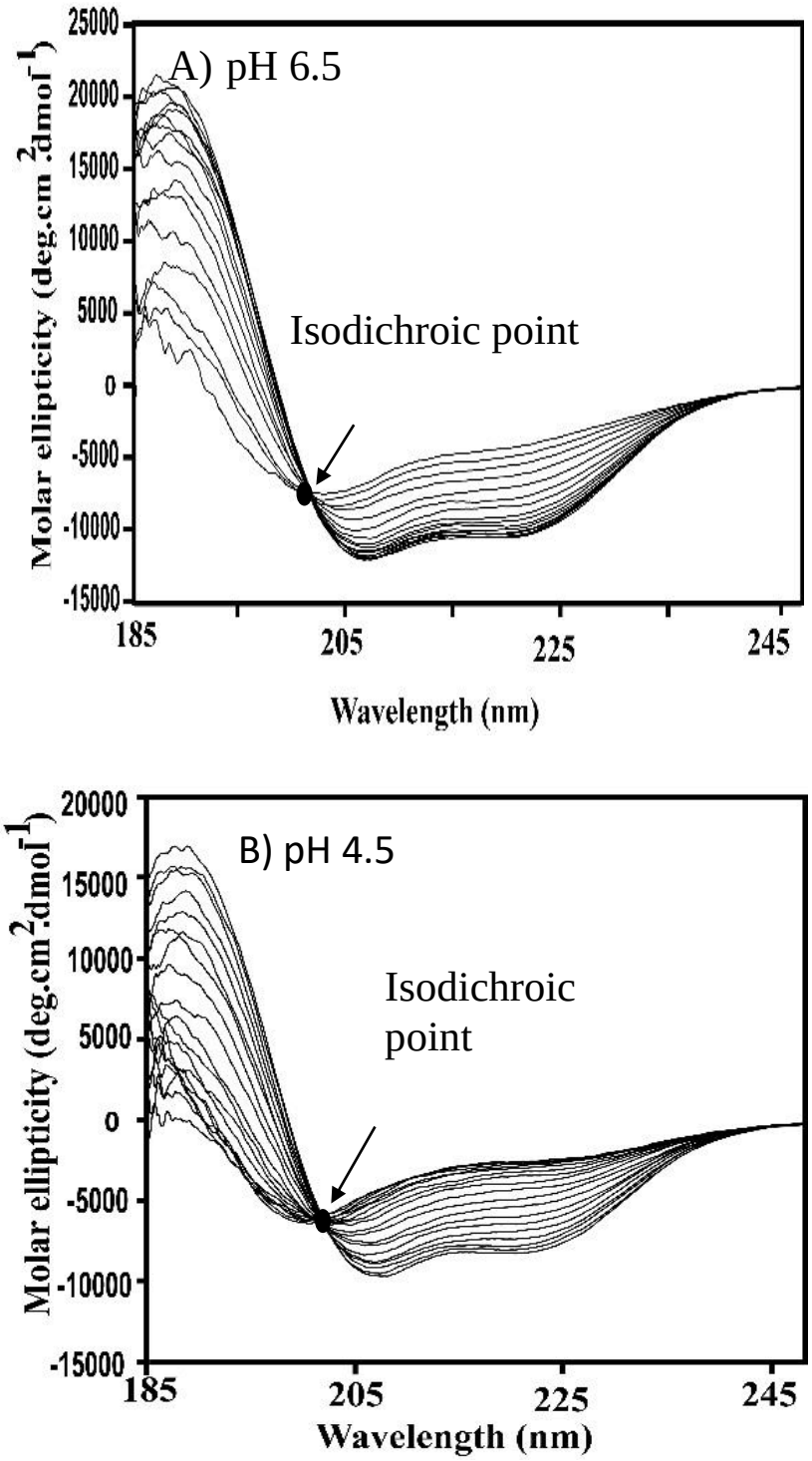


Figure 2.10: Thermal stability of OfurPBP2. Far-UV CD spectra of OfurPBP2 were collected at increasing temperatures from 25 °C to 110 °C. The CD spectrum of OfurPBP2 at pH 6.5 and 4.5 as a function of temperature. A) The CD spectra were overlaid at different temperatures from 25

°C to 110 °C at pH 6.5. B) The CD spectra were overlaid at different temperatures from 25 °C to 110 °C at pH 4.5. The temperature decreases from bottom to up; the spectra are obtained at successive 2 °C intervals**.**

c. **Unfolding of a Protein as a Function of Temperature**

Thermal melting curves monitored the unfolding of the protein as a function of temperature. The unfolding profile of the protein was measured at a wavelength of 222 nm. The melting curves were obtained by plotting the elliptical values at 222 nm for each spectrum against the increasing temperature. The thermal denaturation curve of OfurPBP2 at 222 nm showed a gradual loss of CD signal starting at 77 °C and continuing to 97 °C at both pH conditions. No further significant changes were observed at the higher temperature, indicating the protein achieved a denaturation state, which can be seen from the curve reaching the plateau. From the "S" shape curve (also called the denaturation curve), the melting temperatures obtained were 90.47 °C and 87.01 °C at pH 6.5 and 4.5, respectively (Figure 2.11). While comparing the two pHs, the gradient of unfolding was less for the protein at pH 4.5. The slope of the curve explained the co-cooperativity of folding and denaturation. The steeper the curve indicates the significant cooperativity, while a shallower curve suggests that some parts of the protein will likely denature while other parts are still folded.[150,151]

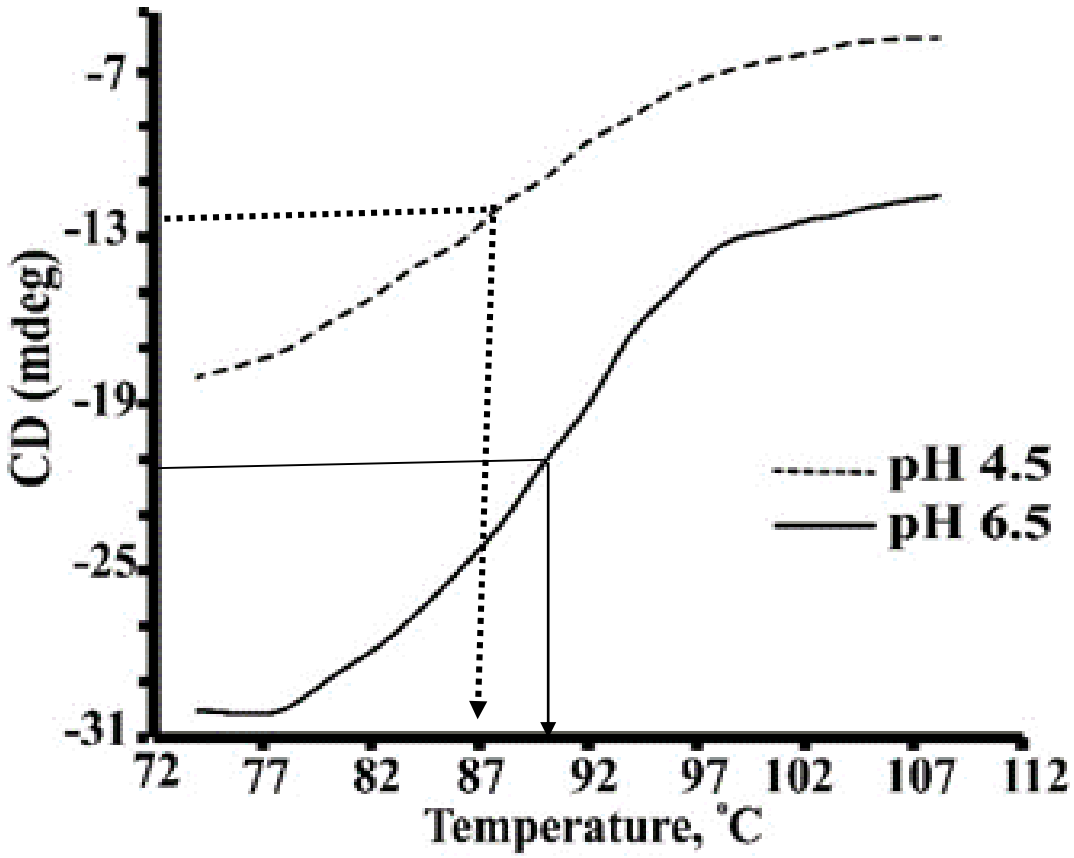


Figure 2.11: S shape curve comparing the melting temperature at pH 6.5 (90.47 °C) and 4.5 (87.01 °C).

There are two mathematical models to analyze the unfolding curve. One is a two-state model, reversible, unimolecular, and equilibrium folding, while the other is a three-state model via the formation of an intermediate like molten globule/compact intermediate between native and unfolded state. The folding process of OfurPBP2 was monitored by far-UV CD spectroscopy. The model that can be explained for equilibrium unfolding is given from equation 2.3.

$$N \text{ (native)} \rightleftharpoons D \text{ (unfolded)} \qquad (2.3)$$

Here, the protein possesses a native (N) and a denatured (D) state. The two-state model can explain the observation of an isodichroic point around 203 nm.[151] The unfolding free energy calculated for the globular proteins lies within the range of 20-50 kJ/mol.[152]

The shape of the unfolding curve qualitatively measures cooperativity. A highly cooperative unfolding transition indicates that the protein existed initially as a compact, well-folded structure valid for pH 6.5. The protein at pH 6.5 shows a single sigmoidal transition. OfurPBP2 shows a two-state unfolding behavior, similar behavior was observed on ubiquitin-ubiquitin interacting motif (UIM).[153] The sigmoidal curve with a smaller slope at pH 4.5 suggests that the protein behaves non-cooperatively.[154] It indicates that the protein existed initially as a very flexible, partially unfolded protein. This indicates less cooperativity, which is due to a non-compact structure.

The thermodynamic parameters for thermal folding equilibria indicate that the protein is thermostable with a ΔG of 49.216 KJ/mol and 25.210 KJ/mol for pH 6.5 and 4.5 respectively (Table 2.60. There is complete reversibility. At low temperatures, $T\Delta S < \Delta H$, means that ΔG is positive and unfolding is not spontaneous. As we increase the temperature, we will eventually get to a point where $T\Delta S > \Delta H$, where ΔG will be negative and ΔH is positive, the unfolding of the protein is spontaneous at high temperature ($T\Delta S > \Delta H$), ΔS is positive. The cooperativity of unfolding reaction can be also measured qualitatively by the shape of the unfolding curve. The appearance of sigmoidal melting curves, indicative of cooperative thermal unfolding (Figure 2.12). A highly cooperative unfolding transition indicates that the protein existed initially as a compact, well-folded structure. The slope of the curve tells about the cooperativity of folding and denaturation. The steeper the curve indicates greater cooperativity of the process, while a shallower curve indicates that some parts of the protein will likely denature while other parts are still folded. CD data conclude that the protein has high structural stability.

Table 2.6: Thermodynamic parameters of OfurPBP2 obtained from the circular dichroism at two different pH

Protein	Melting temp T_m(°C)	ΔH J/mol	ΔS J/mol/K	Free energy of folded states (KJ/mol) @ Room Temperature	K_{eq}
pH 6.5	90.47	272739	750.076	49.216	2.36E-09
pH 4.5	87.01	146066	405.554	25.210	3.81E-05

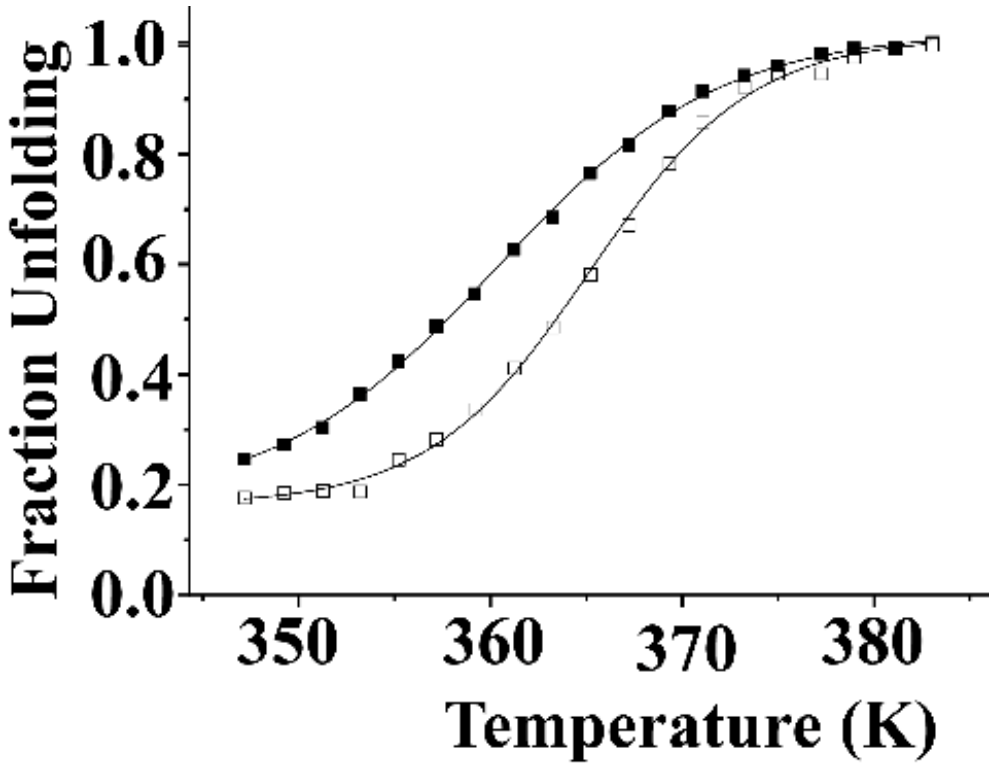


Figure 2.12:The fraction of unfolding from far-UV CD spectra at 222 nm.

2.3.7. **Characterization of OfurPBP2 by Fluorescence Spectroscopy**

To measures, the binding affinity of OfurPBP2, the fluorescent ligand 1-*N*-phenyl-1-naphthylamine (1-NPN) was used. The binding of 1-NPN to delipidated OfurPBP2 at pH 6.5 was measured by monitoring the increase in the NPN fluorescence at 420 nm. N-phenyl-1-naphthylamine (NPN) is a hydrophobic fluorescent probe. The 1-NPN probe was excited at 337 nm, and emission spectra were collected from 370-600 nm on a fluorescence spectrophotometer. The change in fluorescence intensity at different ligand concentrations was used to calculate the relative fluorescence intensity (F_R) in equation 2.2. To determine the dissociation constants, the intensity values corresponding to the relative fluorescence intensities were plotted against the concentration of free ligand. The dissociation constant, Kd, was determined from a non-linear curve fit of the data. The Kd value was calculated as 1.2 ± 0.07 µM (Figure 2.13 and 2.14). Ligand binding of the fluorescent probe 1-NPN to the OfurPBP2 protein exhibit good affinity for NPN with dissociation constants in the micromolar range.

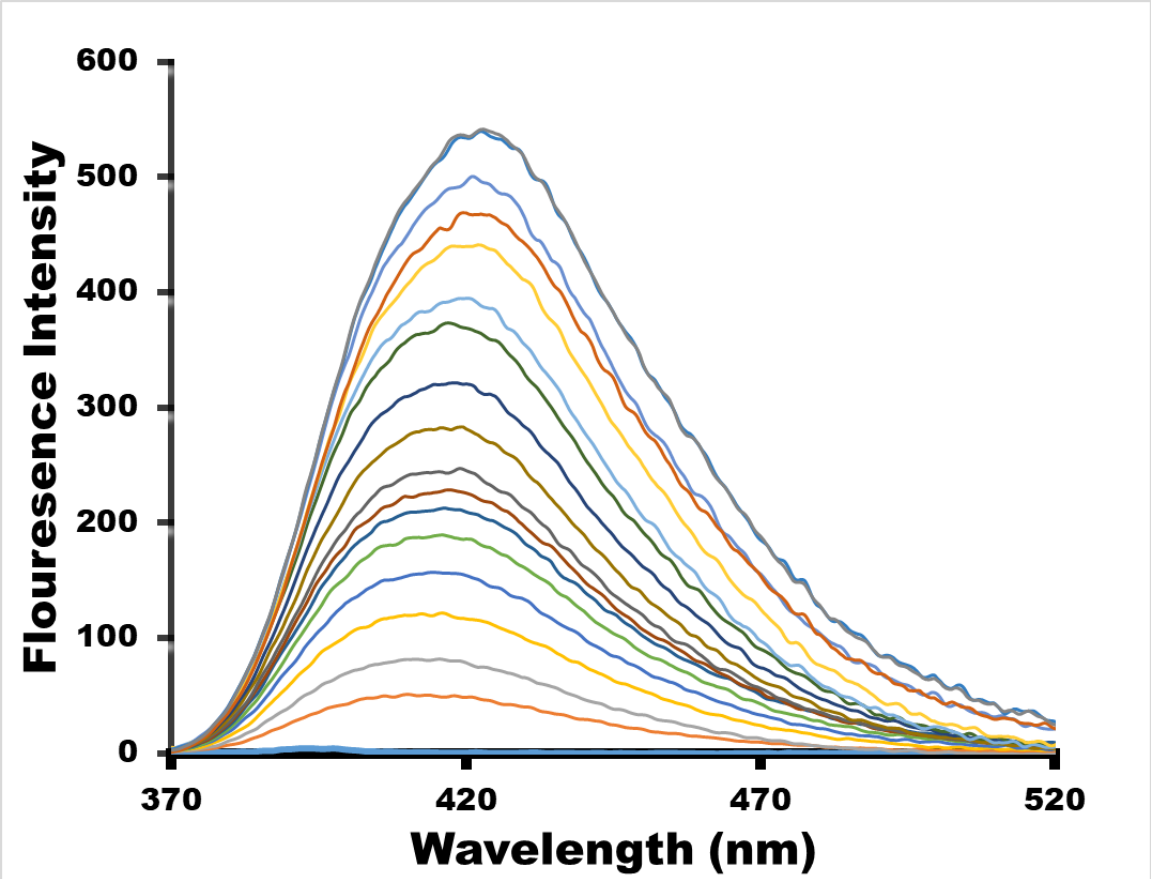


Figure 2.13: Fluorescence spectra of delipidated OfurPBP2. Extrinsic NPN binding fluorescence spectra. The protein concentration was 1 µM in 20 mM phosphate buffer at pH 6.5. Fluorescence emission spectra of OfurPBP2 protein upon addition of different concentrations of NPN.

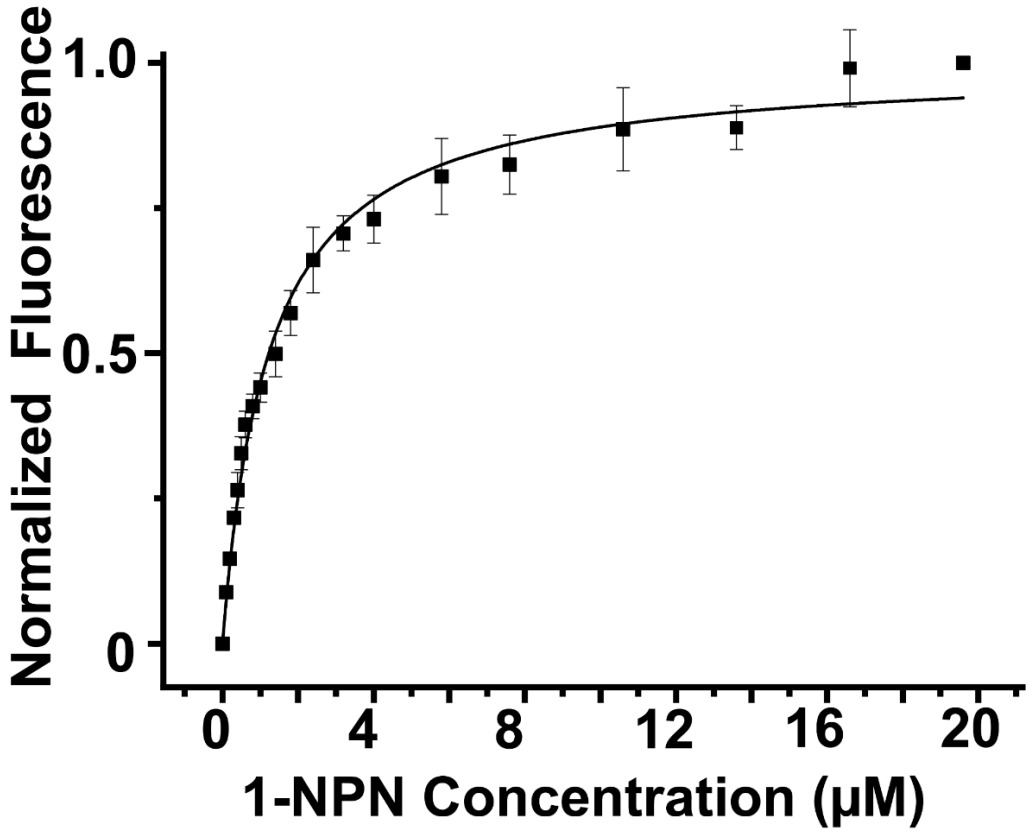


Figure 2.14:The increase in fluorescence intensity was measured at 420 nm. The normalized fluorescence intensity (F^R) plot to 1- NPN concentration (in μM) was used to calculate the Kd value.

During the expression of lipid-binding proteins, the protein picks up a hydrophobic molecule from the host bacterial system. OfurPBP2 was expressed bound to an endogenous hydrophobic ligand from the bacteria. The endogenous lipids were removed from the protein with hydrophobic

interaction column chromatography (HIC). This process is called delipidation. These hydrophobic compounds bind to Lipidex-1000 resin as the protein passes through the column. Ligand-binding assays were carried out with the free OfurPBP2 after the removal of the bacterial ligand through the delipidation process. The N-phenyl-1-naphthylamine (NPN) is a hydrophobic fluorescent probe, typically used to measure the binding affinity and/or probe the hydrophobic pocket/environment of lipid-binding proteins and membranes. The binding of NPN to delipidated OfurPBP2 at pH 6.5 was measured by monitoring the increase in the NPN fluorescence at 420 nm. Competitive displacement of a fluorescent probe with pheromones was performed to determine the binding constants. In the competitive displacement assay using NPN as a fluorescent probe with *Z*12-14: OAc and E12-14: OAc pheromone as the competing ligand. We observed that the delipidated OfurPBP2 with Z12-14: OAc at pH 6.5 has (*Kd* = 33.5 nm) and with *E*12-14: OAc (*Kd* = 47.29 nm) (Figure 2.15). The *E*12-14: OAc pheromone has slightly lower binding affinities than *Z*12-14: OAc pheromone. The *Kd* values at the nanomolar range indicate the strong binding affinity. The binding affinity is similar to that of delipidated ApolPBP1 wt at 6.5 with its 6*E*,11*Z*-hexadecadienyl acetate pheromone (Kd = 50 nm).[65] The previously reported binding affinity of OfurPBP2 by Zhang et al. is in the micromoles range.[155] As they did not delipidate the protein and binding assay were carried in the undelipidated protein. When OfurPBP2 was expressed in *E. coli*, it binds to a hydrophobic ligand that is endogenous to the host cells. The ligand is removed by a delipidation procedure.

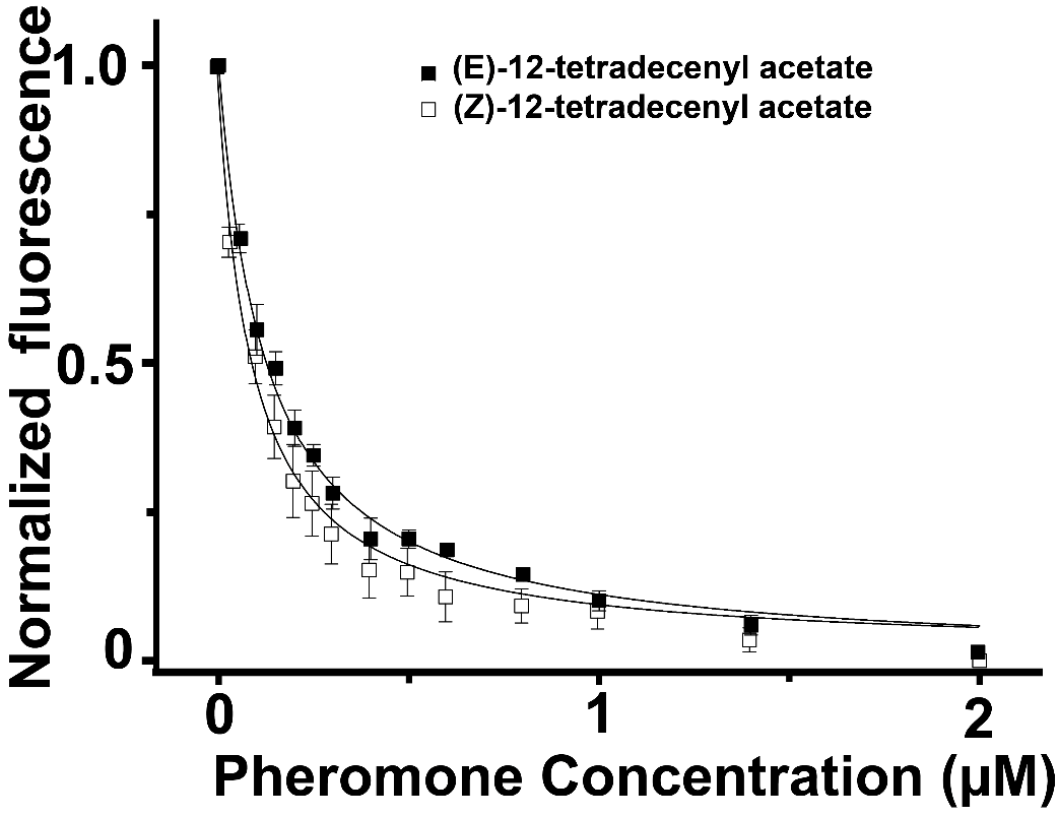


Figure 2.15: Competitive binding of pheromones (Z12-14: OAc and E12-14: OAc) with NPN. The standard deviations are indicated by error bars.

Table 2.7: The dissociation constants of OfurPBP2 with pheromones

Pheromones	IC_{50}	K_i
E-12 tetradecenyl acetate	125 nM	47.29 nM
Z-12 tetradecenyl acetate	88 nM	33.5 nM

2.3.8. **Effect of Temperature by NMR**

To optimize the temperature conditions and investigate the effect of temperature on OfurPBP2, 2D {^{1}H, ^{15}N} HSQC were collected at 298 K, 308 K, and 318K (Figure 2.16). There were no significant changes in the spectra, indicating no significant conformational changes or denaturation. The number of amide resonances at 298 K and 308 K spectrum are equal. However, at 298 K, peaks corresponding to residues; Phe36, Leu68, Asn104, and Arg70 were missing due to line broadening. The reduction of temperature slows down the faster intra- and intermolecular motions, contributing to spin-spin relaxation, which leads to line broadening. As the temperature increased, the correlation time of the protein decreased, and the resonance became narrow. However, at 318 K, some cross-peaks were still missing, likely due to intermediate exchange.

The amide peak intensity of residues Arg46, Asn107, Gly40, and Lys14 decreased or sometimes disappeared. The disappearance of cross-peaks could be due to amide-proton exchange or conformational exchange on the chemical shift time-scale. Along with the increase in temperature, the molecular tumbling rate increases, resulting in sharp and intense signals. The peak intensity of amide involved in the conformational exchange from a µs-to-ms timescale depends upon the relaxation parameter.[156] The relaxation parameter again depends upon the exchange regime, the exchange rate, and the thermodynamic parameters. Although 2D {^{1}H, ^{15}N} HSQC spectra are well dispersed with sharp peaks and less overlap at low and high temperatures, the spectra still suffer from signal loss. The smaller chemical shift change with increasing temperature was due to the thermal fluctuation, which results from the high mobility at the elevated temperatures.[157] NMR sensitivity generally increases with decreasing sample temperature.[158] Similarly, a few of the conformational exchange peaks were also suppressed at low temperatures.[158]

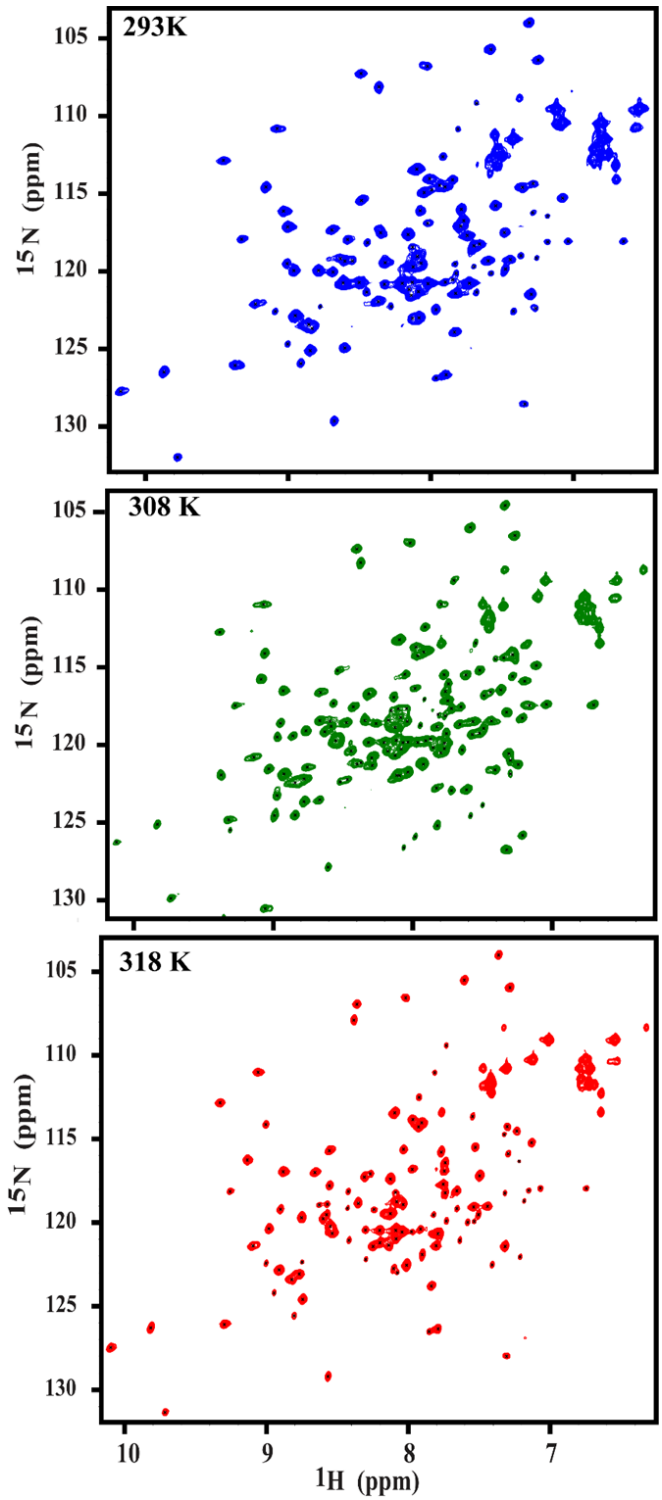


Figure 2.16: The {^{1}H ^{15}N} HSQC spectra collected at 800 MHz at an increasing temperature in the range of 298, 308, and 318 K**.**

2.3.9. **Effect of pH on the Conformation of the OfurPBP2**

NMR was used to characterize isotope-labeled pure recombinant OfurPBP2 protein. The 2D {^{1}H, ^{15}N} heteronuclear single quantum coherence (HSQC) spectrum represents a fingerprint of the protein. HSQC spectra are sensitive to chemical structure and can detect any change in the structure at the level of individual nuclei.[159] Any change in protein structure or conformation is due to mutation/s or ligand binding or changes in pH, temperature, salt concentration, etc., are reflected in the HSQC spectrum. NMR experiments were performed at pH 6.5, 5.5, and 4.5 to study the effect of pH on OfurPBP2 conformation. The HSQC spectrum of OfurPBP2 at pH 6.5 is well-dispersed, suggesting that the protein is well-folded with a stable tertiary structure (Figure 2.17). Spectra collected at pH 6.5 and 5.5 showed no significant changes in chemical shift values, suggesting that there was no change in protein conformation (Figure 2.17). However, a conformational transition occurred between pH 5.0 and 4.5 as indicated by {^{1}H, ^{15}N} HSQC. The fingerprint region of OfurPBP2 at pH 6.5 and 4.5 is significantly different (Figure 2.18). The amide signals were shifted completely below pH 5.5, with a reduction in peak dispersion, causing more overlap in the center of the HSQC spectrum (Figure 2.18). Furthermore, the quality of spectra degraded at pH 4.5. The center of the spectrum is crowded along with a reduction in chemical shift dispersion. The intensity of the peaks was reduced and peak doubling was observed for many resonances, indicating the presence of more than one conformation at pH 4.5.

The exchange between different conformation, peak broadening, and a poorly dispersed HSQC spectrum are the characteristic features of a molten-globule protein.[160] These molten globules exist in the protein under mild denaturing conditions or when the cofactor or ligand is removed. Acidic pH is one of the factors that push the folded protein into a molten globule state.[161] At low pH, OfurPBP2, adopts the conformation of the partially molten globule. In this partially molten globule state, there is a fluctuation of ensembles on the order of milliseconds to microseconds.[162] Due to the structural fluctuations between different conformational states, there is a substantial broadening

of the NMR signals.[163] Similar conformational characteristics were observed for many proteins in the literature, one of the examples is α-lactalbumin.[163] Due to the overlapping of peaks, it is challenging to infer much structural information at pH 4.5. The protein at this pH is prone to aggregation and occasional precipitation.

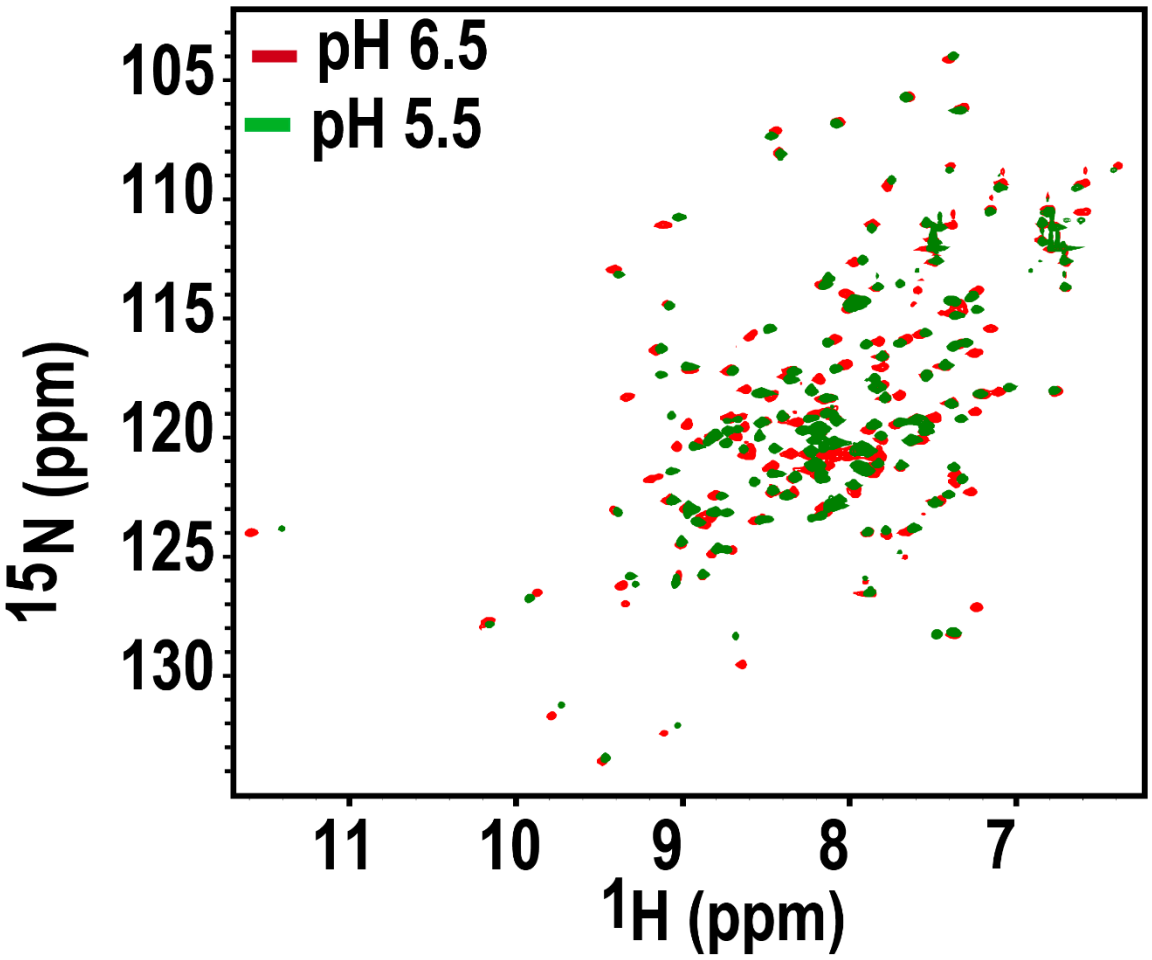


Figure 2.17: Overlay of {^{1}H ^{15}N} HSQC spectra collected at pH 6.5, and 5.5.

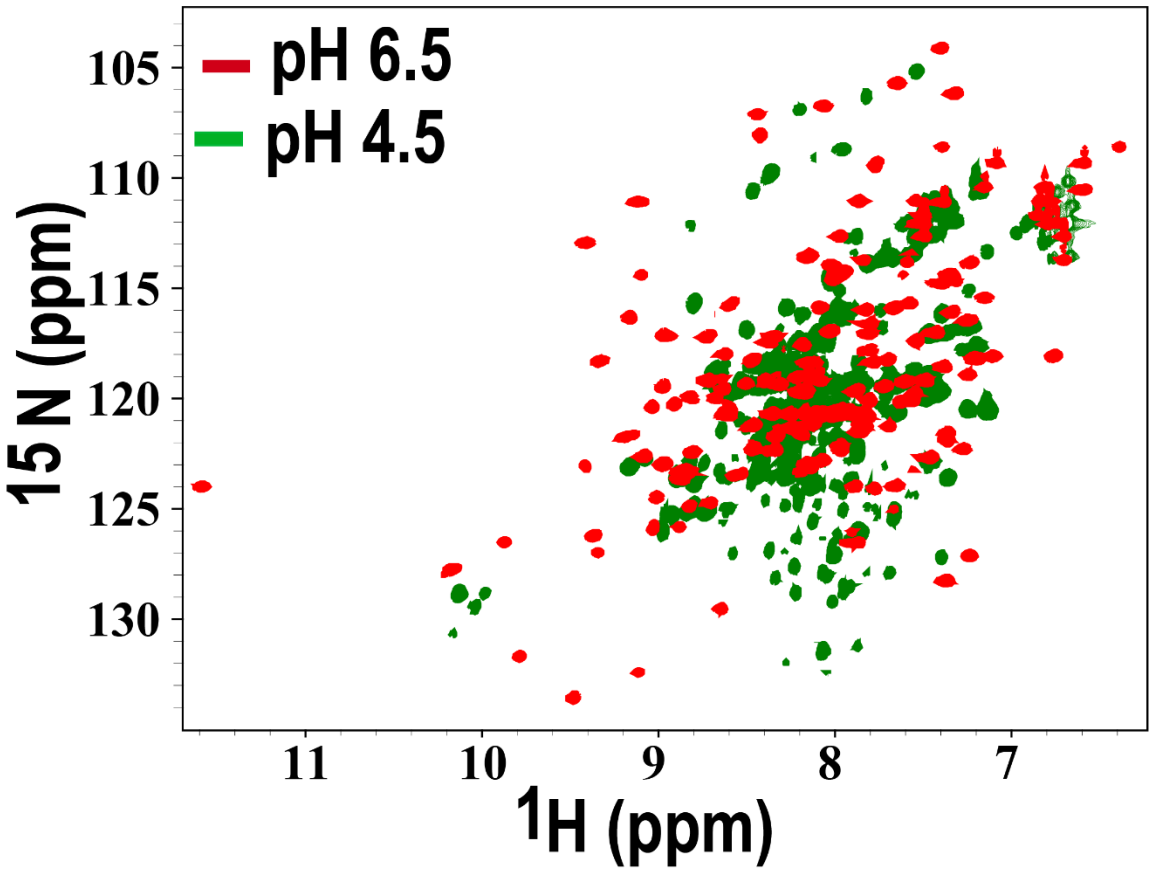


Figure 2.18: Overlay of {^{1}H ^{15}N} HSQC spectra collected at pH 6.5, and 4.5.

To investigate whether OfurPBP2 is denatured at pH 4.5, the pH level was again reversed to 6.5. The HSQC spectrum obtained after raising the pH matched the original spectrum (Figure. 2.19) taken at pH 6.5, suggesting no acid-induced denaturation in OfurPBP2, and the conformational heterogeneity at pH 4.5 was reversible. Although pH titration was reversible, it was not similar to other PBPs. [51,62,64,65,76,137-139]

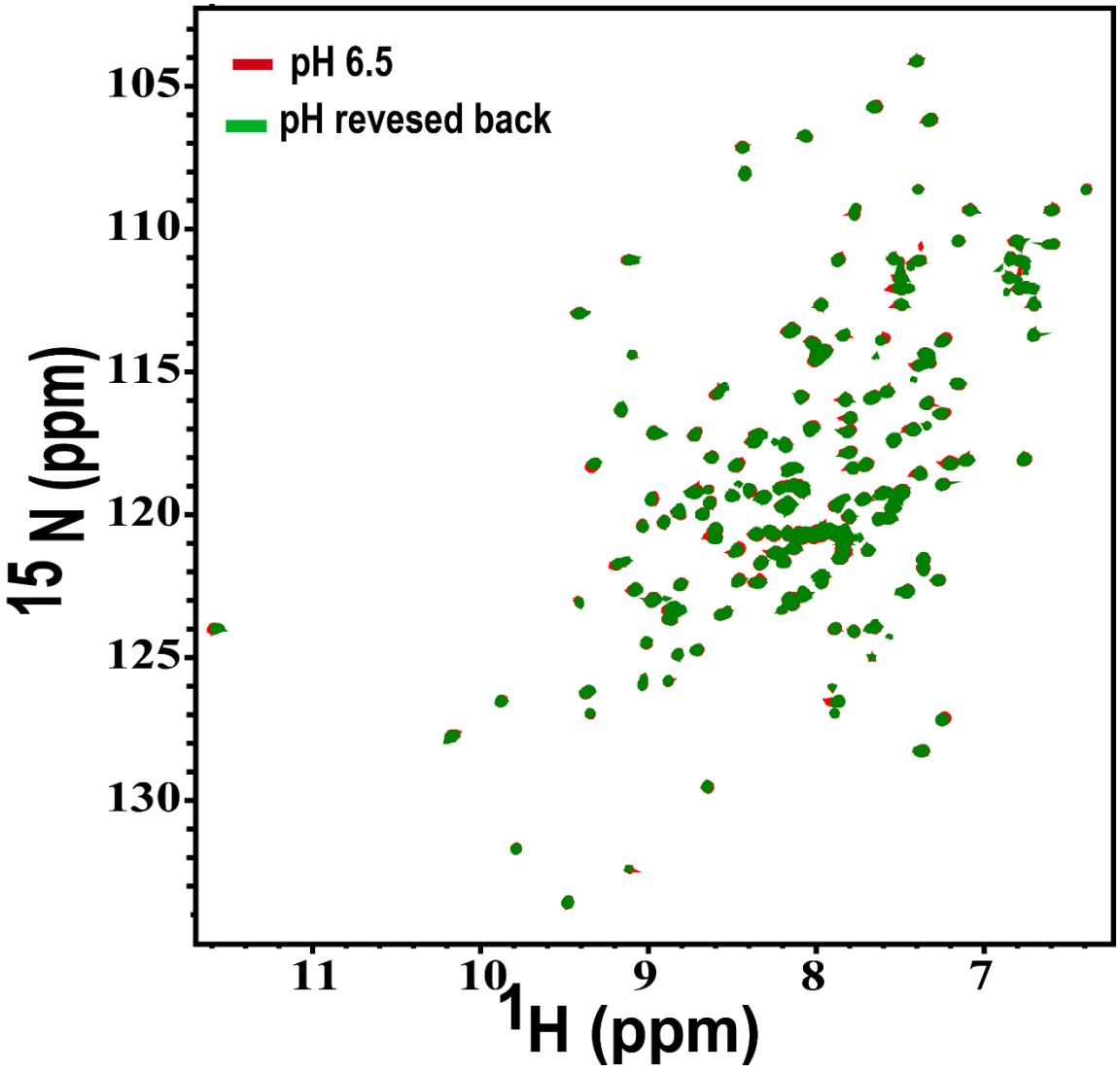


Figure 2.19: Two-dimensional 2D {^{1}H, ^{15}N} HSQC spectra at pH 6.5 overlaid, showing when pH is reversed back to pH 6.5 after pH 4.5. All the peaks returned to their original position.

Conclusion

We have reported the over-expression of recombinant OfurPBP2. The CD data shows that the protein is highly helical at pH 6.5. The unfolding profile showed that the protein is very stable to high temperature with a melting temperature of 90 °C at pH 6.5 and 87 °C at pH 4.5. Since this protein at pH 4.5 is more flexible, the NMR data can be collected at low temperatures to reduce protein dynamics. Based on the pH titration studies, OfurPBP2 undergoes a reversible pH-dependent conformational change. The circular dichroism (CD) and NMR data show a change in both secondary and tertiary structures at low pH. At neutral pH, the protein is well-folded into a relatively rigid conformation. The protein conformation is quite similar at pH 5.5 and above, without significant change in chemical shift. At pH 5.0 and below, the overall peak dispersion is reduced with crowding of peaks due to the protein flexibility and likely partial unfolding. The protein is likely in a molten globule conformation which could account for the extreme line broadening due to conformational fluctuation on the millisecond to a microsecond time scale. Due to the severe overlapping of peaks and missing peaks, the determination of the structure of the protein at pH 4.5 is very challenging.

The behavior of OfurPBP2 at acidic pH of 4.5 is in stark contrast to the current model of pheromone uptake and releases by several well-studied Lepidopteran PBPs[51,62,64,65,76,137-139], including ApolPBP1, BmorPBP, AtraPBP1, and LdisPBP2. In ApolPBP1, the ligand-bound protein is primarily in a PBP^B (bound) conformation above pH 6.0. It is a mixture of PBP^B and PBP^A (bound and free) conformations between pH 6.0–5.0, while primarily in a PBP^A (free) conformation at pH below 5.0. Thus, at pH 4.5, the PBP^A (free) conformation is predominantly present.[65] Similar phenomena have been observed for BmorPBP[63,72,73,164,165], AtraPBP1[83,166], and LdisPBP2.[51] However, in OfurPBP2, the HSQC data quality gradually degrades starting at pH 5.0 and below,

with resonances crowding at the center of the spectrum, unlike the PBPs mentioned above. The pH titration studies by NMR indicated that OfurPBP2 does not behave like other well-studied Lepidopteran PBPs, including ApolPBP1[65,76,138,139], BmorPBP[63,73,164,165,167] AtraPBP1[83,166], and LdisPBPs.[51] Based on the pH-titration data, we hypothesize that OfurPBP2 may release pheromone using a novel mechanism. Further investigation of structure and function is necessary to gain insight into the mechanism of pheromone communication in *Ostrinia furnacalis*.

CHAPTER III

NMR RESONANCE ASSIGNMENTS AND SECONDARY STRUCTURE OF THE *OSTRINIA FURNACALIS* PHEROMONE BINDING PROTEIN 2 (OfurPBP2)

3.1 Introduction

The detailed structural and mechanistic studies of PBPs of *Bombyx mori* [59-62,64,69,168,169], *Antheraea polyphemus*[65,76,80,82], *Amyelois transitella*[83,84], and *Lymantria dispar* [51,137,170] showed the proteins consist of 6 tightly-folded helices enclosing a large hydrophobic pocket with the unstructured C-terminus outside. Although OfurPBP2 has over 50% sequence similarity with the well-studied lepidopteran PBPs mentioned above, there are significant differences in the two biologically important gates: An Arg replaces the His70 in the histidine gate[65,76], and the C-terminus of OfurPBP2 has four additional charged residues. A detailed structural characterization by solution NMR analysis is needed to understand the effect of these critical substitutions on the structure and function of OfurPBP2.

NMR spectroscopy is one of the most powerful tools for determining the structure and function of molecules. The chemical shifts are exquisitely sensitive probes to obtain detailed atomic properties of macromolecules, such as the secondary and tertiary structure of the protein. The protein's fingerprint region suggests that the chemical shifts inherently carry enough information.

to determine such structures at high resolution. Atomic-resolution structure determination is crucial for understanding the protein-ligand interaction at the molecular level.

NMR occupies a unique niche in the biophysical analysis of proteins because of its ability to identify binding sites, affinities and ligand pose at the level of individual amino acids. It also provides structures of flexible regions that often fail to crystallize. The assignment of resonances in the complex NMR spectrum of a protein is the first step in studying protein structure, function, and dynamic. Structure determination by NMR spectroscopy usually consists of several sequential steps. These include ^{13}C, ^{15}N double/ ^{2}H, ^{13}C, ^{15}N-triple labeled sample preparation, NMR data acquisition, data processing, assignment (backbone, side-chain, and NOE assignment), restraints generation, incorporation of NOE information, dihedral angles, structure elucidation, structure, energy minimization, and structure validation. Among these, the resonance assignments are usually a laborious, most time-consuming, and daunting task. The assignment of the protein backbone is the first step of resonance assignments. The assignment is cross-verified by identifying the side chains on protein and the carbons by TOCSY and NOESY experiments. NOE peak assignment is vital for structure determination, as it serves as the primary source of structural constraints for structure calculation. The NOE cross-peaks were assigned with the help of backbone and side-chain resonances. The assignments for undelipidated OfurPBP2 at pH 6.5 were 97% completed for backbone and 88% completed for side-chain resonances. The NMR spectrum of the delipidated OfurPBP2 exhibited more signals than the expected number, and the dispersion of peaks is poor. This might be due to the presence of multiple conformations. One set of resonances could be assigned at a higher counter level, other set could not be possible to assign because of heavy overlapping and very broad peaks or even low intensities. Extra peaks were observed at the lower counter level, reflecting a mixture of other conformation of OfurPBP2 under the current experimental condition. One set of resonances could be assigned at a higher counter level and approximately 85% backbone and 80% side-chain resonances were assigned.

To understand the molecular basis of OfurPBP2, we have initiated solution structural studies using multidimensional heteronuclear NMR spectroscopy. Herein, we report the assignments of the secondary structure of OfurPBP2. The results show that the protein is mainly composed of seven helices, connected by turns or loops. The helical regions are strongly suggested by continuous sections of intense sequential dNN connectivities and are confirmed by the observation of medium-range NOEs. These assignments provide the starting point for the determination of the tertiary structure. This chapter deals with the step-by-step procedures for backbone and side-chain assignment and the process of secondary structure calculation using the backbone chemical shift by using TALOS+[171], CSI (Chemical Shift Index), and SSP[172] for both undelipidated and delipidated OfurPBP2.

3.2. Methods and Materials

a. Protein Sample Preparation

The isotopically labeled (^{15}N and $^{15}N/^{13}C$), recombinant OfurPBP2 was expressed in *E. coli* origami 2 cells. The recombinant proteins were refolded and purified by dialysis, anion exchange DEAE-chromatography, and size exclusion chromatography using a Superdex 75 column fitted to ÄKTA FPLC (GE Healthcare) as described previously[173]. The SDS-PAGE confirmed the purity of the sample. NMR samples used for the structure determination contained 0.4 mM uniformly ^{15}N- and $^{15}N/^{13}C$-labeled OfurPBP2 (95% H_2O/5% D_2O) in 50 mM of phosphate buffer (pH 6.5) containing 1 mM EDTA and 0.01% NaN_3.

The details of the delipidation of OfurPBP2 were explained in Chapter 2. The delipidation of OfurPBP2 was performed by Dr. Mohanty, by modifying the original protocol mentioned by Bette *et al.*[143] and Katre *et al.*[65] We have optimized the protocol by modifying temperature, incubation time, time of shaking, and also the volume of the Lipidex resin. Briefly, the protein was buffer-

exchanged to 50 mM sodium citrate buffer at pH 4.5 (buffer A) and concentrated to 0.8 mL using a Millipore ultrafiltration concentrator with a molecular weight cut-off of 3,000 Da. The 15 mL of Lipidex ™-1000 resin was manually packed and washed 15-20 times with water to remove the residual methanol and equilibrated with buffer, and then equilibrated with a citrate buffer. The protein was loaded in the Lipidex column and incubated at 37℃ for 30 minutes, and then was eluted manually under gravity until the absorbance at A_{280} was negligible. The eluted protein was concentrated to 1 mL and was buffer exchanged to 50 mM sodium phosphate buffer at pH 6.5 with 1 mM EDTA, 0.01% sodium azide, and 10% D_2O. Protein concentrations were determined spectrophotometrically using the theoretical extinction coefficient, A_{280} of 15845 $M^{-1}cm^{-1}$.

b. NMR Data Collection

NMR samples contained 0.4 mM uniformly $^{15}N/^{13}C$-labeled OfurPBP2 (95% H_2O/5% D_2O) in 50 mM phosphate buffer (pH 6.5) containing 1 mM EDTA and 0.01% NaN_3. All NMR data were collected at 35 ^0C on a Bruker AVANCE 800 MHz spectrometer equipped with a cryogenic triple resonance probe at the National High Magnetic Field Laboratory (NHMFL) Tallahassee, FL, and Oklahoma State University. The 2D {^1H, ^{15}N} HSQC spectrum was collected with 256 increments in the ^{15}N dimension and 2048 complex points in the ^1H dimension. For the sequential assignment of ^1HN, $^1H_\alpha$, ^{15}N, $^{13}C_\alpha$, $^{13}C_\beta$, and ^{13}CO resonances: following experiments were used: 2D {^1H, ^{15}N} HSQC, 2D {^1H, ^{13}C }-HSQC, 3D HNCA, 3D HN(CO)CA, 3D HNCO, 3D HN(CA)CO, 3D HNCACB, 3D CACB(CO)NH. The side-chain assignments were carried with 3D CC(CO)NH, 3D H(CCCO)NH, 3D HCCH-TOCSY, 3D ^{15}N-edited HSQC-TOCSY experiments. For the NOE distance restraints 3D ^{15}N-edited HSQC-NOESY with mixing times of 85 ms, aliphatic ^{13}C-edited HSQC-NOESY, and aromatic ^{13}C-edited HSQC-NOESY with mixing times of 120 ms were used. The ^{13}C carrier frequency in the aliphatic (44 ppm) and aromatic (125 ppm) regions were collected.

The list of the standard experiments needed for the structure determination by NMR are listed in Figure 3.1. The detail for parameters used in NMR experiments is listed in Table 3.1.

Table 3.1: Parameters used in NMR experiments for undelipidated OfurPBP2

Spectrum	Nuclei	Data size (Complex points) $^{15}N\times^{13}C\times^{1}H$	Sweep width(SW) (ppm) $^{15}N\times^{13}C\times^{1}H$	Carrier frequency $^{15}N\times^{13}C\times^{1}H$	Number of scan (NS)
2D {^{15}N ^{1}H } HSQC	^{15}N, ^{1}H	256×2048	36×16 ($^{15}N\times^{1}H$)	119×4.7 ($^{15}N\times^{1}H$)	32
2D {^{13}C ^{1}H }HSQC	^{13}N, ^{1}H	256 × 2048	70×16 ($^{13}C\times$ ^{1}H)	45 ×4.7 ($^{15}N\times^{13}C\times^{1}H$)	8
3D HNCACB	^{15}N, ^{13}C, ^{1}H	40×128×2048	36×70×16 ($^{15}N\times^{13}C\times^{1}H$)	119×45×4.7 ($^{15}N\times^{13}C\times^{1}H$)	32
3D CBCACONH	^{15}N, ^{13}C, ^{1}H	40×128×2048	36×70×16	119×45×4.7 ($^{15}N\times^{13}C\times^{1}H$)	16
3D HNCA	^{15}N, ^{13}C, ^{1}H	48×128×2048	36×70×16	119×45×4.7 ($^{15}N\times^{13}C\times^{1}H$)	8
3D HN(CO)CA	^{15}N, ^{13}C, ^{1}H	32×128×2048	36×70×16	119×45×4.7 ($^{15}N\times^{13}C\times^{1}H$)	8
3D HNCO	^{15}N, ^{13}C, ^{1}H	40×128×2048	36×70×16	119×45×4.7 ($^{15}N\times^{13}C\times^{1}H$)	8
3D HN(CA)CO	^{15}N, ^{13}C, ^{1}H	40×128×2048	36×70×16	119×45×4.7 ($^{15}N\times^{13}C\times^{1}H$)	8
3D-TOCSY-HSQC	^{15}N, ^{1}H, ^{1}H	40×160×2048	36×16×16	119×4.7×4.7 ($^{15}N\times^{1}H\times^{1}H$)	32
3D-HCCH-TOCSY,	^{13}C, ^{13}C, ^{1}H	128×220×2048	36×36×16	45×45×4.7 ($^{13}C\times^{13}C\times^{1}H$)	2
3D-CCCONH	^{15}N, ^{13}C, ^{1}H	140×256×2048	36×70×16	119×45×4.7 ($^{15}N\times^{13}C\times^{1}H$)	4
3D-H(CCCO)NH	^{15}N, ^{1}H, ^{1}H	96×170×2048	36×16×16	119×4.7×4.7 ($^{15}N\times^{1}H\times^{1}H$)	4
3D- ^{15}N NOESY-HSQC Mixing time 85 ms	^{15}N, ^{1}H, ^{1}H	128×256×2048	70×16×16	45×4.7×4.7 ($^{15}N\times^{1}H\times^{1}H$)	32
3D-^{13}C-NOESY-HSQC (Aliphatic) mixing time 110ms	^{13}C, ^{1}H, ^{1}H	128×200×2048	70×16×16	45×4.7×4.7 ($^{13}C\times^{1}H\times^{1}H$)	24

3D-^{13}C-NOESY-HSQC (Aromatic) Carrier frequency 125ppm, and mixing time 110ms	^{13}C, ^{1}H, ^{1}H	128×200×2048	70×16×16	125×4.7×4.7 (^{13}C×^{1}H×^{1}H	24

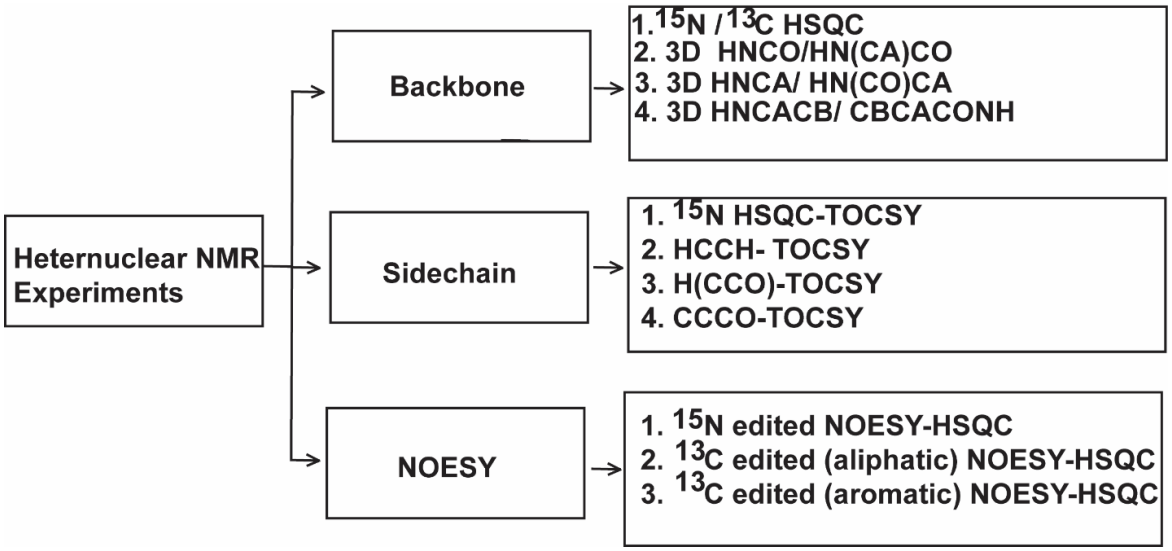


Figure 3.1: The list of standard NMR experiments needed to calculate the 3D structure of the protein.

c. Data Processing

There are standardized conversion scripts to convert the time-domain data into the frequency domain using NMRpipe.[174] Conversion of the Bruker data to NMRpipe format was accomplished

with the bruk2pipe script, which is incorporated on NMRpipe. The processing scripts were executed in the UNIX terminal. The residual water signal was minimized by time-domain deconvolution. The direct ^{1}H dimensions were zero-filled to 1024 complex points. For the 3D data, where the indirect dimensions were collected with only a few points, the number of points and digital resolutions were increased by using linear prediction before apodization with a sine bell window function and Fourier transformation. A window function is generally applied to reduce the artifacts caused by the incomplete sampling of the decaying NMR signal. NMRdraw was used to set and adjusting phase values. The frequency-domain spectra were converted to the sparky format. Sparky was used to visualize and analyze NMR spectra.

d. Sequential Assignment

1. Assessing the Quality from 1D

Although SDS-PAGE and SEC show monomeric protein, high concentration could lead to aggregation/precipitation in the NMR tube, resulting in signal loss. The 1D proton experiment is routinely used to detect protein signals and also the signals from the additive present in the buffer. The information on aggregation and oligomerization can be obtained by using the dilution test while collecting 1D or 2D {^{1}H, ^{15}N} HSQC, and signal intensities as a function of the protein concentration, which is monitored. The non-linear relationship between the protein concentration versus signal amplitudes indicates (transient) oligomerization/aggregation at a higher concentration or an increase in the solution viscosity.[175] The 1D spectra are routinely used to assess the quality of the protein.

2. 2D {^{15}N-^{1}H}-HSQC

The 2D {^{1}H, ^{15}N} HSQC is the starting point for resonance assignments. It is the fingerprint region of the protein. Assessing the quality of 2D {^{1}H, ^{15}N} HSQC is essential before collecting other data sets. Suppose more than 10% of the expected cross-peaks are missing. In that case, precautions should be taken, and the reason for the absence of the peak should be investigated to overcome the problem before continuing with the assignment process.

3. Spin System Numbering

The first step in defining spin systems consists of peak picking all H/N signals in a 2D {^{1}H, ^{15}N} HSQC spectrum. The corresponding peaks, or {H, N} systems, will then be used as a basis for seeking sequential correlations. Here, the letter 'i' represents the reference residue on the spin system, and i-1 and i+1 represent preceding and succeeding amino acid residues, respectively (Figure 3.2). For residues like alanine and serine/threonine, complete spin system identification is straightforward. These residues have a unique chemical shift used as a checkpoint. The HNCACB and CBCA(CO)NH pair can be used to identify some of the amino acid residue types or narrow down the possibilities via 13 Cβ chemical shifts. For example, alanine, serine, and threonine have a Cβ of ~ 20 ppm, ~ 63 ppm, and ~ 70 ppm, respectively, while glycine has no Cβ with a Cα of ~ 45 ppm. Other residues like asparagine, aspartate, phenylalanine, and tyrosine have characteristic chemical shifts where their Cβ appearing around 40 ppm. The isoleucine and valine can be identified by their high Cα chemical shift values lie around 62-63 ppm. For other residues, including arginine and lysine, complete spin system identification is more difficult because of high chemical shift degeneracy.

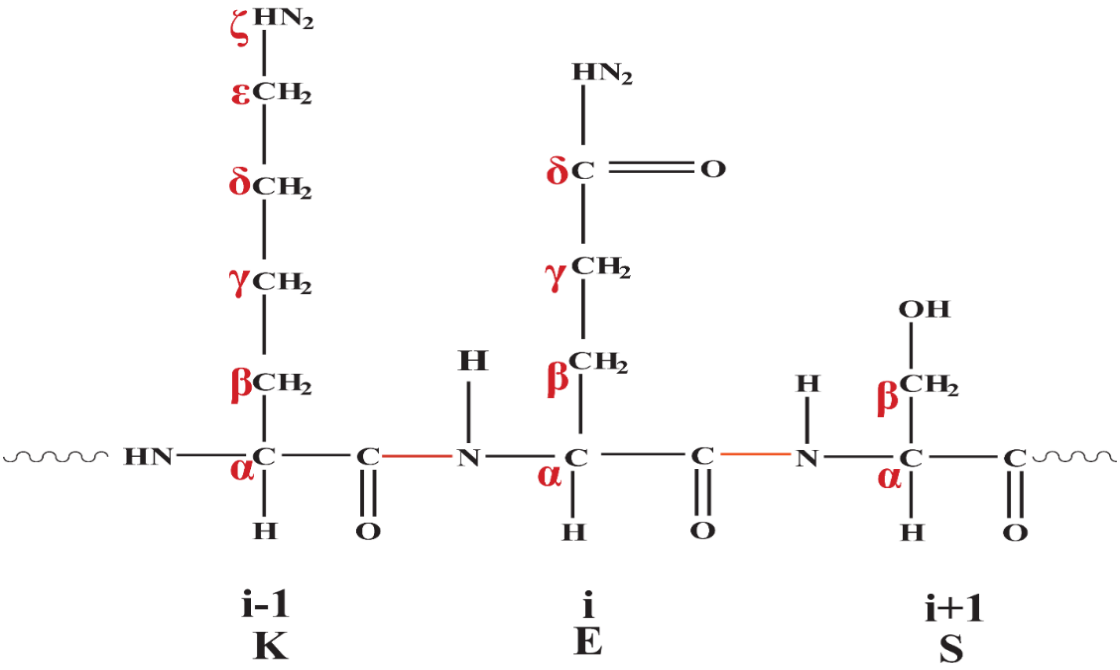


Figure 3.2: Three spin systems in the primary protein structure**.**

4. Backbone and side-chain assignment using 3D experimental data sets

The resonances assignment on the 2D {^{1}H, ^{15}N} HSQC needed six different triple resonance spectra. These six backbone experiments are divided into three complementary pairs: HNCACB and CBCACONH, HNCA and HN(CO)CA, and HN(CA)CO and HNCO. A detailed description of magnetization transfer has been discussed in Chapter 1. Similarly, proton and carbon side-chain resonances were assigned with the help of HCCH-TOCSY, (H)CC(H)-TOCSY, ^{15}N-edited HSQC-TOCSY, and CC(CO)NH TOCSY experiments.

3.3. **Results and Discussion**

Introducing NMR active stable isotopes ^{13}C and ^{15}N into the protein has been tremendously useful. The preparation of proteins enriched with two active nuclei is accomplished by heterologous expression of the protein in the *E. coli*, grown in the media where carbon and nitrogen are fully ^{13}C labeled and ^{15}N labeled, respectively. The purity and mass were confirmed by SDS-PAGE (Figure 3.3).

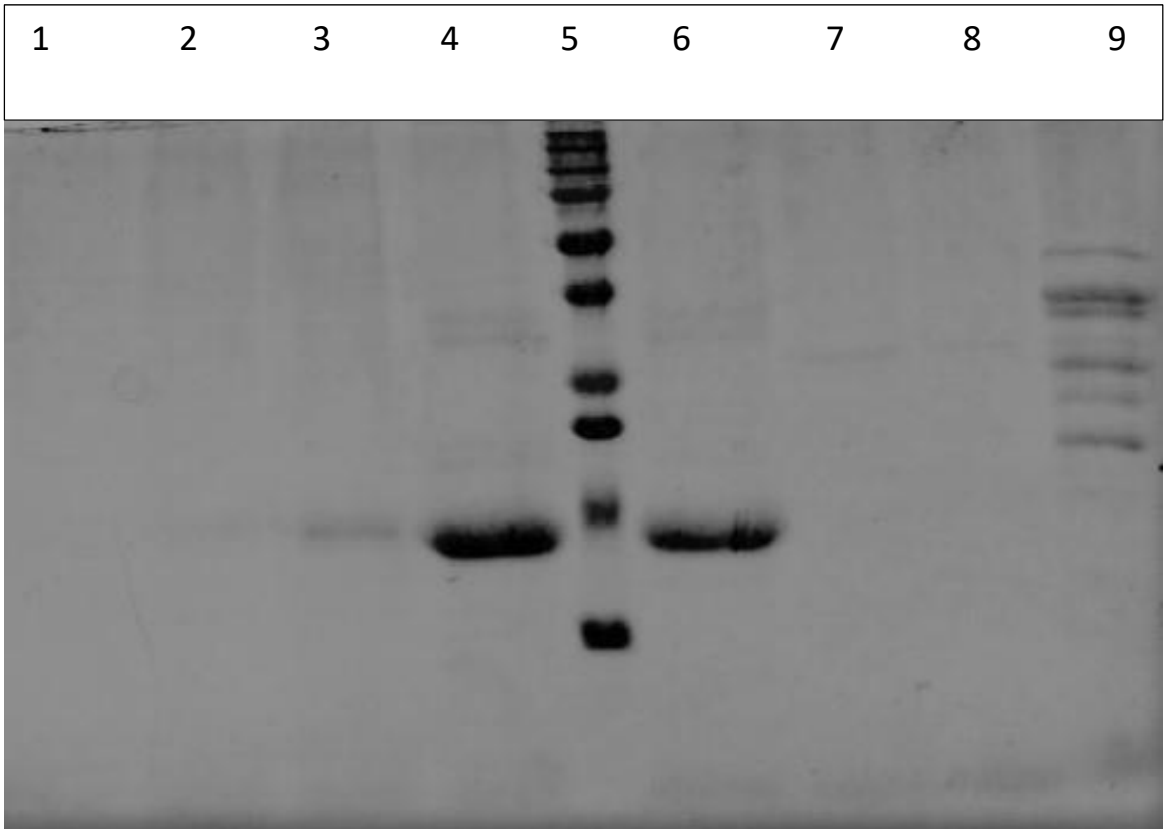


Figure 3.3: The purification profile of double-labeled protein after SEC. Lane 4 and 6 show the pure protein, and Lane 5 is a protein marker.

3.3.1. NMR Data Acquisition

Protein NMR spectra suffer from signal overlapping. Increasing the dimension of NMR spectra resolves the degeneracy problem. Along with the increasing dimensionality of the spectra, experimental time also increases drastically. It takes a couple of days to a week to acquire 3D experiments. Some reasons for the extension of the times include multiple scans, phase cycling, and time between scans (recycle delay). The signal-to-noise ratio (S/N) is directly proportional to the square root of the number of scans (NS). Although a higher magnetic field is needed for better resolution, increasing the number of points in the indirect dimensions is crucial for obtaining highly resolved spectra. For data acquisition, pulse calibrations, spectral parameters, temperature adjustment, deuterium lock, $^{1}H/^{15}N/^{13}C$ tuning, and shimming are done routinely. The standard 2D and 3D pulse sequences can be found in pulse sequence libraries from the NMR spectrometer or otherwise available on the BioMagResBank (BMRB) website. The important parameter like sweep width (SW), number of the experiment (TD), number of scans (NS), dummy scan (DS), carrier frequency, receiver gain value (RG), incremental delay, and mixing time should be optimized according to the type of experiment and pulse program. The ^{1}H carrier frequency was chosen at 4.7 ppm equivalent to the water peak. To optimize the proton sweep width, the 1D ^{1}H spectrum was collected with 70 ppm. The ^{15}N sweep width and ^{15}N carrier frequency were kept at 36 ppm and 119 ppm respectively. The sweep width was optimized to avoid peak folding. Similarly, for ^{13}C, sweep widths were kept at 70 ppm and carrier frequency was kept at 45 ppm. The pH used was 6.5 which is slightly acidic where all the backbone amides were observed in the HSQC spectrum. Generally, when pH is above 7.0, the intrinsic exchange rate of NH may become so fast that some of the peaks are lost.[176] All the data should be collected on a single sample as it reduces the spectral variations. This makes spectral references, peak picking, and assignments convenient. Spectral crowding occurs for proteins with a larger number of residues, intrinsically unfolded proteins, and alpha-helical proteins. If the protein has a flexible region, it can degrade the spectral quality because there is a gradient of evolution times in indirect dimensions for the structured and unstructured

regions.[175] Short evolution time for the unstructured regions leads to the truncation of the free induction decay (FID), resulting in the truncation artifacts (sinc wiggles), which degrade the quality of the spectrum.[175] As the size of the protein increases, the likelihood of repetitions of same/similar amino acid residue/s in the primary sequence increases, resulting in degeneracies of resonances/peaks in the NMR spectra. At higher pH, the intrinsic HN exchange rate may become so fast that the peak will disappear. When the pH is slightly acidic most of the peaks from the backbone amide will be observed.[177] Thus, pH and temperature optimization are very much essential before data collection

3.3.2. Backbone Assignment

All six different experimental sets are needed for the sequential assignment. Sometimes, due to missing signals (for low sensitivity), or signal overlap, or amide exchange, due to solvent or conformational exchange, and the presence of proline residues, the sequential backbone assignments will be complicated. Sequential NMR spin system connectivity is established using 3 pairs of 3D NMR experiments. They are HNCA, HN(CO)CA, HN(CA)CO, HNCO, HNCACB, and CACB(CO)HN. The nuclei enclosed in brackets only participate in magnetization transfers, the signals from these nuclei do not observe in the experiments.

The first pair includes 3D CBCA(CO)NH, which provides preceding cross-peaks of C_α and C_β, and 3D HNCACB, which provides intra-residual and preceding cross-peaks of C_α and C_β. The intra-residual peak has a higher intensity, and C_α will have the opposite sign of C_β. A strip of HNCACB from residue Ala28-Gly40 is shown in Figure 3.4. The major drawback of HNCACB is that it contains four peaks per residue, which still results in peak overlap. The ambiguities can be resolved by HN(CO)CA and HNCA. In the strip plot of HNCA, two peaks corresponding to Cα i and Cα i−1 are observed, whereas the strip plot of HN(CO)CA provides a single cross peak corresponding

to Cα i−1 is observed. Furthermore, ambiguities can be resolved with HNCO and HN(CA)CO experiments. The strip plot of HNCO single peak is observed corresponding to the carbonyl from i−1 residue whereas, in HN(CA)CO two peaks are observed corresponding to CO i and CO i−1 residue. The HN(CA)CO is the least sensitive because of the fast relaxation of transverse Cα magnetization[175]. The linkage of the spin system is done by a sequential walk using HNCO and HN(CA)CO experiments. Comparison of the chemical shifts between these experiments allows identification of backbone connectivities. A strip of HNCO and HN(CA)CO from residue His80-Ala85 is shown in Figure 3.5. Furthermore, the assignments were further confirmed by CCCONH and HCCCONH experiments, which correlate all aliphatic side-chain carbons and protons of preceding residue, (i-1)th with the amide of ith residue respectively. These experiments are more efficient than 3D HSQC-TOCSY experiments because coherence transfer takes via a larger ^{13}C-^{13}C coupling.[176]

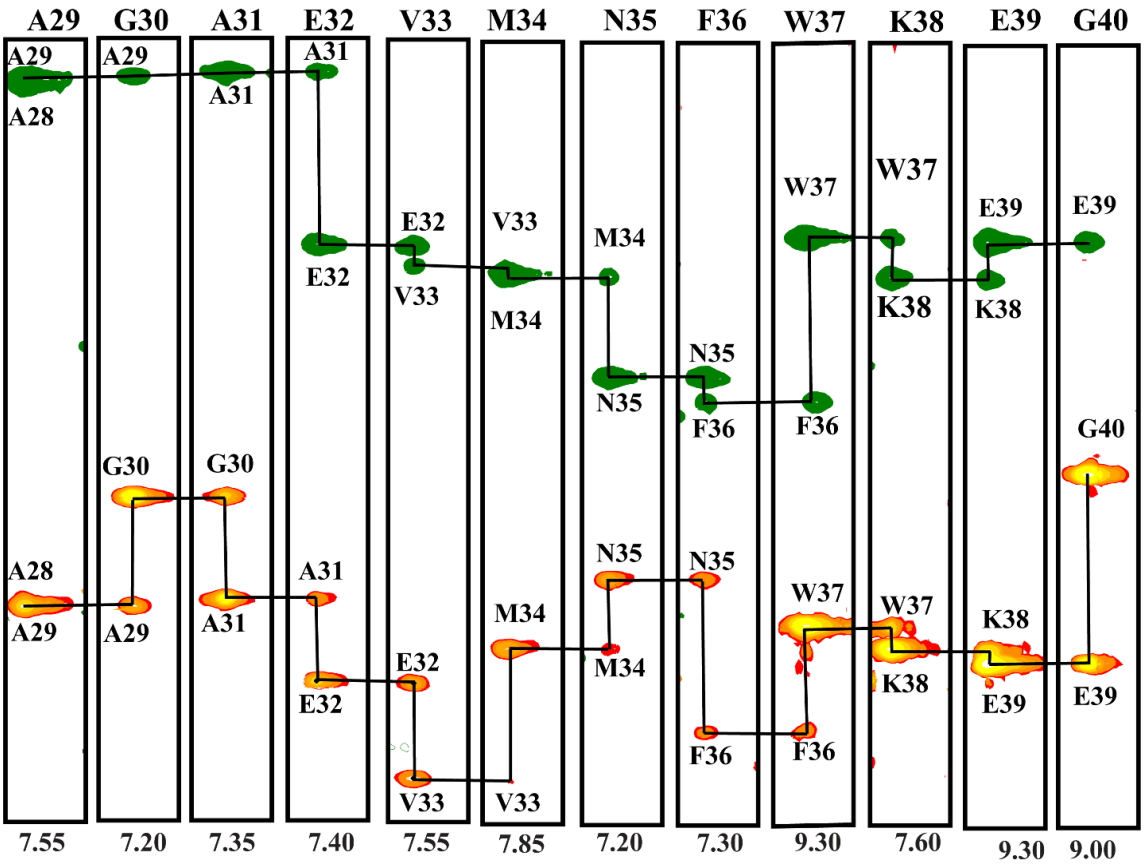


Figure 3.4: Sequential assignments showing the plot of the strip of HNCACB experiment from Ala29 to Gly40 in OfurPBP2. Only the Cα and C_β atoms of the residues were connected with black lines to show the sequential assignment. Positive signals are green and negative signals are orange**.**

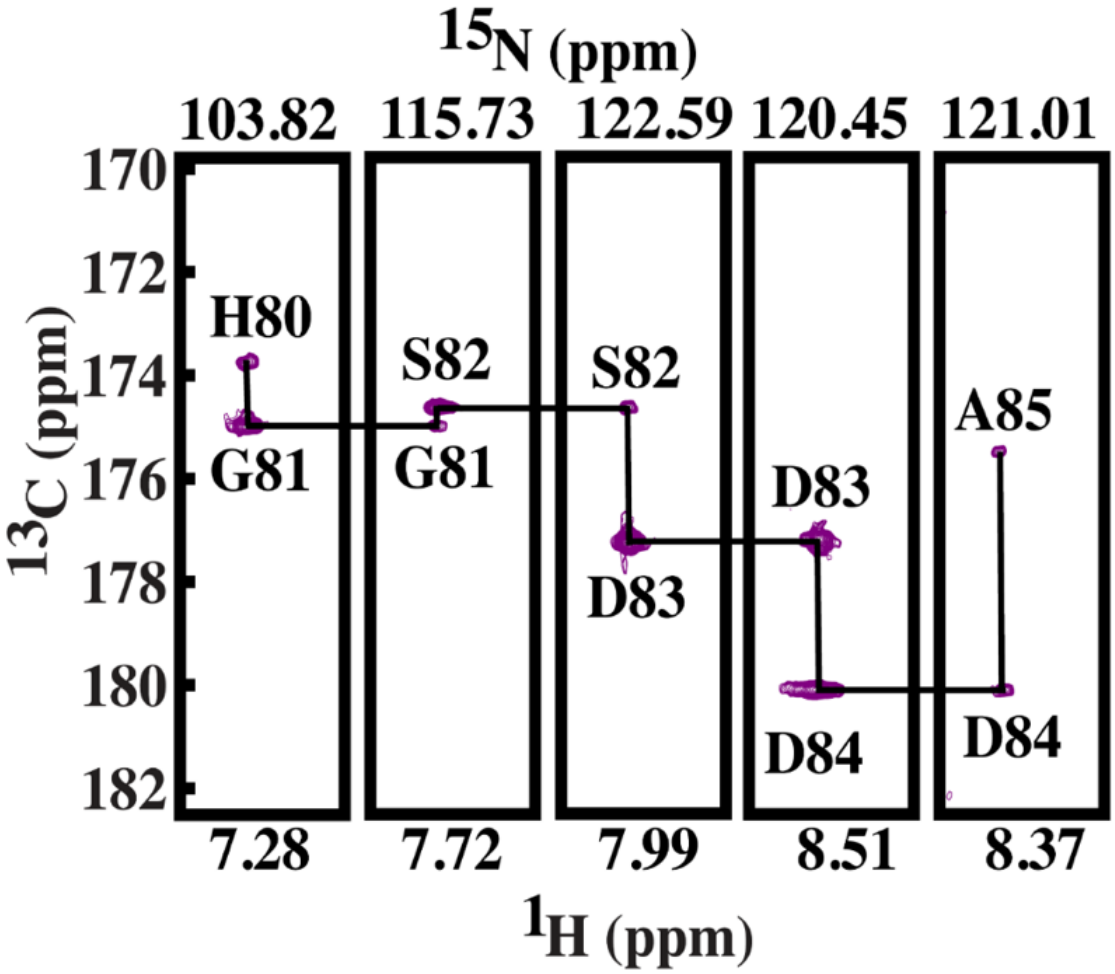


Figure 3.5: Sequential walk showing on the HN(CA)CO spectrum. The intra-residual carbonyl shift is stronger than the preceding ones.

The ^{15}N HSQC-NOESY experiment provides amide to amide NOE cross-peaks. The NH peak of the ith residues is always observed as the diagonal peak. The two other cross-peaks are observed from the nearest neighboring NH, one is from i-1, and the other is from i+1 residue. For the helical section, NOE from NH(i) is visible not only to NH (i±1) but also from NH (i ±2) peaks. The assignments were also cross checked based on sequential NOE connectivities observed in the ^{15}N NOESY spectrum. This linking of spin is called NH-NH sequential walk. The presence of strong sequential NOEs in the amide region indicates the presence of alpha-helices. HN-HN walks from 3D ^{15}N-edited HSQC TOCSY for back-bone assignment showing from residues Asp130-Glu141 is shown in Figure 3.6. NOESY pattern is the indication to identify the characteristic secondary structure of the protein. For the backbone assignment of OfurPBP2, ^{15}N-edit NOESY and ^{15}N-edited TOCSY experiments were used because of the helical nature of the protein. The HNCACB and CBCA(CO)NH spectra are used to assign backbone. However, ^{15}N-edit NOESY and ^{15}N-edited TOCSY spectra are used to confirm the assignment by using NH-NH (amide-amide walking strategy.[178]

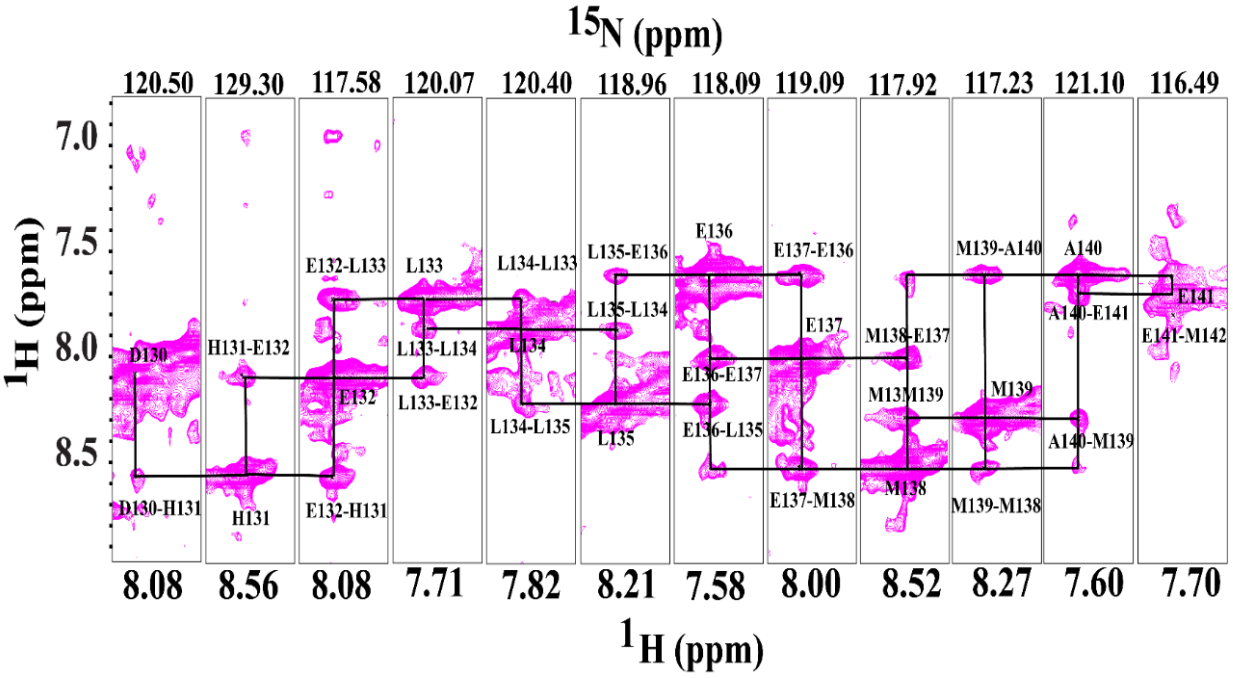


Figure 3.6: Strip plots obtained from 3D ^{15}N-edited HSQC NOESY for back-bone assignment showing from residues Asp130-Glu141. The lines connecting the amide proton cross-peaks demonstrate the NH-NH walking strategy. The assignments of the cross-peaks are shown in the residue number.

In the 2D {^{1}H, ^{15}N} HSQC spectrum the cross-peaks are well-dispersed, which shows that OfurPBP2 is a well-folded protein. In OfurPBP2, the N-terminus Ser1 and Gln1 of the protein could not be assigned in the 2D {^{1}H, ^{15}N} HSQC spectrum because of the missing peak due to the exchange of amide protons with the deuterated solvent. Due to its unique cyclic structure, five proline residues were absent from the 2D {^{1}H, ^{15}N} HSQC spectrum. However, in general, due to the *cis-trans*-isomerization of the proline residues, the neighboring residues experienced two different chemical environments, consequently appearing at two different chemical shift positions. In OfurPBP2, all prolines were found to have trans-conformations, which can be deduced from the observation of a single set of resonances and their chemical shift values are within the range trans

form.[179] The *cis-trans* isomerization on the proline acts as a molecular switch controlling several physiologically important processes.[180,181] The appearance of resonance at 11.5 ppm (downfield) for the amide proton of Asn72 residue could be due to the formation of hydrogen bonding interactions with the solvent.[182] The backbone resonance assignments (^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ, and ^{13}CO) were completed for all residues in the 2D {^{1}H, ^{15}N} HSQC except Ser1, Gln2, and Lys143. All six cysteine residues were in the oxidized state as indicated by their ^{13}Cβ chemical shifts. The C_β shifts can be used to determine the redox state for most of the proteins and are diagnostic of disulfide bond formation. The chemical shift value for oxidized cysteine is within the range of 38.4±3.2 ppm to 43.0±4.2 ppm.[183]

3.3.3. Side-Chain Assignment

The side-chain assignments typically require greater attention because of incomplete data and the degeneracy of aliphatic chemical shifts. The NMR experiments are collected in 90% H_2O, 10% D_2O. The very strong water signal appears in the middle of the spectrum (~ 4.7 ppm). The Hα protons peaks are obscure which are close to the water resonance. The problem of loss of cross-peak due to the saturation of H_α resonances beneath the water resonance can be overcome using pre-TOCSY-COSY or SCUBA-COSY methods. Alternatively, the spectrum can be collected at 10 ^{o}C above or below the previous temperature (25 ^{o}C); the H_2O resonance shifts significantly with temperature, whereas most H_α resonances do not. The TOCSY methods on the large protein (10-20 kDa) suffer from both reductions in sensitivity due to short T_2 values and a small coupling constant. Shorter T_2 (transverse relaxation times) for undeuterated proteins diminishes the sensitivity of the TOCSY based experiments (Figure 3.7). This experiment used isotropic mixing times and more magnetization transfer steps in the pulse sequence, resulting in signal loss. Moreover, sensitivity depends on the coupling constant since a small coupling constant reduces sensitivity. The amount of information in these spectra depends on the length of TOCSY mixing

time, the line width of the proton resonance, and the size of the HN- $H_{\alpha,}$ $H_{\alpha,}$ and H_{β} coupling constants. The side-chain carbon and proton resonances were assigned using 3D CC(CO)NH TOCSY, HCC(CO)NH TOCSY, HCCH TCOSY, and HSQC-TOCSY.

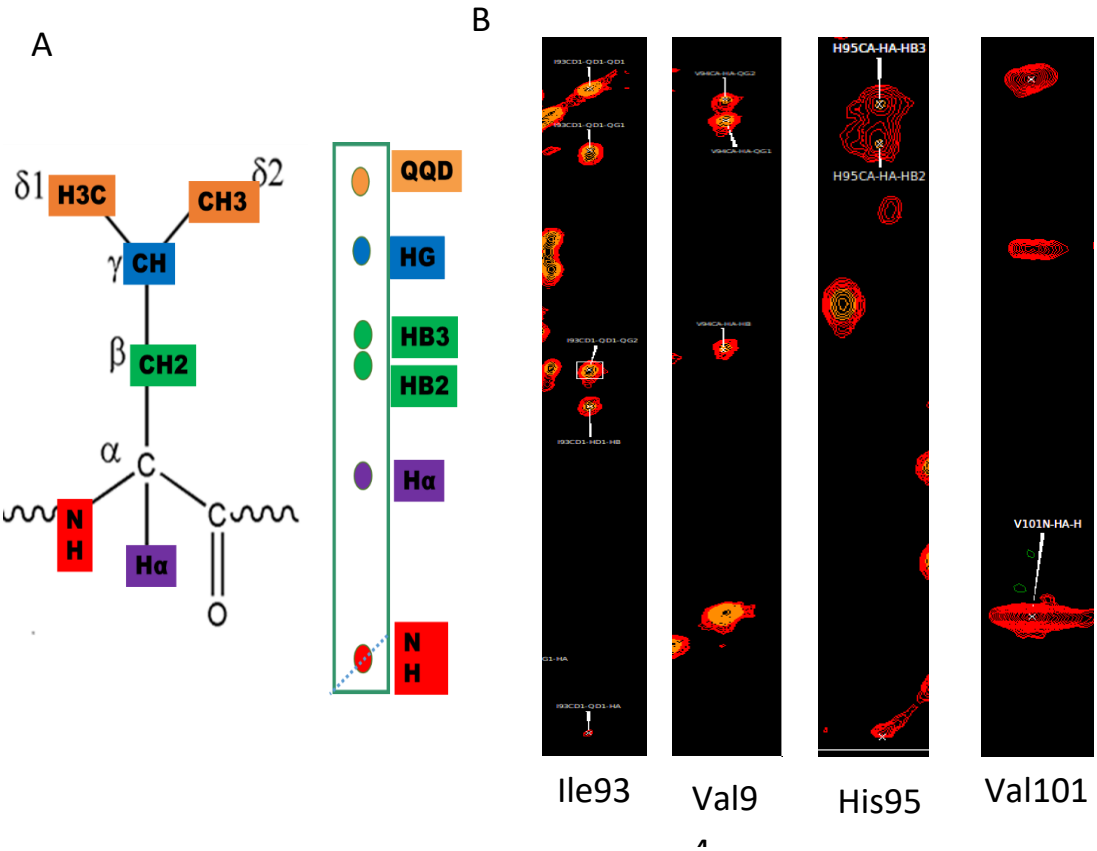
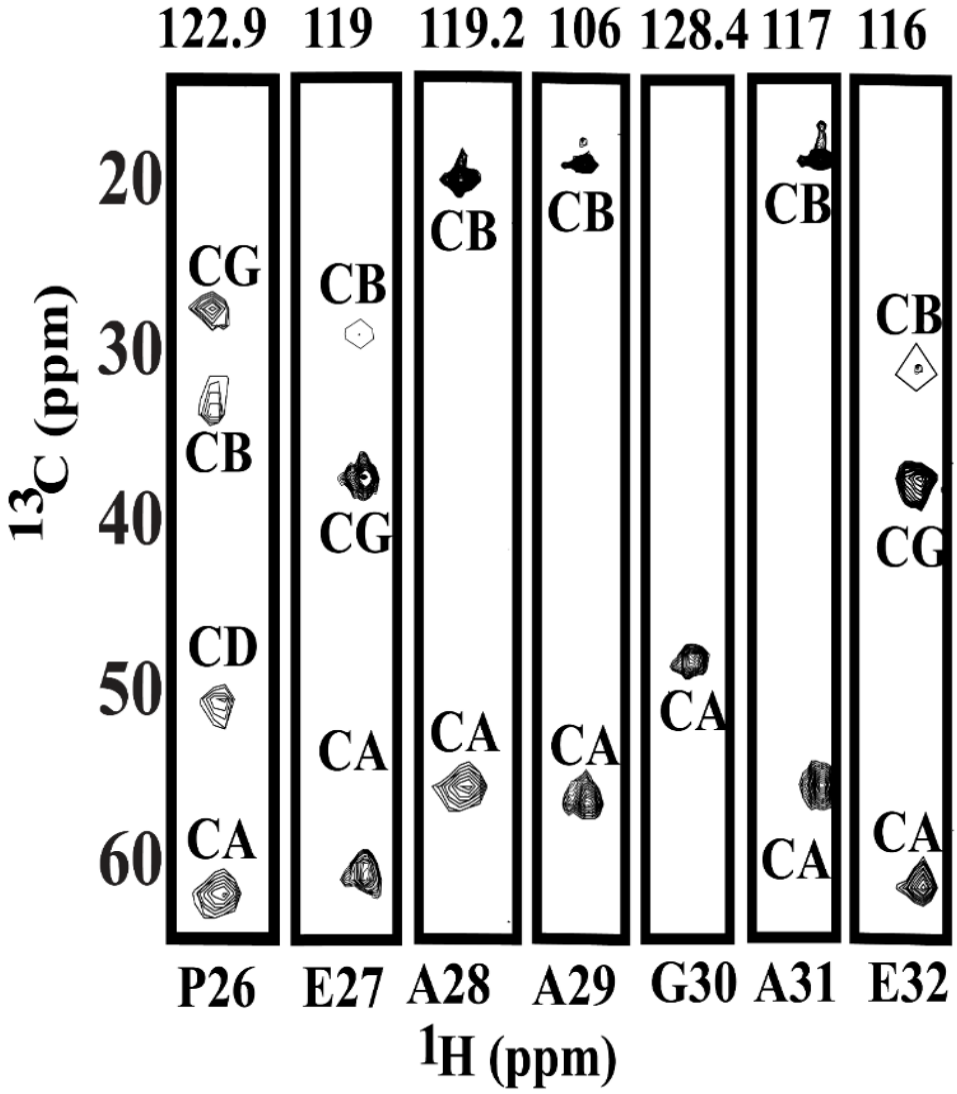


Figure 3.7: A) Schematic showing the TOCSY peaks B) HCCH-TOCSY (strip plot) assignment for the side chain for Ile93, Val94, His95, Val101 (from left to right).

The CC(CO)NH TOCSY provides all the side-chain aliphatic carbon chemical shifts in the amino acid (Figure 3.8). This spectrum is enormously helpful in the assignment process. The time needed for complete magnetization transfer increases with the increasing length of the side-chain of amino

acid in the protein.[184] Other experiments include; HCCH-TOCSY (for proton and carbon resonances), (H)CC(H)-TOCSY (for carbon resonances), ^{15}N-edited HSQC-TOCSY, and H(CCCO)NH-TOCSY. Each cross-peak in the TOCSY type experiments correlates the chemical shifts of each side-chain residue of a particular spin system. The TOCSY spectrum (Figure 3.8) shows the coupling between all the hydrogen nuclei within the amino acid. The advantage of using the 3D HCCH-TOCSY experiment is that both proton and carbon side chains can be obtained simultaneously. The assignment of the aromatic side chain was complicated and could not be assigned only from 3D HCCH-TOCSY. This is because the aromatic CH group has a fast transverse ^{13}C relaxation in aqueous samples. Therefore, other additional experiments, e.g., 3D (HB)CB(CGCD)HD (correlates the side-chain from C_β with the H_δ of aromatic protons in ^{13}C-labeled proteins in D_2O) and 3D (HB)CB(CGCDCE)HE (correlates the sidechain from C_β with the H_ε of aromatic protons) are needed. One of the bottlenecks in NMR structure determination lies in the laborious and time-consuming side-chain resonance assignments. The HCCH-TOCSY of a large protein has low resolution because of the fast ^{13}C transverse relaxation. The relaxation problem could be alleviated by partial deuteration of the sample using deuterium decoupling.[185] The gyromagnetic ratio of deuterium is six times smaller than that of hydrogen. The spin-spin relaxation rate of ^{13}C nuclei is forty times slower in the deuterated protein. Experiments like H(CCCO)NH-TOCSY, CCCONH-TOCSY, and ^{13}C dispersion NOESY-HSQC are advantageous to assign side-chain proton and carbon in partially deuterated protein, where sensitivity is gained in a dramatic way.[186] The CCCONH TOCSY assignment for the aliphatic carbon side-chain of residues from Pro26 to Glu32 and residues from Pro103 to Asp106 is shown in Figure 3.8. The chemical shift is evolved simultaneously on all side-chain carbon nuclei of preceding residue with the amide of ith residue. The nomenclature used was CA, CB, CG, and CD corresponds to C_α, C_β, C γ, Cδ respectively.

A


122.9
119
119.2
106
128.4
117
116
20
30
40
50
60
^{13}C (ppm)
CG
CB
CB
CD
CA
CB
CG
CA
CB
CB
CB
CA
CA
CA
CB
CG
CA
P26
E27
A28
A29
G30
A31
E32
^{1}H (ppm)

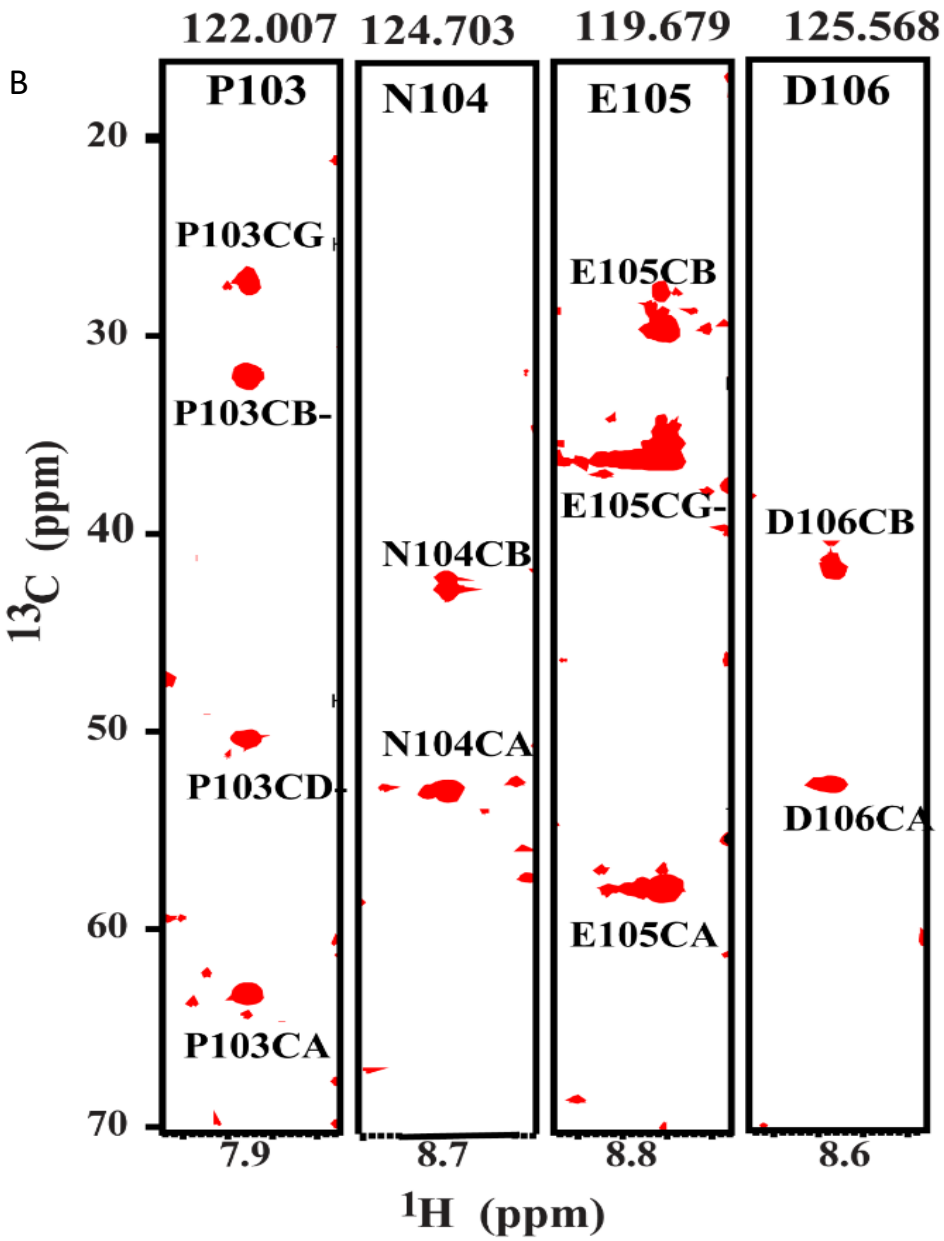


Figure 3.8: The CCCONH TOCSY assignment for the aliphatic carbon side chain. The chemical shift is evolved simultaneously on all side-chain carbon nuclei of preceding residue with amide nitrogen and hydrogen nuclei for ith residue. A) Residues from Pro26 to Glu32 B) residues from Pro103 to Asp106. The nomenclature used CA, CB, CG, and CD corresponds to C_{α}, C_{β}, Cγ, Cδ respectively.

3.3.4. Assignment Challenges

1. The α-helices proteins have comparatively less dispersion of amide chemical shift in the 2D {^{1}H, ^{15}N} HSQC spectrum. The overlapping of resonances complicates the backbone assignment.

2. The signal overlap in the center of the spectrum is due to linewidth broadening, and this peak broadening reduces the sensitivity of the experiment.

3. The resolution of the spectra could be improved by increasing the acquired data points in the indirect dimension. However, the time needed for data collection rises dramatically.

4. The data collection in lower field NMR suffers from the signal overlap. There is undetectable weak and medium-range NOE that gets buried under the strong short-distance intra- and inter-residual NOEs.

5. As the experiment was collected in 90% H_2O and 10% D_2O, the 3D ^{13}C-NOESY-HSQC experiment suffered from the residual water peak, where the H_α and H_β protons close to the water peak were buried.

6. Because of the missing peak and breakage of the sequential connection, we relied on 3D H(CCCO)NH TOCSY which gives all the chemical shifts of all the proton side-chain Hα, H_β, Hγ, and Hδ from the preceding residue and 3D HSQC-TOCSY which gives the chemical shifts of the side-chain protons Hα, H_β, Hγ, and Hδ from the ith residue. The combination of two data sets facilitates the unambiguous linking of two amino acid residues $(i-1)^{th}$ and i^{th}. We could able to connect the primary sequences in the spectra based on the Hα region, which is generally called HA-HA sequential walk.

7. We have used 3D ^{15}N HSQC-NOESY for the NH–NH walk to validate the backbone assignment as well. The ^{15}N HSQC-NOESY experiment provides amide to amide NOE cross-peaks. The NH peak of the ith residues is always observed as the diagonal peak. The two other cross-peaks are observed from the nearest neighboring NH, one is from i-1, and the other is from i+1 residue. For the helical section, NOE from NH(i) is visible not only to NH (i±1) but also from NH (i ±2) peaks. The assignments were also cross verified based on sequential NOE connectivities observed in the ^{15}N NOESY spectrum. This linking of spin is called NH-NH sequential walk.

3.4. Secondary Structure Calculations for Undelipidated OfurPBP2

3.4.1. Torsion Angle Likelihood Obtained from the Shift and Sequence Similarity (TALOS$^+$)

The TALOS$^+$ is a hybrid method for predicting protein backbone Ψ and Φ torsion angles from the backbone chemical shift (^{1}HN, ^{15}N, Hα, ^{13}Cα, ^{13}Cβ, and ^{13}CO). The algorithm is based on a database mining approach, including a neural network. The neural network analyzes the chemical shifts and sequence to estimate the likelihood of a given residue being in a β-sheet, α-helix, or loop conformation.[187] The 2D {^{1}H, ^{15}N}-HSQC spectrum of OfurPBP2 at pH 6.5 is shown in Figure 3.9.

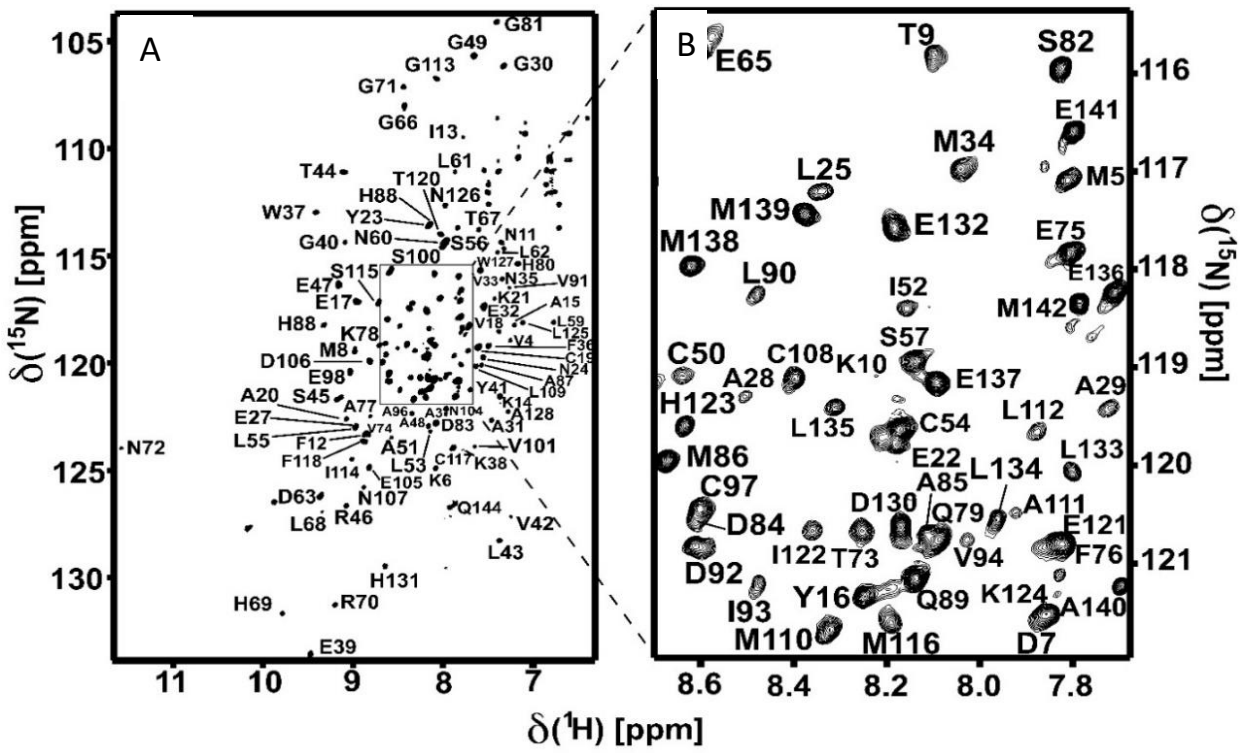


Figure 3.9: (A) The 800 MHz 2D {^{1}H, ^{15}N} HSQC spectrum of uniformly ^{15}N/^{13}C-enriched OfurPBP2 at pH 6.5 at 308 K. The primary structure of OfurPBP2 contains 144 residues. Backbone amide cross peaks have been labeled with residues type and sequence number. (B) The expanded region of the 2D {^{1}H, ^{15}N} HSQC spectrum is shown in the rectangular inset.

Both the backbone chemical shifts and side-chain chemical shifts assignment have been deposited in the BioMagResBank (*www.bmrb.wisc.edu*) under accession number 50074. The ϕ and ψ backbone torsional angles and the secondary structural elements were obtained from TALOS^{+} (Figure 3.10). Based on TALOS^{+} calculations, eight helical regions were observed in the following peptide segments of the protein: 3-23, 27-34, 46-58, 71-79, 84-96, 108-116,119-125, 131-141. A quick inspection of the secondary structure elements of OfurPBP2 showed that the C-terminus had an α-helical structure at pH 6.5 (Figure 3.10). This observation was quite surprising and contrasted with previously reported lepidopteran PBPs, including ApolPBP1[138], BmorPBP[188], LdisPBP2[173],

and AtraPBP1.[85] The C-terminus of these PBPs is a random coil and is exposed to the solvent in the ligand-bound conformation at pH > 6.0. However, the ligand is released at a lower pH (<5.0) near the olfactory receptor neuron site through a pH-dependent conformational switch, where the C-terminus switches to a helix and outcompetes the pheromone for the pocket.

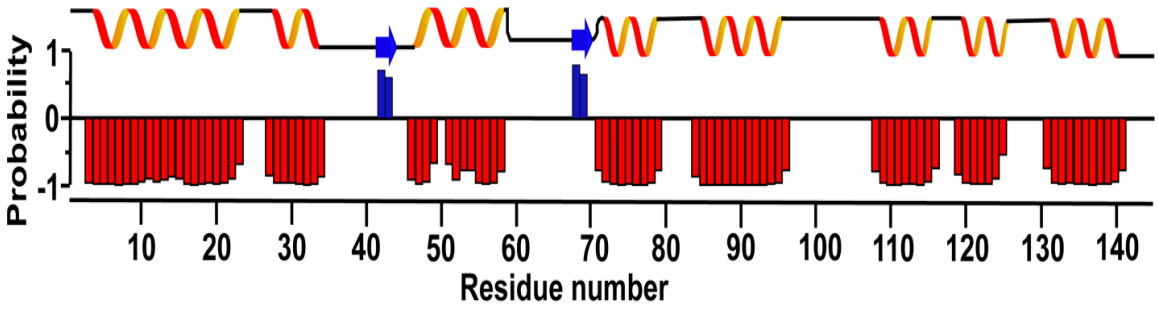


Figure 3.10: Secondary structure prediction of OfurPBP2 obtained with TALOS+[171] using the ^{1}H, ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and ^{13}C backbone chemical shifts. The secondary structure prediction is shown as red bars for α helices and blue bars for β strands, with the height of the bars representing the probability of the secondary structure (-1 for α-helix, 0 for random coil, 1 for β-strand).

3.4.2. Secondary Shift Propensity (SSP) Using the Chemical Shifts

Secondary Structure Propensity (SSP) is an algorithm that combines chemical shifts from different nuclei into a score, which reflects the expected fraction of α- or β-structure.[172] In SSP, the sequence-dependent deviations of experimental resonance assignments from the random-coil chemical shift were calculated. The deviations of specific chemical shifts from their expected random coil values, ($\Delta\delta = \delta_{observed} - \delta_{coil}$), are a useful measure of secondary structure. The $\Delta C_{\alpha} - \Delta C_{\beta}$, ΔC_{α}, and ΔC_{β} values were plotted against the protein sequence (Figure 3.11). Here, positive values indicate the α-helical structure. Negative values indicate β-strand or extended structure. The stretches of positive values indicate eight helical regions α_1 to α_8, which are mostly separated by shorter

stretches lacking well-defined secondary structures (most likely loops), while a negative value indicates the presence of β-sheet structure.

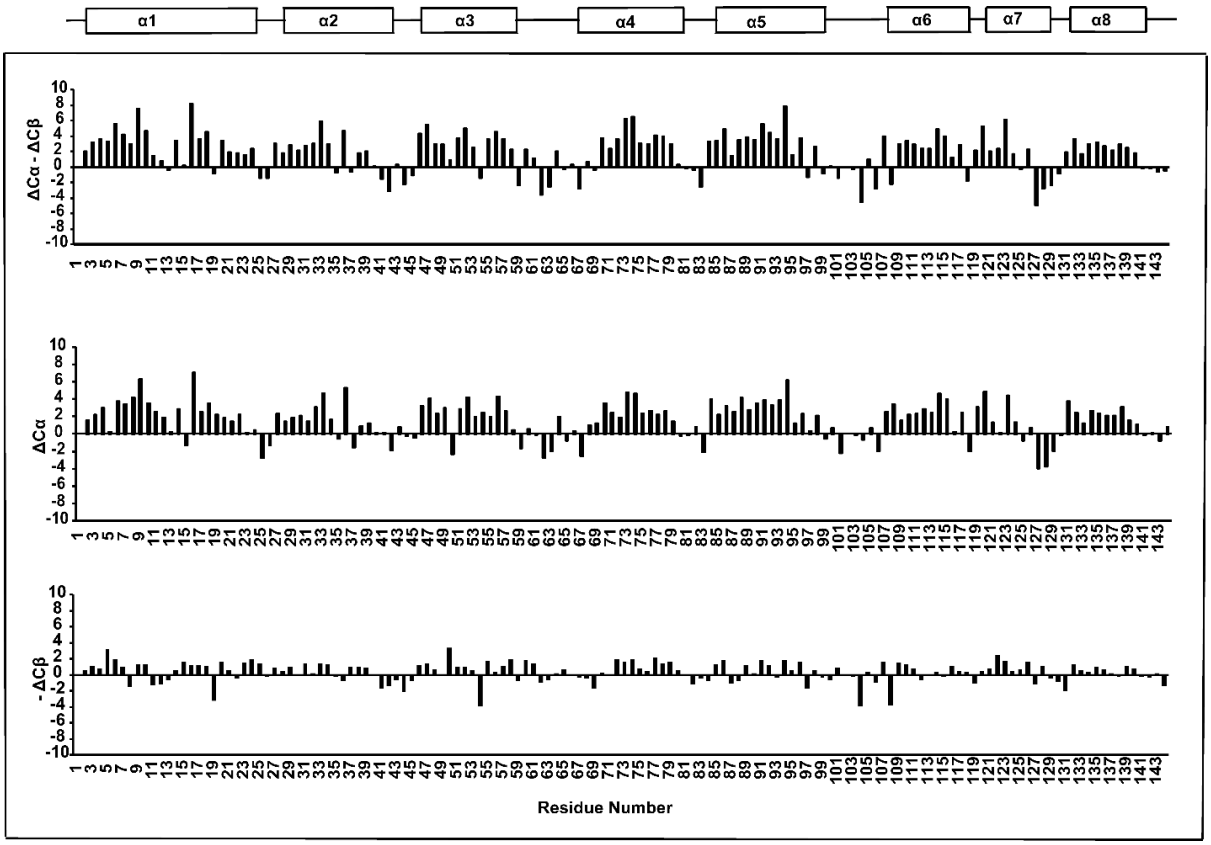


Figure 3.11: Secondary chemical shift, ΔCα-ΔCβ, $\Delta C_{\alpha,}$ and - ΔC_{β} are plotted against the linear amino acid sequence. ΔC_{α} and ΔC_{β} are calculated by subtracting random coil values from the C_{α} and C_{β} shift.[189] The helical regions are shown at the top.

3.4.3 Chemical Shift Index CSI 3.0

The Chemical Shift Index or CSI 3.0 (*http://csi3.wishartlab.com*) is a web server used to determine the secondary and super-secondary structures in protein chains using backbone chemical shifts and

protein sequences. CSI 3.0 identifies 11 different types of secondary and super-secondary structures, including; helices, β-strands, coil regions, five common β-turns (type I, II, I′, II′ and VIII), β hairpins as well as interior and edge β-strands.[190] The CSI 3.0 uses the latest algorithm, which incorporates CSI 2.0, TALOS, and RCI (Random Coil Index). This can calculate the flexibility and order parameters of a protein sequence as well. The server generates a text-based secondary structure assignment and a colorful CSI plot (Figure 3.12). In this method, the experimentally observed chemical shift is compared with a residue-specific standard random coil chemical shift. Three different indexes (1, 0, -1) are used to explain the chemical shift information. When the index was plotted against the protein sequence, we could identify secondary structural elements like beta strands (clusters of +1 values), alpha helices (clusters of -1 values), and random coil segments (clusters of 0 values).

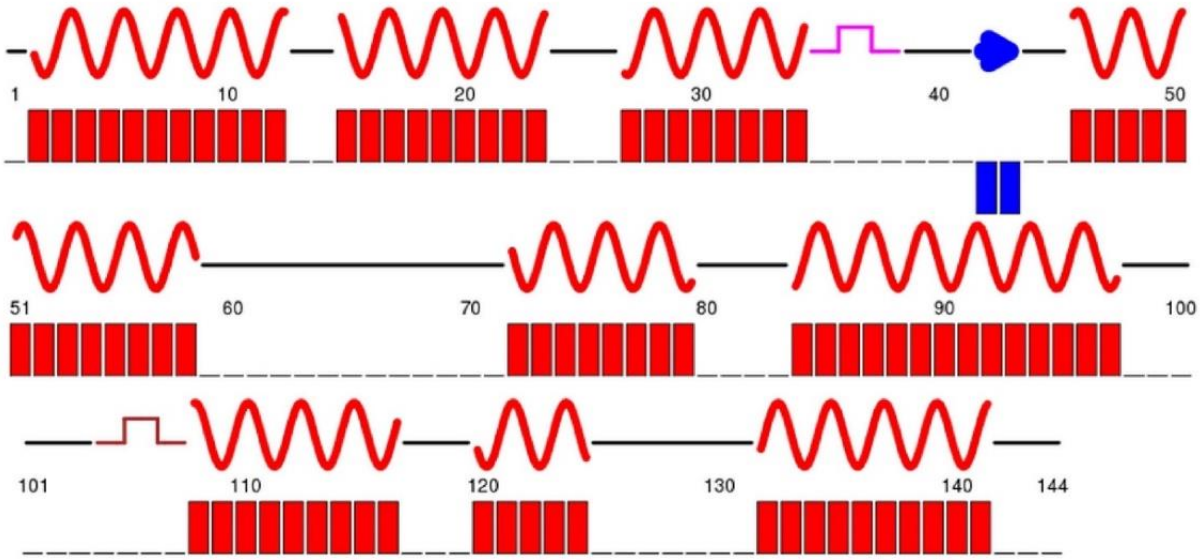


Figure 3.12: The output of the CSI 3.0 servers[190] presented in a residue specific-way (red block stands for α-helix, blue block stands for β-strand and black line stands for random coil and purple color showing turn I).

3.5. NMR data collection of delipidated OfurPBP2

NMR data were collected on 450 uM protein concentration in 50 mM sodium phosphate buffer, pH 6.5 with 1 mM EDTA, and 0.01% NaN_3, and 10% D_2O. The NMR data were collected at 35 ^0C on a Bruker AVANCE 800 MHz spectrometer equipped with a cryogenic triple resonance probe at the Oklahoma State University. For the sequential backbone assignment of ^1HN, $^1H_\alpha$, ^{15}N, $^{13}C_\alpha$, $^{13}C_\beta$, and ^{13}CO resonances: 2D {^1H, ^{15}N} HSQC, 2D {^1H, ^{13}C }-HSQC, 3D HNCA, 3D HN(CO)CA, 3D HNCO, 3D HN(CA)CO, 3D HNCACB, 3D CACB(CO)NH experiments were collected. Similarly, for proton and carbon side-chain resonances assignment 3D H(CCCO)NH-TOCSY, 3D CC(CO)NH-TOCSY, 3D HCCH-TOCSY, and HSQC-TOCSY experiments were collected. For NOE distance restraints 3D ^{15}N HSQC-NOESY and ^{13}C-edited HSQC-NOESY with mixing times of 85 and 120 ms respectively. The ^{13}C carrier frequency in ^{13}C-edited HSQC-NOESY is 44 ppm for the aliphatic 125 ppm for aromatic regions. The backbone assignments were confirmed by 3D ^{15}N-edited HSQC-NOESY. The detail for parameters used in NMR experiments is listed in Table 3.2.

Table 3.2: Parameters used in NMR experiments for undelipidated OfurPBP2

Spectrum	Nuclei	Data size (Complex point) $^{15}N\times^{13}C\times^{1}H$	Sweep width(SW) (ppm) $^{15}N\times^{13}C\times^{1}H$	Carrier frequency $^{15}N\times^{13}C\times^{1}H$	Number of scan (NS)
2D {^{15}N ^{1}H } HSQC	^{15}N, ^{1}H	300×2048	36×16 ($^{15}N\times^{1}H$)	119×4.7 ($^{15}N\times^{1}H$)	8
2D {^{13}C, ^{1}H} HSQC	^{13}C, ^{1}H	256×2048	70×16 ($^{13}C\times^{1}H$)	45 ×4.7 ($^{15}N\times^{13}C\times^{1}H$)	8
3D HNCACB	^{15}N, ^{13}C, ^{1}H	220×80×2048	36×70×16 ($^{15}N\times^{13}C\times^{1}H$)	119×43×4.7 ($^{15}N\times^{13}C\times^{1}H$)	8
3D CBCACONH	^{15}N, ^{13}C, ^{1}H	190×80×2048	36×70×16	119×45×4.7 ($^{15}N\times^{13}C\times^{1}H$)	8
3D HNCA	^{15}N, ^{13}C, ^{1}H	150×60×2048	36×70×16	119×45×4.7 ($^{15}N\times^{13}C\times^{1}H$)	8
3D HN(CO)CA	^{15}N, ^{13}C, ^{1}H	128×40×2048	36×70×16	119×45×4.7 ($^{15}N\times^{13}C\times^{1}H$)	8
3D HNCO	^{15}N, ^{13}C, ^{1}H	128×60×2048	36×70×16	117×45×4.7 ($^{15}N\times^{13}C\times^{1}H$)	8
3D HN(CA)CO	^{15}N, ^{13}C, ^{1}H	140×56×2048	36×70×16	117×45×4.7 ($^{15}N\times^{13}C\times^{1}H$)	16
3D-TOCSY-HSQC	^{15}N, ^{1}H, ^{1}H	70×127×2048	36×70×14	117×4.7×4.7 ($^{15}N\times^{13}C\times^{1}H$)	8
3D-CCCONH	^{15}N, ^{13}C, ^{1}H	40×220×2048	36×74×16	117×43×4.7 ($^{15}N\times^{13}C\times^{1}H$)	4
3D-H(CCCO)NH	^{15}N, ^{1}H, ^{1}H	60×160×2048	36×14×14	117×4.7×4.7 ($^{15}N\times^{1}H\times^{1}H$)	16
3D- ^{15}N NOESY-HSQC Mixing time 85 ms	^{15}N, ^{1}H, ^{1}H	70×256×2048	36×14×14	117×4.7×4.7 ($^{15}N\times^{1}H\times^{1}H$)	8
3D-^{13}C-NOESY-HSQC (Aliphatic) Carrier frequency 45ppm, and mixing time 110ms	^{13}C, ^{1}H, ^{1}H	256×70×2048	74×14×14	45×4.7×4.7 ($^{13}C\times^{1}H\times^{1}H$)	8
3D-^{13}C-NOESY-HSQC (Aromatic) Carrier frequency 125ppm,	^{13}C, ^{1}H, ^{1}H	40×128×2048	60×14×14	125×4.7×4.7 ($^{13}C\times^{1}H\times^{1}H$	8

and mixing time 110ms					

3.6. Chemical shift assignment of delipidated OfurPBP2

The NMR spectrum of the delipidated OfurPBP2 exhibited more signals than the expected number, and the dispersion of the peak is poor. This is likely due to the presence of conformational heterogeneity. The signal intensities were not uniform which is also an indication of conformational dynamics or multiple conformations.[191] At the lower counter level, two sets of peaks are observed which is likely from the minor conformation of the protein. The minor sets of the peaks could not be assigned because of peak broadening, severed overlapping, and low intensity. The molten globule state might consist of a mixture of multiple conformations which is characterized by broadened peaks. One of the reasons for the reduction in the resonance intensity in the spectra is the presence of a heterogeneous ensemble of structures, such as a molten globule. Molten globules protein still consists of secondary structure; however, they lack tertiary folding. The low dispersion of the chemical shifts and the broadening of NMR lines suggests that the delipidated OfurPBP2 likely to have a molten globule state which was similar to that of α-lactalbumin.[191]

The NMR studies of the molten globule state are mainly based on the secondary structure which could be obtained from secondary chemical shifts. The molten globule is thought to have dynamic hydrophobic core and NMR signals for the aromatic groups broaden. There is the absence of a well-packed hydrophobic pocket on the molten globule protein which is the characteristic of unliganded OfurPBP2. The characterization of molten globule states of protein by NMR is difficult. The molten

globule state consists of many conformational states and also dynamics on the NMR time scales that correspond with an intermediate exchange, which leads to broadening of the NMR lines and consequently loss of signals.

The overlay of 800 MHz 2D {^{1}H, ^{15}N} HSQC spectrum of uniformly ^{15}N/^{13}C-enriched delipidated OfurPBP2 at pH 6.5 and 35 °C at different counter level (Figure 3.13). The green color is the peaks showing at the lower counter level. The red color showing at higher counter level. The primary structure of OfurPBP2 contains 144 residues. At the lower counter level (green color peak) there are more numbers of peaks labeled with a star. These peaks suffer from severed overlapping and low intensity. It shows that there are presences of more than one conformation. The peak corresponds to the major conformation that could be assigned at a higher counter level shown in figure below.

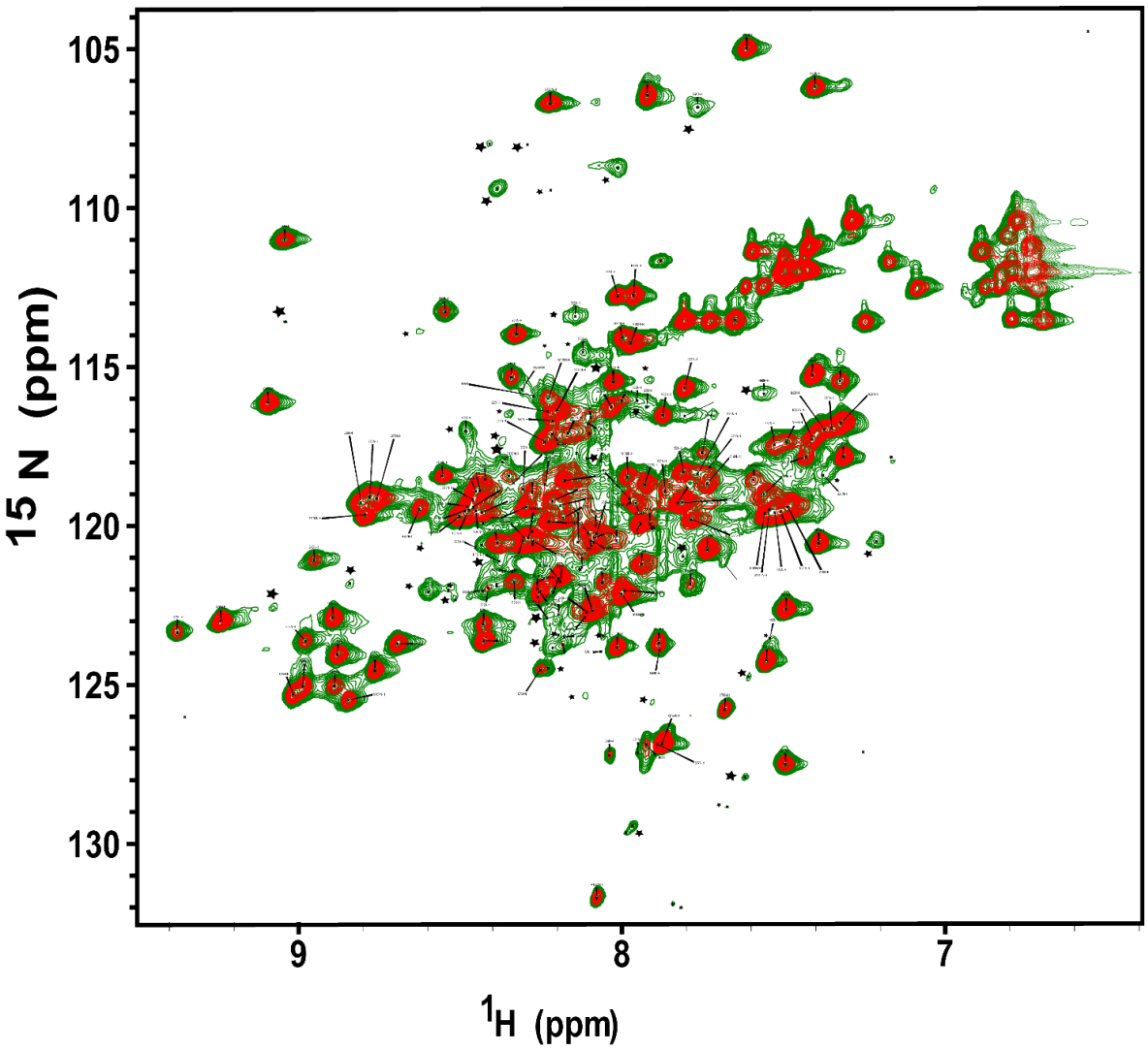


Figure 3.13: The overlay of 800 MHz 2D {^{1}H, ^{15}N} HSQC spectrum of uniformly ^{15}N/^{13}C-enriched delipidated OfurPBP2 at pH 6.5 and 35 °C at different counter level. The green color is the peaks showing at the lower counter level. Red color shows peaks at the higher counter level. The primary structure of OfurPBP2 contains 144 residues. At the lower counter level (green color peak) there are more numbers of peaks labeled with star. These peaks suffer from severed overlapping and low intensity. It shows that there are presences of more than one conformation. We have assigned only one conformation at a higher counter level which is shown below.

The backbone assignment was done by using 2D {^{1}H, ^{15}N} HSQC, 3D HNCA, 3D HN(CO)CA, 3D HNCO, 3D HN(CA)CO, 3D HNCACB, 3D CACB(CO)NH experiments. The details of the magnetization transfer were explained in Chapter 1. The sequential NMR spin system connectivity is established using CBCACONH and HNCACB which are two complementary experiments. The 3D CBCACONH provides peaks for C_α and C_β for the preceding peaks. The strip plot of CBCACONH provided peaks for C_α and C_β from preceding residue only. The HNCACB provided peaks for C_α and C_β for both preceding, (i-1) and intra residual, ith peaks. Sequential assignments were done as explained above for undelipidated OfurPBP2. The strip of the HNCACB experiment showing from Ile13 to Lys21 is shown in Figure 3.14. The Cα and C_β atoms of the residues were connected with black lines to show the sequential assignment. Positive signals are green belongs to Cα and negative signals are red belongs to C_β. The strip of HNCACB consists of four peaks C_α and C_β peaks from preceding residue have lower intensity whereas, C_α and C_β of inter-residual peaks has a higher intensity.

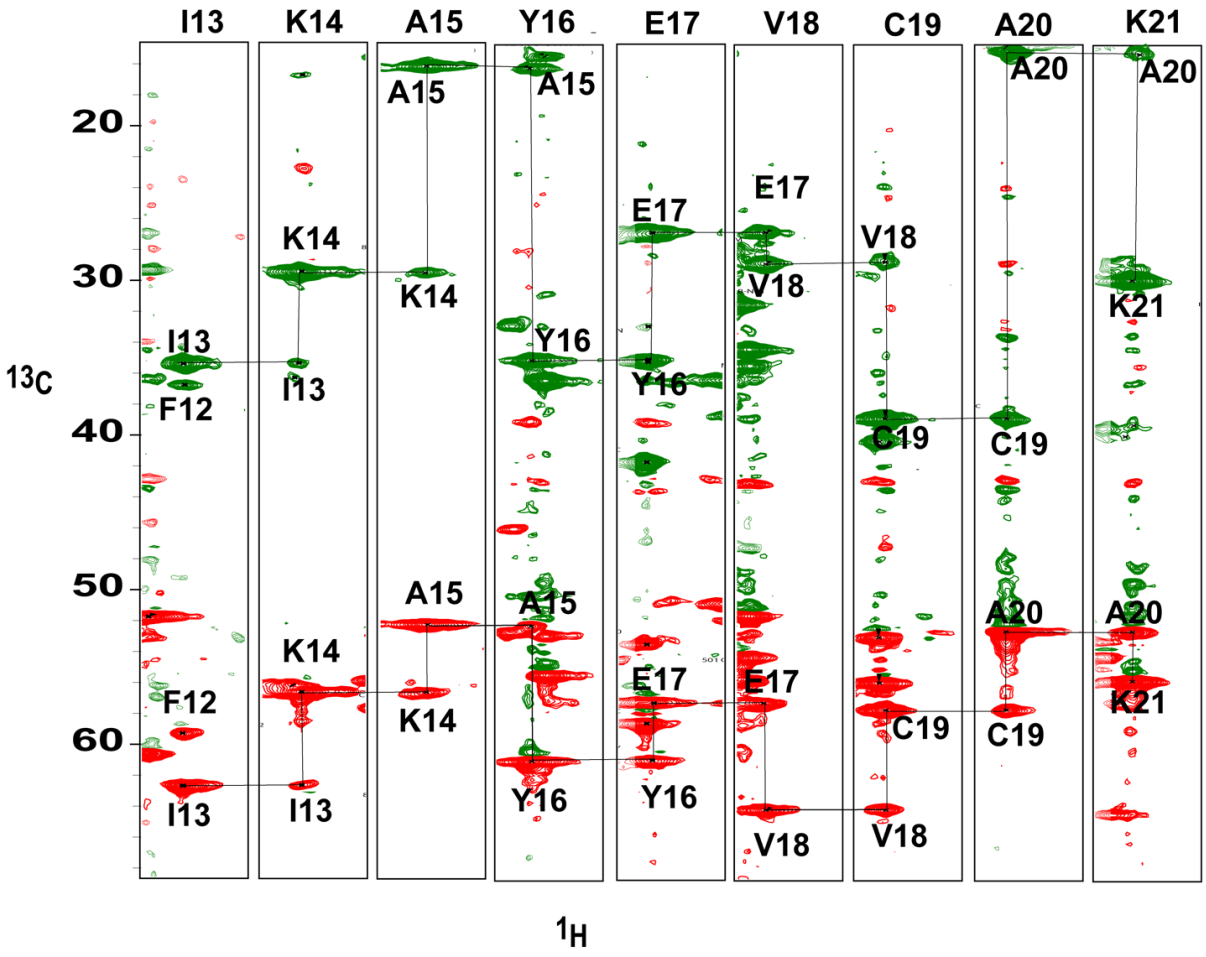


Figure 3.14: Sequential assignments showing the plot of the strip of HNCACB experiment from Ile13 to Lys21 in delipidated OfurPBP2. The Cα and C_β atoms of the residues were connected with black lines to show the sequential assignment. Positive signals are green and negative signals are red.

After the assignment of backbone, we assign the side-chain nuclei of delipidated OfurPBP2. As mentioned above we have used 3D H(CCCO)NH-TOCSY, 3D CC(CO)NH-TOCSY, 3D HCCH-TOCSY, and HSQC-TOCSY experiments. The 3D CC(CO)NH-TOCSY and H(CCO)NH-TOCSY provide correlations linking either aliphatic carbons and protons with backbone amide group chemical shifts respectively. The 3D CC(CO)NH-TOCSY experiment provides all the aliphatic carbon side chains from preceding (i-1) residue correlated with the amide of i^{th} residue (Figure 3.15). This experiment is very useful for validating the sequential assignment of backbone. All the side chain carbon chemical shifts including valine, leucine, isoleucine, arginine, lysine, and proline can be assigned unambiguously.

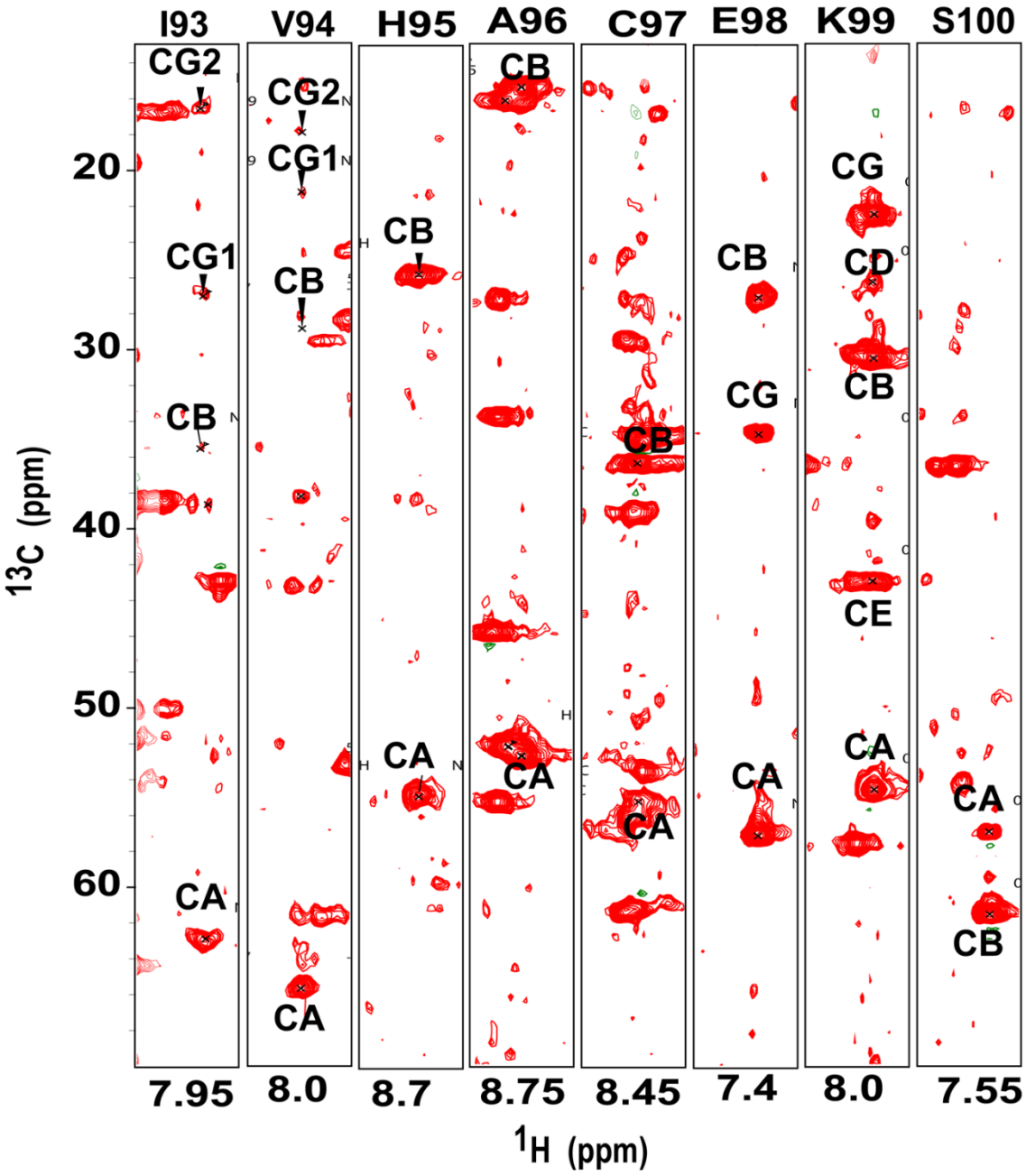


Figure 3.15: The strip plot of the CCCONH spectrum showing the sequential assignment of carbon side-chain from residue Ile93 to Ser100. The chemical shift is evolved simultaneously on all side-chain carbon nuclei of preceding residue with amide from ith residue. The nomenclature used CA, CB, CG, and CD corresponds to Cα, Cβ, C γ, Cδ respectively.

Similarly, assignments of side-chain protons of (i-1) residues were accomplished using 3D H(CCCO)NH-TOCSY experiment. The side-chain protons of ith residues were assigned using the 3D HSQC-TOCSY experiment (Figure 3.16). These two spectra were used simultaneously side-by-side for assigning side-chain protons. The use of alpha proton from the i-1 and ith residue to confirm the sequential assignment of the backbone is known as HA-HA sequential walk. The nomenclature HA, HB2, HB3, and HG corresponds to Hα, Hβ2, Hβ3, and Hγ respectively and QB is used to represent a degenerate chemical shift for HB1, HB2, and HB3 of alanine residue.

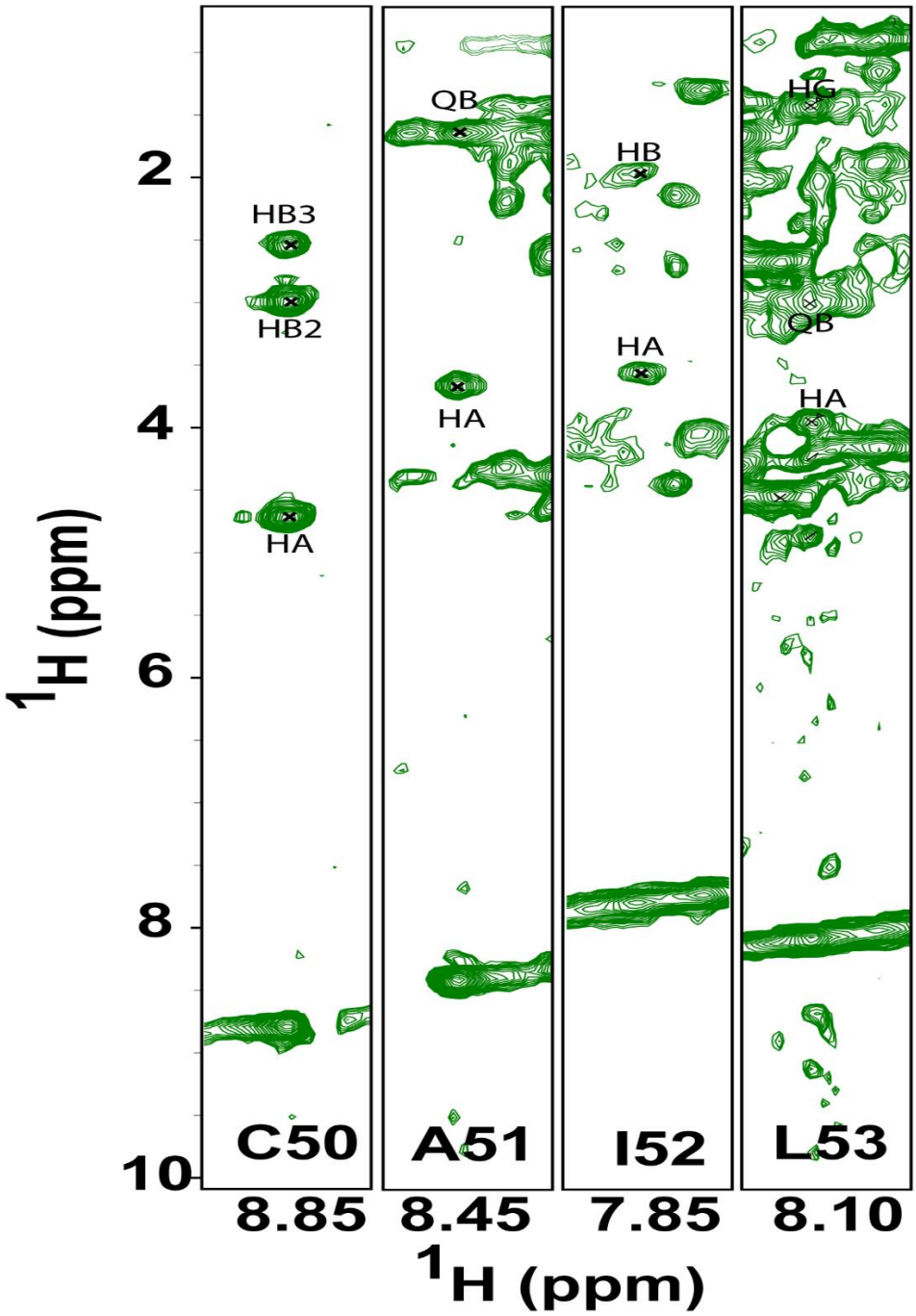


Figure 3.16: The 3D HSQC-TOCSY spectra showing side-chain assignments of the protons from residue Cys50 to Leu53. The nomenclature HA, HB2, HB3, and HG corresponds to Hα, Hβ2, Hβ3, and Hγ respectively and QB is used to represent degenerate chemical shift for HB1, HB2, and HB3 of alanine residue.

The resonances from the N-terminus Gln1 to Lys6, and could not be assigned which is likely due to intermediate exchange resulting in broadening of the signals and sometimes beyond detection which is likely due to conformational fluctuations. Similarly, the C-terminus residues: Leu134, Leu135, Glu136, Glu137, and Lys143 could not be assigned due to the lack of observable resonances. The missing resonances on this segment are likely due to conformational exchange (the exchange of backbone amide protons with solvent protons, or exchange between different protein conformers, or due to the conformational fluctuations on the C-terminus), which resulted in the broadening of the resonances beyond detection.

3.6 Secondary Structure for Delipidated OfurPBP2

After the delipidation of the endogenous ligand, the conformation of the protein is drastically changed. The 2D {^{1}H, ^{15}N} HSQC spectrum of delipidated OfurPBP2 with complete assignments is shown in Figure. 3.17. The NMR spectrum of the delipidated OfurPBP2 exhibited more signals than the expected number, and the dispersion of peaks is poor. This might be due to the presence of multiple conformations. Doubling of resonances was noted for when we go to the lower counter level, reflecting the mixture of other conformation of OfurPBP2 under the current experimental condition. One set of resonances could be assigned at a higher counter level, other set could not be possible to assign because of heavy overlapping and very broad peaks or even low intensities. The molten globule state might consist of a mixture of multiple conformations which is characterized by broadened peaks. One of the reasons for reduction in the resonance intensity in the spectra is due to the presence of a heterogeneous ensemble of structures, such as a molten globule. Molten globules protein still consists of secondary structure; however, they lack tertiary folding. The NMR studies of the molten globule state are mainly based on the secondary structure which could be obtained from secondary chemical shifts. The molten globule has thought to have dynamic hydrophobic core and NMR signals for the aromatic groups are broadening. There is likely an absence of a well-packed hydrophobic pocket on the molten globule protein which is the

characteristic of unliganded OfurPBP2. The characterization of molten globule states of protein by using NMR is difficult. The molten globule state is consists of many conformational states and also dynamics on the NMR time scales that correspond with intermediate chemical exchange resulting in broadening NMR signals and even sometimes beyond detection which is observed in molten globule states of α-lactalbumin and apomyoglobin.[192]

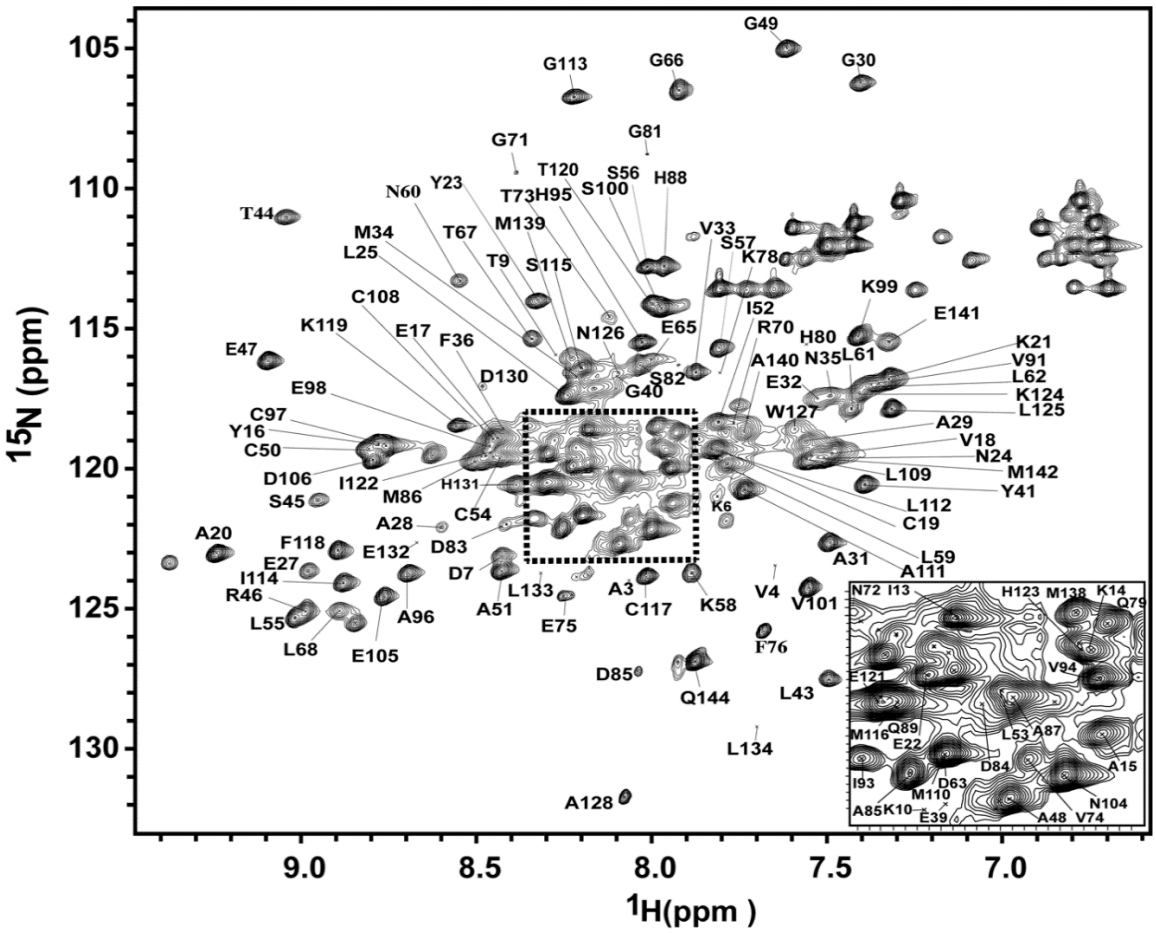


Figure 3.17: The 800 MHz 2D {^{1}H, ^{15}N} HSQC spectrum of uniformly ^{15}N/^{13}C-enriched delipidated OfurPBP2 at pH 6.5 and 35 °C. The primary structure of OfurPBP2 contains 144 residues. The backbone amide cross peaks have been labeled with residues type and sequence number. The expanded region of the 2D {^{1}H, ^{15}N} HSQC spectrum is shown in the rectangular inset.

The 2D {^{1}H, ^{15}N} HSQC shows the reduction in peak dispersion and overlapping, indicating that the protein is behaving as a molten globule. The assignment of both N- and C-terminal regions is extremely complicated, which may be potentially due to the flexibility of these regions resulting in peak broadening. Despite the heavily overlapping spectra, nearly 85 % of the backbone and 80 % side-chain assignment was achieved. The residues Met5, Met8, Phe12, Trp37, Val42, Asp63, Arg70, Ala77, Leu134, Leu135, Glu137, and Lys143 could not be assigned. Both the assignment for backbone and side-chain chemical shifts of delipidated OfurPBP2 have been deposited at the BMRB (*http://www.bmrb.wisc.edu*). The backbone torsional angles (ϕ and ψ) and the secondary structural components were obtained from TALOS^{+}. The TALOS+ calculations (Figure 3.18), showed that the delipidated protein consisted of eight helical regions 11-21, 27-34, 46-49, 51-54, 76-79, 84-96, 108-116,119-125, and 138-141. The secondary structure elements were also calculated with secondary chemical shifts: $\Delta C\alpha - \Delta C\beta$, $\Delta C\alpha$, and $-\Delta C\beta$ (Figure. 3.19). Again, seven α-helices were obtained with breakage on the α3 region. The C-terminal and N-terminal region showed the absence of secondary structure elements. The secondary structure elements of delipidated OfurPBP2 showed that both the N- and C-terminus were mainly random coils. The residues from the C-terminal residues were not observed or were difficult to assign due to peak broadening, probably due to the intermediate exchange with the solvent. Similarly, the N-terminus residues also suffered peak broadening, which is again likely due to the intermediate exchange. NOEs assignments will further confirm the detailed secondary component on the C and N terminal.

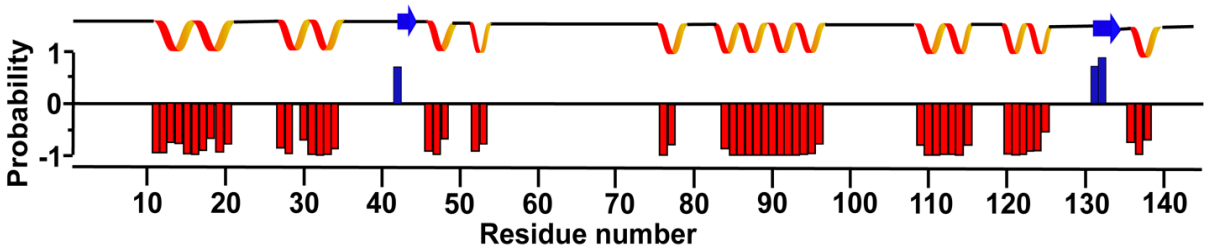


Figure 3.18: Secondary structure prediction of delipidated OfurPBP2 obtained with TALOS+ using the ^{1}H, ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and ^{13}C backbone chemical shifts. The secondary structure prediction is shown as red bars for α-helices and blue bars for β-strands.

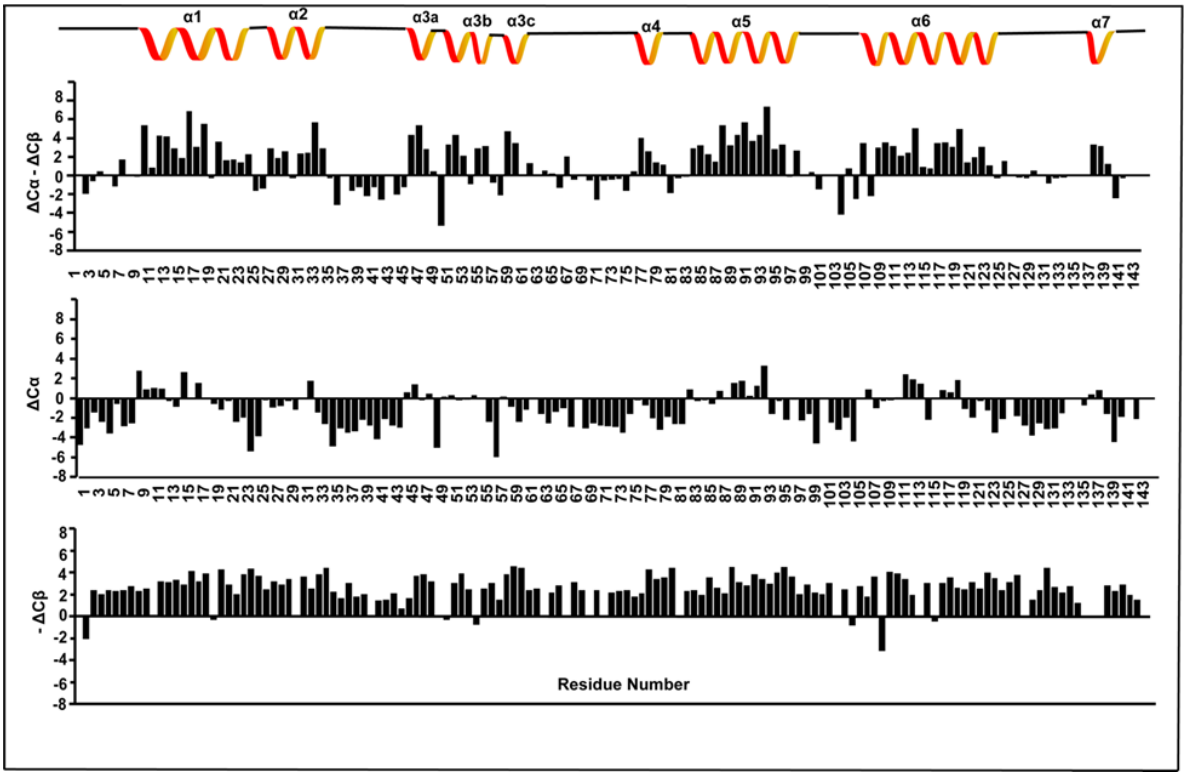


Figure 3.19: Secondary chemical shifts, ΔCα − ΔCβ, ΔCα, and − ΔCβ, are plotted against the linear amino acid sequence. ΔC_{α} and ΔC_{β} are calculated by subtracting random coil values from the C_{α} and C_{β} shift.[189] The helical regions are shown at the top**.**

Comparing undelipidated OfurPBP2 protein with delipidated revealed a few significant differences. One difference was at the C- and N-termini of the undelipidated protein, which has a well-defined helix, whereas the delipidated protein is an unstructured coil. Similarly, the undelipidated protein on the flap region from residue 61-69 was a beta loop, whereas the delipidated protein was again unstructured. The backbone NMR assignments of both delipidated and undelipidated OfurPBP2 serve as a baseline for assessing tertiary structure. This data set is the foundation for studying the tertiary structure, dynamics, and ligand binding properties of OfurPBP2.

3.7. Disulfide Bond Mapping from Chemical Shifts

The ^{13}C NMR chemical shifts are excellent probes to detect disulfide bonds. The reduced (free) and oxidized (disulfide-bonded) state of cysteine can be separated based on ^{13}C chemical shifts.[183] The C_α shifts for reduced cysteine fall into two distinct regions with minimal overlap (α-helix: 62.6±1.7 and β-strand: 56.6±1.8 ppm), whereas for oxidized cysteine there is significant overlap (α-helix: 57.6±2.3 and β-strand: 54.8±2.1 ppm). In reduced cysteine, the observed C_β shifts for the α-helix and β-strand were 26.5±1.1 ppm and 29.7 ± 2.0 ppm, respectively, whereas, for oxidized cysteine, they were 38.4±3.2 ppm and 43.0±4.2 ppm, respectively. The C_β shifts can be used to determine the redox state for most of the proteins and are diagnostic of disulfide bond formation. The deviation on the overlap region's chemical shift values was usually from oxidized cysteines in α-helices and reduced cysteines in the β-strands.[183] Assignment of the single random coil shift value is difficult, as the cysteine chemical shifts are sensitive to pH and solution conditions. The chemical shift values of C_α and C_β of OfurPBP2 were listed in Table 3.3. These values confirm that all the cysteine residues were oxidized.

Table 3.3: List of C_α and C_β chemical shift values for cysteine residues for undelipidated and delipidated OfurPBP2.

	Undelipidated OfurPBP2 Oxidized C_β **(**38.4±3.2 ppm to 43.0±4.2) Cα 57.6±2.3		**Delipidated OfurPBP2** Oxidized C_β **(**38.4±3.2 ppm to 43.0±4.2) Cα 57.6±2.3	
Cysteines position	**Cα**	**C_β**	**Cα**	**C_β**
Cys19	59.835	41.519	57.641	38.783
Cys50	55.245	35.048	52.527	38.765
Cys54	60.036	42.214	57.381	39.207
Cys97	57.873	39.989	55.31	36.326
Cys108	61.059	43.984	58.526	41.589
Cys117	59.98	37.983	57.475	34.763

Here, the C_β shift of Cys 50 lies near the overlap region (33-34). However, the C_β the other 5 cysteines chemical shift lies in the oxidized region. After consideration of the secondary structure, the signal at 35.048 ppm was assigned to the oxidized region.[183]

3.8. Dihedral Angles from Chemical Shifts

After the completion of the chemical shift assignments, the constraints for the backbone φ and ψ dihedral angles were generated using the TALOS+ program.[171] Alternatively, 3J-scalar coupling measurements could be used to estimate dihedral angles from empirically-defined relation.[193] The

TALOS^{+} uses chemical shift of ^{1}HN, ^{15}N, Hα, ^{13}Cα, ^{13}Cβ, and ^{13}CO and make quantitative predictions for the protein secondary structure, Φ and Ψ torsion angles, sidechain χ1 torsion angles, and hydrogen bonding patterns in secondary structure.

3.9. Conclusion

To understand the structural mechanism of pheromone binding and release in OfurPBP2 protein, we initiated the characterization of the protein by solution NMR. The backbone resonance assignments and the secondary structural elements of undelipidated OfurPBP2 at pH 6.5 were obtained using double-labeled protein with triple resonance NMR experiments. The assignments are 97% completed for the backbone and 88% completed for the side-chain resonances. The secondary structure of OfurPBP2, based on backbone chemical shifts, consists of eight α-helices, including a well-structured C-terminal helix. The secondary structure elements of undelipidated OfurPBP2 show that the C-terminus segment, His131-Met142, has an α-helical structure at pH 6.5. This observation was quite surprising and contrasted with several well-studied lepidopteran PBPs, including ApolPBP1[80], BmorPBP[63], and AtraPBP1.[166] The C-terminus of each of these PBPs exists as a coil that is exposed to the solvent in the ligand-bound conformation of the protein at pH > 6.0. However, the ligand is released at a lower pH (< 5.0) near the site of the olfactory receptor neuron through a pH-dependent conformational switch, where the C-terminus switches to a helix and outcompetes the pheromone for the pocket.[80, 63, 166]

Similarly, for delipidated OfurPBP2, we have assigned approximately 85% backbone and 80% of side-chain resonances. These assignments provide the starting point for a detailed solution state characterization of this dynamic protein by NMR methods. Although OfurPBP2 is a highly helical protein, it is also highly dynamic. The secondary structure was determined based on the chemical shift values. These resonance assignments are essential for assessing protein-ligand interactions,

and for the determination of 3D structure. The secondary structure elements of delipidated OfurPBP2 show that the C-terminus is mainly a coil and a very short helix. The residues from C-terminal residues were not observed, because of peak broadening, probably due to the intermediate exchange with the solvent. Similarly, the N-terminus residues also suffer peak broadening, which is also likely due to the intermediate exchange. NOE assignments confirmed detailed secondary components on the C- and N-terminus. The undelipidated OfurPBP2 protein in comparison with the delipidated, a significant difference was observed at the C and N termini. This study helps in the determination of the high-resolution three-dimensional structure of OfurPBP2. The data set is the foundation of studying dynamics and ligand binding properties of OfurPBP2.

CHAPTER IV

STRUCTURAL INSIGHT INTO PHEROMONE-BINDING PROTEIN 2 IN *OSTRINIA FURNACALIS* AT PHYSIOLOGICAL pH

4.1. Introduction

Olfaction plays a crucial role in perception and communication in animals. Many animals use their sensory system for survival. Lepidopteran moths have a highly efficient olfactory system to detect small quantities of semiochemicals. The olfactory system in the insect is an excellent model in neuroscience. Male moth antennae are housed in hair-like structures called sensilla, which are involved in detecting chemical signals. Pheromones are highly volatile semiochemicals released by female moths, which attract and trigger male moths for mating.[36] Male antennae are extremely sensitive to detect the female-secreted sex pheromone.[35] The highly hydrophobic pheromone molecules enter the sensillar pores of the cuticle to reach the aqueous sensillar lymph. Pheromone-binding proteins (PBPs) located in the sensillum lymph at high concentrations bind and transport the pheromone to the dendritic membrane of the olfactory receptor neuron (ORN)[40]. The ORN transduces the pheromone stimulus to a nervous signal, prompting the behavioral response of the male to locate the female moth for mating. PBPs are a class of odorant-binding proteins (OBPs) integral to the pheromone olfaction cascade of several insect orders.

Many lepidopterans are voracious agricultural pests costing billions of dollars in crop yield loss and control costs annually.[127,194] The Asian corn borer (*Ostrinia furnacalis*, Guenée) is one of the most destructive insect pests of maize and hundreds of other stored grains, cereals, and seeds.

This highly polyphagous pest causes a complete loss of crops and is widespread throughout Asia, Australia, and Oceania. Pesticides and insecticides are not effective at controlling these pests because they have developed pesticide resistance.[195] In these voracious pests, females lay eggs on crop plants, where they grow to caterpillars that bore into stems, leaves, and fruits and feed on silk and kernels, causing extensive damage. Conventional control methods, such as the use of insecticides, often prove ineffective. Insecticide application is challenging because of the height of the crop and a short window to kill eggs and larvae before they bore into the stalk. Excessive use of pesticides can result in severe environmental damage, adverse human health, development of pest resistance, and secondary pest outbreaks. Disrupting the mating process via sensory inhibition of male moths presents an alternative control method. A comprehensive understanding of PBPs, including the structural basis of pheromone recognition and the mechanisms of action of pheromone binding and release, will aid in the bio-rational development of pheromone mimics for the lepidopteran pest. However, much of this knowledge remains elusive. To unravel the mechanism of pheromone reception, we have started three-dimensional structure determination of pheromone binding protein 2 from *Ostrinia furnacalis*. Understanding pheromone reception mechanisms at the molecular level in *Ostrinia* allows for the development of novel bio-rational crop protection by manipulating insect behavior through sensory inhibition. This control strategy can potentially be applied to prevent insect vector spread of human disease.

Lepidopteran PBPs are extracellular water-soluble proteins of around 130-150 amino acids with molecular weights of 13-20 kDa. They are composed of six or seven helices surrounding a hydrophobic binding cavity. Despite over 50% sequence identity and six strictly conserved cysteine

residues that form the disulfide bonds[49] (Figure 4.1), PBPs from different moth species have species-specific substrate specificity even in a minute difference in pheromone blend ratio. The pheromone binding and release mechanism of *Antheraea polyphemus* PBP1 *(ApolPBP1)*[65,66,76,80,82,196], *Bombyx mori* (BmorPBP)[58-64,67,69,71,168,197,198], *Amyelois transitella* (AtraPBP1)[83-85,199], and *Lymantria dispar* (LdisPBP2)[51,88,137,170,200-202] have been investigated in detail.

PBPs are bound to pheromones with high affinity at neutral pH in the open conformation (also known as ligand-bound, or PBP^B conformation) where the ligand occupies the hydrophobic pocket while the unstructured C-terminus is extended to the solvent.[59,61,65,80] They undergo a conformational switch to the closed conformation (also known as the ligand-free or PBP^A conformation) at acidic pH, where the extended C-terminus is inserted inside the pocket as a helix.[10,11,18,20,28] Ligand binding and release occur through a concerted pH-dependent mechanism of two molecular switches: a histidine gate consisting of His70 and His95 and the C-terminus. [61,65,82,198] At acidic pH, PBP^B undergoes a conformational switch to PBP^A where the histidine gate opens due to repulsion between the protonated histidine residues[9,11,17], while the C-terminus inserts into the pocket as a helix ejecting the ligand through the opened histidine gate.[64,65,67,198] Conversely, PBP^A can adopt PBP^B conformation at neutral pH in the presence of a ligand.

The genus *Ostrinia* is an excellent model system to elucidate the molecular mechanism underlying pheromone specificity in the male moth olfactory system.[203] Although there are five PBPs reported in *O. furnacalis*[107,125], only OfurPBP2 and OfurPBP3 have male-biased expression and sex pheromone detection.[107] OfurPBP2 shares about 50% sequence identity to the well-studied lepidopteran PBPs and retains six strictly conserved cysteine residues. However, the majority of moth pheromone components were characterized as long hydrocarbon chains (10-18C) with acetate, alcohol, or aldehydes functional group containing 1-3 double bonds located at different

positions with variable geometric configuration.[204,205] The *Ostrinia* female sex pheromone consists of a blend of monounsaturated tetradecenyl acetate, varying in a double bond position at 9, 11, 12, and geometry cis (*Z*) or trans (*E*) between species. *O. furnacalis* has uniquely evolved to use a blend of *E*- and *Z*-12-tetradecenyl acetate (*E*-12- and *Z*-12-14: OAc) pheromone component, whereas all other Ostrinia species use *E*- and *Z*-11-tetradecenyl acetate (*E*-11- and *Z*-11-14: OAc).[110,111,114,206] This subtle change in the pheromone structure may impart the species specificity for the pheromone and pheromone binding proteins.[115]

Furthermore, OfurPBP2 has key differences in both biological gates. One of the histidine-gate residue His70 is replaced by arginine and the C-terminus of OfurPBP2 contains seven charged residues making the C-terminus more hydrophilic, whereas there are only three charged present on the other lepidopteran PBPs[61,63,80,83,88] (Figure 4.1). The mechanistic role of Arg70 in the histidine gate and the impact of four additional charged residues on the C-terminal gate remains to be unraveled. The orientation of the C-terminus in the OfurPBP2 seems to be interesting. The role of these 7 charged residues needs to be understood at the molecular level. In addition, if Arg70 plays any role in the formation of a gate that regulates the ligand binding or release in OfurPBP2, needs to be investigated. The structural and mechanistic details of acetate pheromone recognition by OfurPBP2 is an intriguing question.

To understand the impact of these key differences in the protein sequence on OfurPBP2 structure and function, we initiated a detailed NMR investigation. This Chapter discusses the solution NMR structure of OfurPBP2 at pH 6.5. The protein has a globular fold where six helices, α1a (2-14), α1b (16-22), α2 (27-37), α3 (46-60), α4 (70-80), α5 (84-100), and α6 (107-124) are supported by the disulfide bridges between residues 19-54, 50-108, and 97-117 enclose a large hydrophobic binding pocket. Our structure reveals that the C-terminus forms a well-structured helix outside the hydrophobic pocket, a major deviation from other previously reported lepidopteran PBPs.[61,63,80,83,88] Furthermore, docking studies provide important insight into similarities and differences in the

protein binding interactions to both pheromones. Molecular dynamics simulations and relaxation studies explain flexibility regions. Ligand binding studies showed the slow exchange, which is a characteristic feature for explaining the formation of the protein-ligand tighter complex. Similarly, the perturbation table was obtained by using ligand-bound and ligand-free protein chemical shift assignment showing the important residues and loops region which are affected by ligand binding.

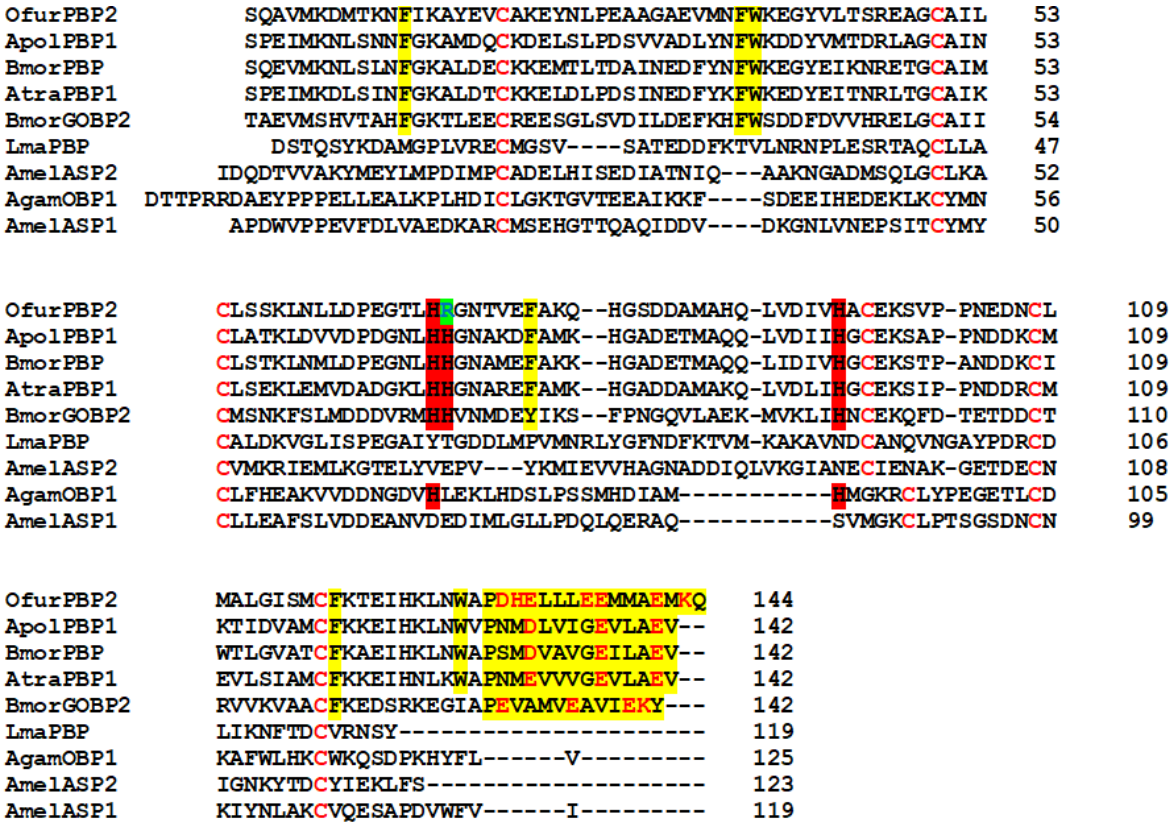


Figure 4.1. Primary sequences of the PBPs of the moths: *Ostrinia furnacalis (*OfurPBP2, Acc. Num. LC027679), *Antheraea polyphemus* (ApolPBP1, Acc. Num. X17559), *Bombyx mori* (BmorPBP, GenBank Accession Number X94987), *Amyelois transitella* (AtraPBP1, Acc. Num. GQ433364), *Bombyx mori* General odorant-binding proteinBmorGOBP2 (BmorGOBP2, Acc. Num. X94989), Cockroach *Leucophaea maderae* pheromone binding protein, LmaPBP (LmaPBP Acc. Num. AY116618), *Aedes aegypti* odorant binding protein, AgmOBP1 (AgamOBP1, Acc. Num. DQ440077), *Apis mellifera*, AmelASP2 (AmelASP2, Acc. Num. NM001011591), *Apis mellifera*, AmelASP1 (AmelASP1, Acc. Num NM001011590), Conserved cysteine residues are

shown red, histidine gate residues (His70 and His95) are shown in the red background. The C-terminal gate is shown in the yellow background, with charged residues being highlighted in red.

4.2.Experimental Procedures

4.2.1. Sample Preparation

The uniformly isotopically labeled (^{15}N and $^{15}N/^{13}C$) recombinant OfurPBP2 was expressed in *E.coli* and purified by ion exchange and size exclusion chromatography as described previously.[173,207] NMR samples used for the structure determination contained 0.4 mM protein solution in 90% H_2O/10% D_2O in 50 mM phosphate buffer at pH 6.5 containing 1 mM EDTA and 0.1% NaN_3. The delipidation of OfurPBP2 was performed by Dr. Mohanty, by modifying the protocol mentioned by Bette *et al.*[143] and Katre *et al.*[65] The details of delipidation procedures were mentioned in Chapter 2.

4.2.2. **NMR Spectroscopy and Resonance Assignment**

NMR spectra were obtained at 35 ^oC using a Bruker Avance II 800 MHz spectrometer equipped with a triple resonance H/C/N cryoprobe TCI with pulse field gradients at the National High Magnetic Field Laboratory (NHMFL) at Tallahassee, FL. For the sequential assignment of ^1HN, $^1H\alpha$, ^{15}N, $^{13}C\alpha$, $^{13}C\beta$, and ^{13}CO resonances, the following experiments were performed as previously reported[207]. 2D {^1H, ^{15}N} HSQC, 2D {^1H, ^{13}C} HSQC, 3D HNCA, 3D HN(CO)CA, 3D HNCO, 3D HN(CA)CO, 3D HNCACB, 3D CACB(CO)NH, 3D CC(CO)NH, 3D H(CCCO)NH, 3D HCCH-TOCSY, and 3D ^{15}N-edited HSQC TOCSY. The nuclei enclosed in brackets participate in magnetization transfers but their signals do not observed during experiments. ^{15}N-edited HSQC-

NOESY and ^{13}C-edited HSQC-NOESY spectra were collected with two different mixing times of 85 ms and 120 ms respectively. The ^{13}C carrier frequency was set to 44 ppm and 125 ppm for the aliphatic and aromatic region, respectively on ^{13}C-edited NOESY. The details of NMR experiment parameters were listed in Chapter 3. The NMR data were processed using NMRPipe[174] and analyzed with NMRFARM sparky.[208] The secondary structural elements were calculated using TALOS^{+}[171] and CSI[190], based on assigned chemical shifts. The resonance assignments of the backbone were obtained with standard triple resonance spectra and confirmed by ^{15}N-edited HSQC-NOESY spectra. The overall secondary structure was obtained from the chemical shift index and NOE connectivity.

4.2.3. NOE Assignments and Structure Calculation

The structure determination of a protein by NMR relies mainly on the distance restraints derived from the NOESY spectra(Figure 4.2 and 4.3). NOE peaks were manually peak-picked. From the assigned NOESY spectra, NOE peaks list files were used as distance constraints. Along with the assignment of these cross-peaks intensities/volumes are used as distance information from the NOESY spectra in XEASY format. The TALOS+ chemical shift file was used in the form *.prot format. The angle restraints *.aco file was generated from TALOS+. The distance constraints were converted into upper limits of distance constraints as *.upl files. During the structure calculation process, both (*.upl and *.peak) files were used during an iterative process. During each of the CYANA run, the output files (*.ovw) were analyzed for possible improvement in the next cycle(Figure 4.4). The violations were listed in *.ovw files which were addressed one at a time. After each run, the target function and number of violations were checked from *.ovw files (Figure 4.2). The assignments were confirmed or corrected with a CYANA module, using the standard protocol consisting of seven cycles of iterative NOE assignment and structure calculation.[209] CYANA algorithm is based on the fast torsion angle dynamics and simulated annealing coupled to a temperature bath which is cooled down slowly from its initial high temperature. A total of 260

upper-limit constraints based on backbone dihedral angles (φ and ψ) and 198 side-chain dihedral angles (χ1 and χ2) were obtained from TALOS^{+}. The standard upper and lower distant constraints for three disulfide bonds were used. The experimentally derived upper distance restraints were derived from the NOE intensities of the assigned peaks. There are 402 long-range NOEs. The preliminary structure was calculated without the incorporation of restraints from disulfide bridges. The structure was found to converge with the target function 3.72 ±0.35 Å^2, and the resultant 20 conformers had rmsd (Ala3-Asp130) of 0.72 Å and 1.21 Å for backbone and heavy atom, respectively. The structure calculation showed that the disulfide topology was possible based on the observed NOEs. The six cysteine residues were paired unambiguously between residue Cys19-Cys54, Cys50-Cys108, and Cys97-Cys117 in structure calculations. These well-conserved pairs of disulfide bonds were determined by biochemical methods in BmorPBP[169,210]. Disulfide restraint was introduced for the final structure calculation. The default calibration functions 6.7E6 and 8.0E4 were used for the backbone and side-chain resonance respectively. The length of the annealing was 10,000 steps. The random seed values of 434726 were taken. The calculation was started with 100 randomized conformers and the 20 best-calculated CYANA structures with the lowest target function values obtained from the final cycle were used for energy minimization with the explicit solvent with YASARA[211] using YASARA force field.[212] The 20 structures with the lowest potential energy and Ramachandran plot score were assessed by Molprobity[213] and PROCHECK.[214] The structures were visualized with VMD, and figures were created using Pymol and Chimera. The average root means squared deviations (rmsd) from the idealized geometry for the bonds and angles are 0.009 Å and 1.5 Å, respectively. None of the distance constraints were violated by more than 0.40 Å and there were no dihedral angle restraint violations. The global rmsd deviation value relative to the mean coordinate was 0.46 Å, calculated for the backbone residues from Ala3-Asp130 and 1.10 Å for all heavy atoms. The quality of the model was analyzed by using PROCHECK[214], which shows 97.7 % of the residues were in the most favored region, and the rest were in the

additionally allowed region of the Ramachandran plots shown in the Figure results section. The complete statistics are given in Table 4.1.

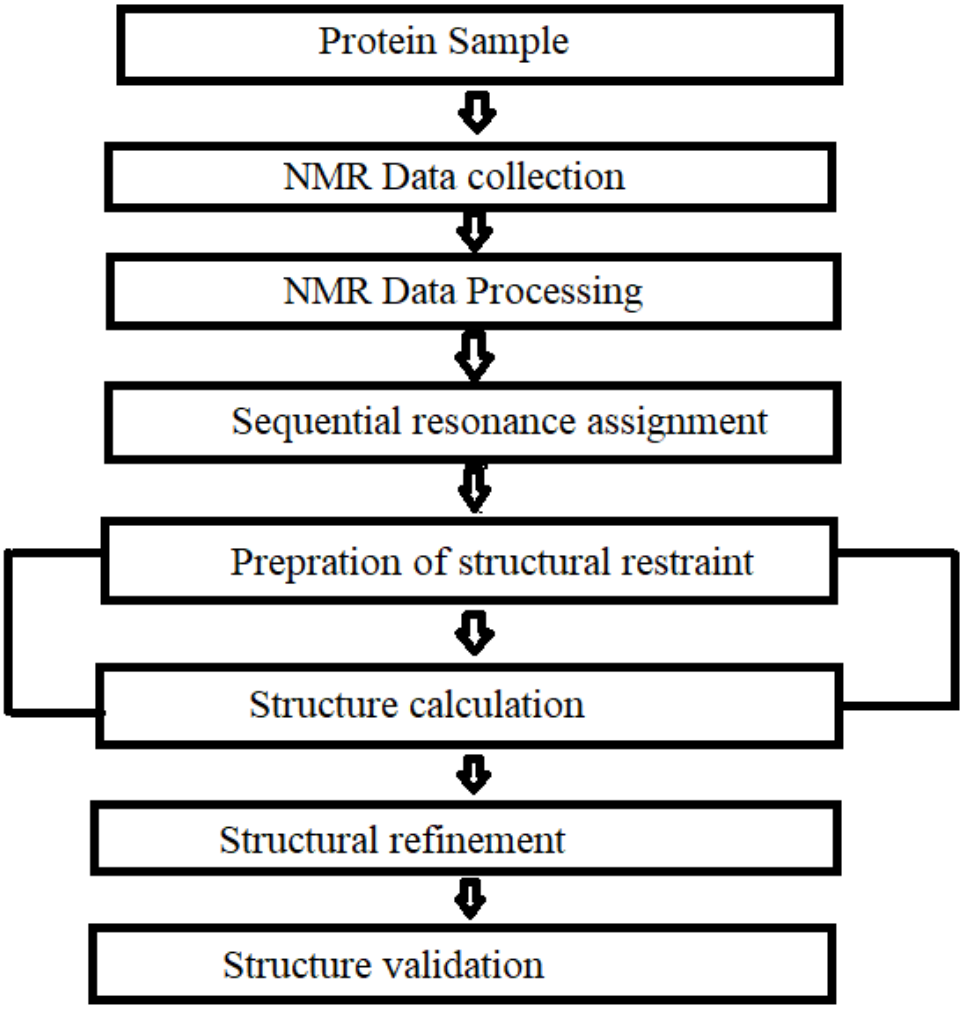


Figure 4.2 : Flow chart showing the steps for the structure calculation.

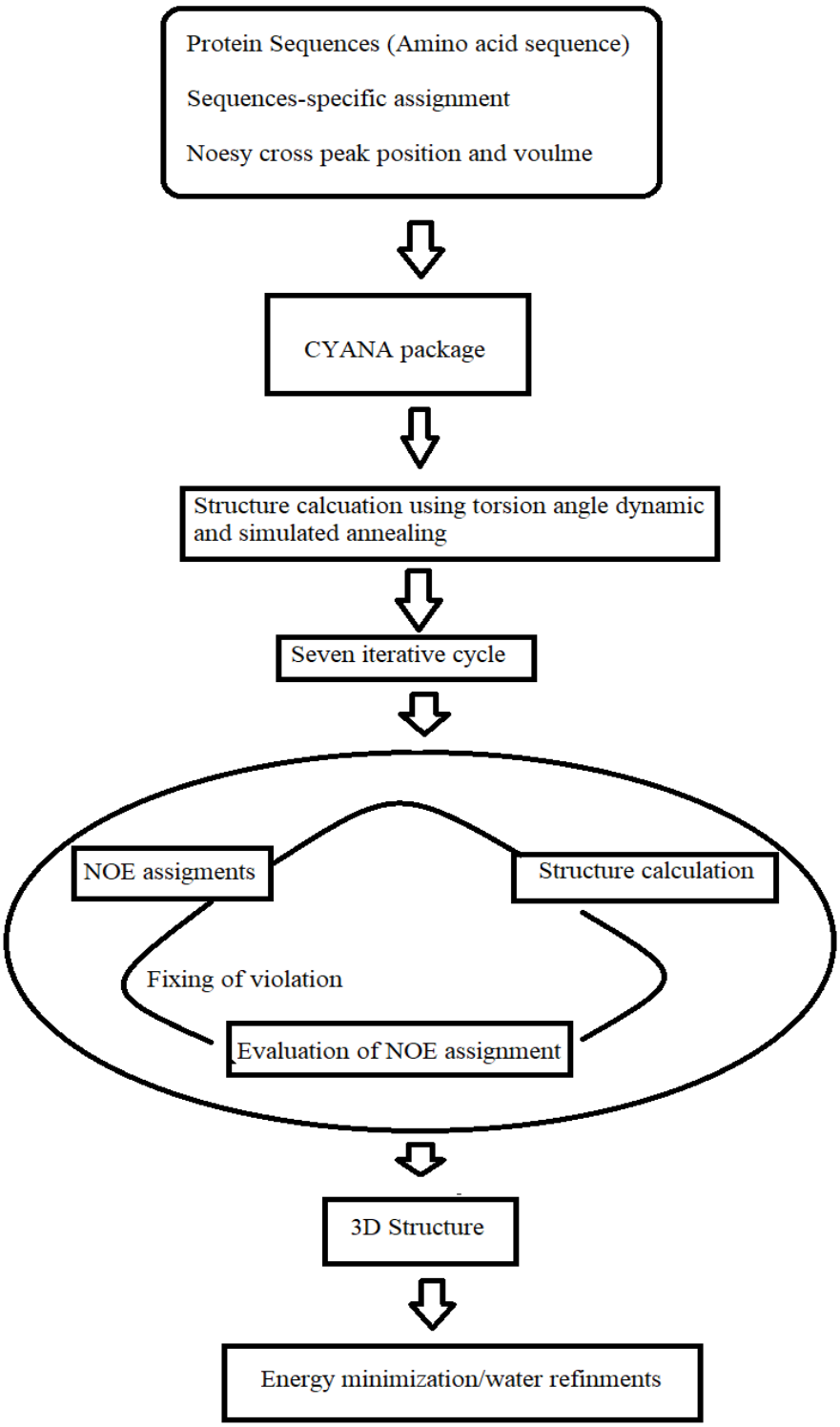

Protein Sequences (Amino acid sequence)
Sequences-specific assignment
Noesy cross peak position and voulme
CYANA package
Structure calcuation using torsion angle dynamic and simulated annealing
Seven iterative cycle
NOE assigments
Structure calculation
Fixing of violation
Evaluation of NOE assignment
3D Structure
Energy minimization/water refinments

Figure 4.3: General scheme of NOESY assignment and structure calculation using CYANA.

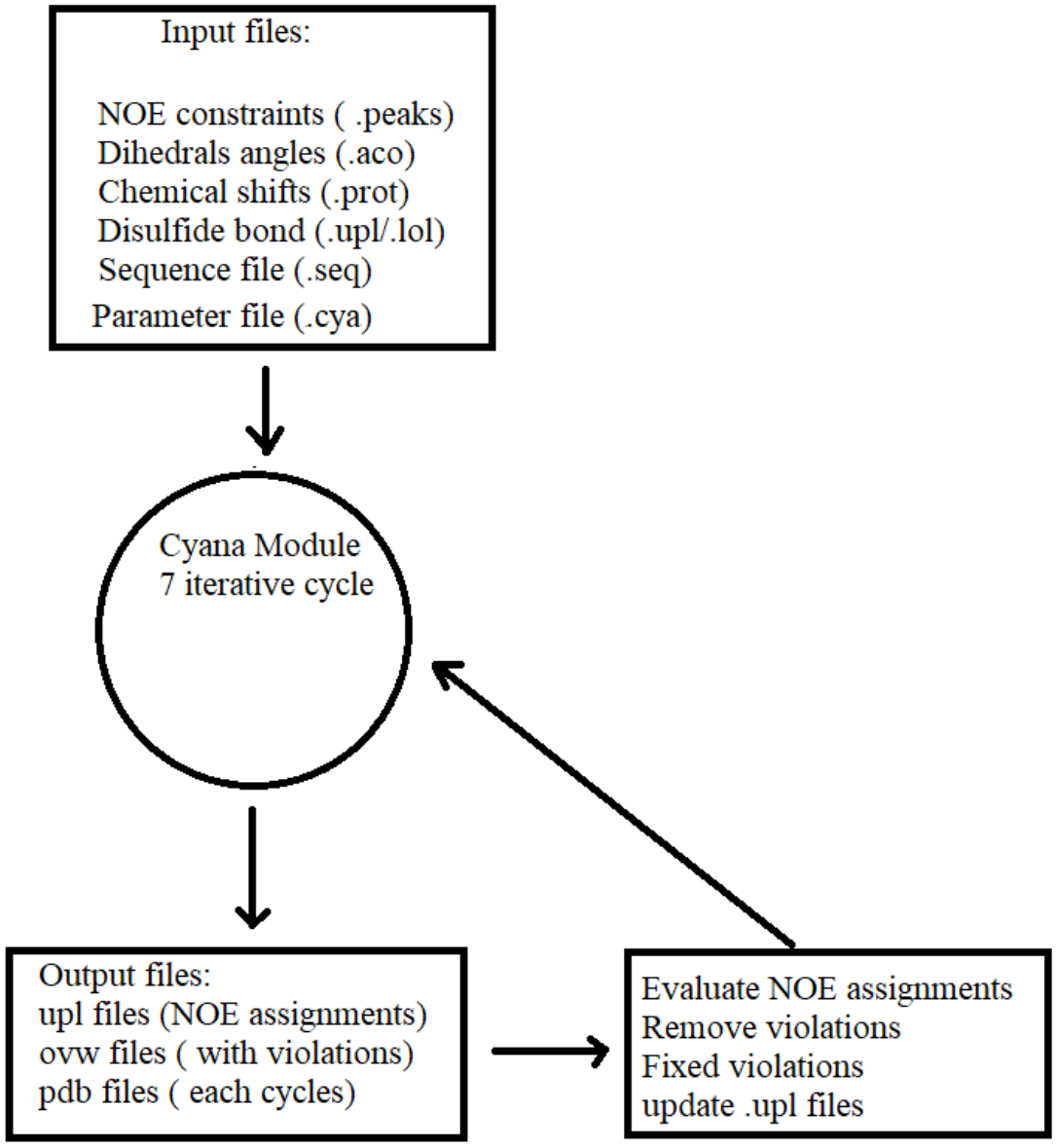


Figure 4.4: General scheme of input and output for the structure calculation using CYANA.

4.2.4. Statistics and Visualization

Visualization, root-mean-square distance, hydrogen bond, and helix packing angle calculations were performed with the program PYMOL, Chimera, and VMD. Ramachandran plot statistics were calculated by PROCHECK. The internal cavity was analyzed with the program Computed Atlas of Surface Topography of Proteins.[215]

Table 4.1: Restraints & Structural Statistics

Property	Value
Restraints	
Total experimental distance restraints	2557
Sequential li-jl <1	1314
Medium range 1< li-jl < 5	841
Long range li-jl > 5	402
Dihedral angle restraints (Φ and Ψ)	260
Hydrogen bond distance restraints (a*)	80
Residual restraint violations after simulated annealing	
Distance restraint violation greater than 0.01 A˚	6.0
Maximum distance restraint violation (A˚)	0.33
Dihedral angle restraint violations	0.0
RMS deviations from the averaged coordinates (A˚)	
The backbone of the regular secondary structure	

RMSD from ideal geometry, Bond length (Å)	0.009
Bond angles (°)	1.5°
All heavy atoms of the regular secondary structure	
Backbone of the residues 1–130	0.48
All heavy atoms of the residues 1–130	1.10
Residual target function value (A˚2)	2.92
Total energy KJ/mol	-18360.68
Ramachandran plot statistics (%)	
Most favored regions	97.7
Additionally allowed regions	2.3
Generously allowed regions	0.0
Disallowed regions(b*)	0.0

4.2.5. Molecular Docking Studies

Molecular docking studies were performed to predict the potential binding mode and to estimate the free energy of binding of the OfurPBP2 and the pheromones molecule. Docking was performed using AutoDock [216] and AutoDock Vina.[217] AutoDock Tools 1.5.6 was used to create the PDBQT (Protein Data Bank, Partial Charge (Q), & Atom Type (T)) format. The PDBQT format is similar to the PDB format but it includes partial charges (Q) and atom types (T). Polar hydrogens were added using the AutoDock Tools interface. The Kollman charges were added. Similarly, for the ligand, all the torsion was released except around the double bond and saved in PBDQT format. The default AutoDock force field was used.[218] The whole protein was covered by a grid box with a spacing of 0.375 Å and saved as (*.gpf) file format. The docking parameter file was prepared and saved as (*.dpf) file format. Four different files: protein.pdbqt, ligand.pdbqt, grid.gpf, and

parameter.dpf files were saved on the working directory. The step-by-step protocol was followed as explained by Rizvi SM et.al.[219] The ligand was subjected to 100 Lamarckian genetic algorithm[220], which allows handling a large number of degrees of freedom with 25 x 10^6 evaluations in each and the rest of the parameters were default. The root means square deviation (rmsd) tolerance of the resulting docked structures was ≤ 2 Å. AutoDock performs cluster analysis based on all-atom mean square deviation (rmsd). The resulting families of docked conformations were ranked according to increasing energy. The pose with the most populated cluster was selected for analysis.[221]

4.2.6. Molecular Dynamics Simulation

Interaction studies of the pheromone molecules (*E*-12-tetradecenyl acetate and *Z*-12-tetradecenyl acetate) with the protein were performed by Molecular Dynamics (MD) simulation using the GROMACS v5.1 software package.[222] Coordinates for pheromone molecules were generated using Discovery Studio v17.2.0.1.16349. The pheromone molecules were sketched and edited to provide the correct geometry by using the sketch and edit window of the Discovery Studio, and the coordinates were saved. Topology files for the pheromone molecules were obtained from ATB topology builder.[223] The topology file of the protein was prepared using the pdb2gmx tool incorporated in GROMACS using the Gromos54a7 force field.[224] The coordinates of protein and ligand were merged, solvated with simple point charge (SPC) water molecules, energy minimized, and equilibrated. The covalent bond lengths were constrained using the LINCs algorithm[225], and the time step was set to 0.002 ps following a published protocol.[226] The molecular dynamics (MD) simulation was carried out for 150 ns. All simulations were performed using the Cowboy high-performance computer (HPC) at Oklahoma State University. Pymol was used to analyze the protein and ligand interactions.

4.2.7. Ligand Titrations Study by NMR

For ligand titration experiments, uniformly ^{15}N-labeled delipidated OfurPBP2 (300 μL of 530 μm protein in 50 mm phosphate buffer pH 6.5, containing 5% D_2O, 1 mm EDTA, and 0.01% (w/v) NaN_3) was titrated with increasing concentrations of pheromones and the corresponding two-dimensional {^1H, ^{15}N} HSQC spectra were recorded. For the titration studies, 50 mM stock solutions of pheromones *E*- and *Z*-12-tetradecenyl acetate (*E*-12- and Z-12-14: OAc) were prepared in methanol. We took protein sample into the NMR shape tube, required volume of the pheromone (Table 4.2) was added into the protein sample in the NMR shape tube with the help of a long gel loading tip and mixed properly by inverting the tube 3 to 4 times and incubate for 30 minutes at room temperature. After the addition of the required pheromone, the corresponding two-dimensional HSQC spectra were recorded at each titration point and were processed using NMRPipe and analyzed using Sparky.[227]

Table 4.2: Protein and pheromone ratio used in NMR titration experiments

Protein:Pheromone	Pheromone added (ul)	Total pheromone added (ul)
1:0	0.0 ul	0.0 ul
1:0.2	0.64ul	0.64 ul
1:0.6	1.26 ul	1.90 ul
1:1	1.28 ul	3.18 ul
1.2	3.18 ul	6.36 ul
1:4	6.36 ul	12.72 ul
1:6	6.36 ul	19.08 ul
1:10	12.72 ul	31.8 ul

4.2.9. PDB and BRMB Accession Codes

The atomic coordinates of OfurPBP2 have been deposited in the Protein Data Bank† (accession code 6XCW) and BMRB ID 30762. The assigned chemical shifts have been deposited in the BioMagResBank‡ (accession code 57004).

4.3. Results and Discussion

4.3.1 Resonance Assignments and NMR Structure Determination

The {^{1}H, ^{15}N} HSQC NMR spectrum of ^{13}C/^{15}N doubled-labeled OfurPBP2 at pH 6.5 displayed the expected number of amide resonances corresponding to the number of amino acids in the protein indicating that the protein is homogenous. The large dispersion of chemical shifts indicates that the protein is stably folded.[173] The sequence-specific chemical shifts of ^{1}HN, ^{15}N, ^{13}Cα, $^{13}C_{\beta}$, and ^{13}CO were assigned with three-dimension triple-resonance experiments as reported in Chapter 3.[207] Both backbone and side-chain chemical shift assignments have been deposited to the BioMagResBank (BMRB) repository (accession no. 50074). The observation of strong sequential dαδ NOEs indicates that all prolyl residues are in the trans conformations.[228] All six cysteine residues are in the oxidized state, as indicated by their $^{13}C_{\beta}$ chemical shifts. Disulfide bonds were determined by NOE patterns between the linked Cys residues and by the characteristic β-methylene ^{13}C chemical shift[229]. The NOEs were observed between the β-protons of the disulfide-linked Cys residues which provide the evidence for disulfide bond connection.[230] One of the examples of determining the connectivity of NOEs across disulfide bonds was explained by Takeda et al.[231] The NOEs between carbon H_{β}/H_{β} and carbon Hα/Hβ were shown to have positive predictive values for the characterization of disulfide links.[232] The three-dimensional structure of OfurPBP2 was derived from the NMR chemical shifts and NOE data. The structure obtained had a resolution of 1.5 Å with root mean square deviation (rmsd) of 0.48 Å and 1.1 Å for backbone and heavy atoms, respectively.

Table 4.1 summarizes the structural statistics of the 20 lowest energy conformers. The atomic coordinates obtained from three-dimensional structures of OfurPBP2 have been deposited in the RCSB Protein Data Bank. The strong and medium-range NOE connectivity data indicated the presence of seven helices, with the C-terminal helix, α7, formed by the polypeptide segment 131–142. Lack of long-range NOEs, from His131 to Gln144 implies a flexible C-terminus; however, several NN (i, i + 2), αN (i, i + 2), αN (i, i + 4), and αβ (i, i + 3) NOEs in this region confirmed an α-helical C-terminus (Figure 4.5). The amide signals of residues Trp37, Glu39, and Thr44 exhibited peak splitting due to slow exchange as observed in the 3D ^{15}N-resolved the {^{1}H, ^{1}H} NOESY spectrum. Line broadening was observed in Lys6, Arg46, and Arg70 in the {^{1}H, ^{15}N} HSQC, but they were readily assigned in ^{15}N- edited {^{1}H, ^{1}H} NOESY spectra. In NOESY spectra of an α-helical protein, strong and medium intensity d_{NN}(i, i ± 1) NOEs are usually found. For backbone atom assignment, d_{NN} NOEs provide sequential connectivities.[233] The amide proton range lies in the 6.5 to 11.5 ppm range. The amide-amide cross-peak pattern of dNN(i -1, i), diagonal peak, and dNN(i, i+1) cross-peaks for residue i, clearly indicate that residues are adopting α-helix.[233] All the seven helices were well characterized by the numerous dNN(i,i+1), dNN(i,i+2), dαβ(i,i+3), dαN(i,i+3) and dαN (i,i+4) NOEs and further supported by continuous sequential NN NOEs connectivities, which are the diagnostic features for the formation of an α-helix[234] as shown in Figure 4.5. The strong dNN(i,i + 1) NOE together with a dαN(i,i + 3), dαβ(i,i + 3), and dαN(i,i + 4) and a weaker dαN(i,i + 1) clearly indicate that the C-terminus of OfurPBP2 is in an α-helical configuration. NOEs between amide protons of consecutive residues and between Hα and the amide proton of subsequent residues are represented by bars connecting the residues. For NN(i, i+1) and αN(i, i+1) NOEs, the thickness of the bar qualitatively represents the relative intensity (weak, medium, or strong) of the NOE (Figure 4.5). Similarly, the antiparallel β-stand only short distance strong dNN(i, i+1) and dαN(i, i+1) NOEs are observed from residue Leu61 to His69. There is the presence of inter-strand NH–NH, NH–αH, and αH–αH NOEs between residues are the indication of antiparallel β –strand[235].

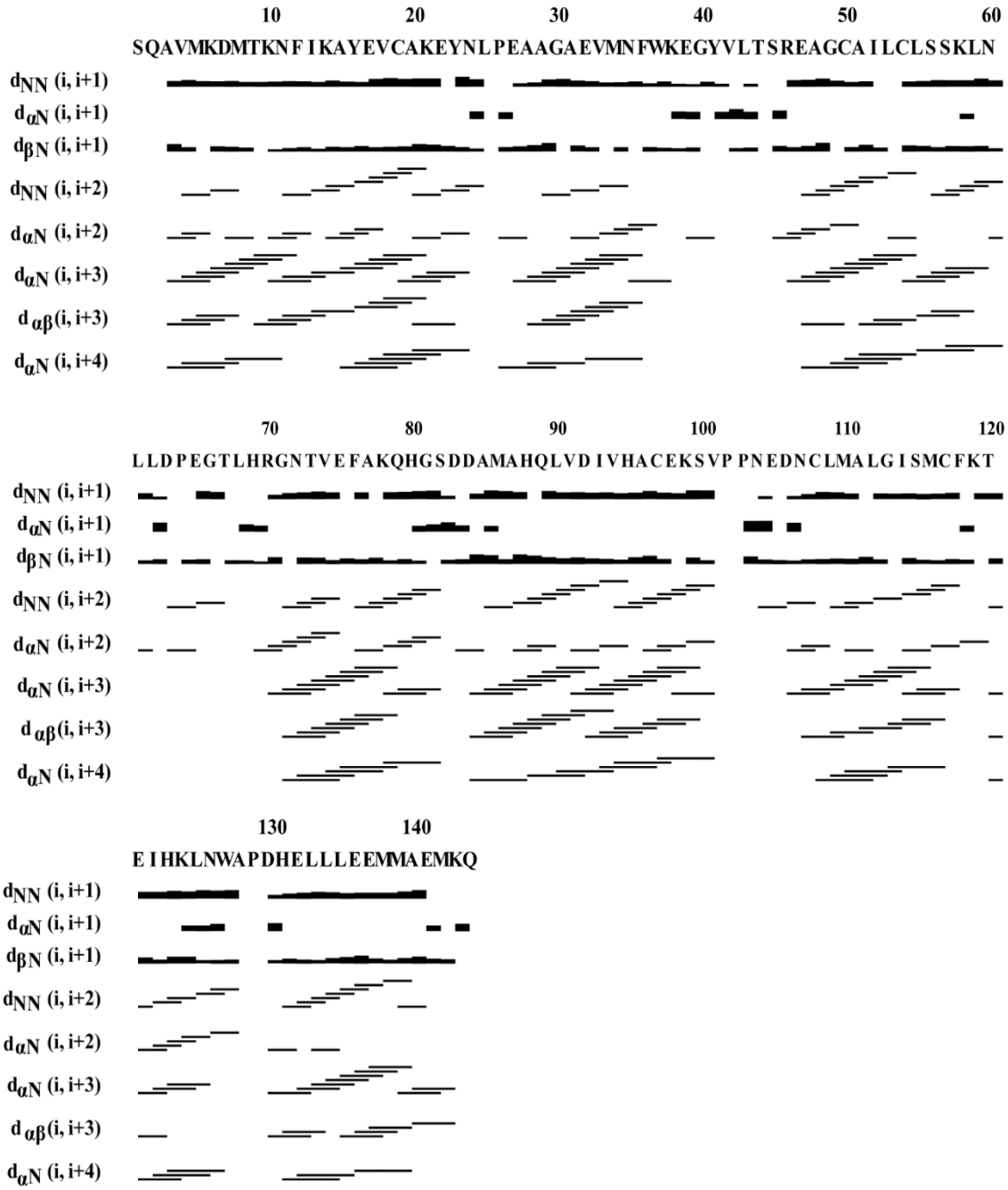


Figure 4.5: Summary of secondary structure information from sequential and medium-range NOEs for OfurPBP2. NOE intensities are represented by the line (indicated with thick or thin lines for strong or weak NOEs, respectively). The medium-range connectivity is shown by lines starting and ending at the positions of the residues related to the NOE.

4.3.2 Water Refinement

The NMR solution structure of a protein is represented by an ensemble of 20 conformers. The 20 conformers with the lowest target function obtained from the final cycle from the cyana were used for energy minimization. The protein structures are generally calculated in a vacuum. Water has a significant effect on protein structures. Refinement was done by molecular dynamics simulation in water. YASARA runs molecular dynamics simulations of models in explicit solvent using a knowledge-based all-atom force field. The server performs an energy minimization using the YASARA force field[212] derived from Amber whose parameters have been optimized as described by Krieger et.al.[236] The PDB file was used as input. The server prepares the PDB/structure for simulation, predicts amino acid side-chain rotamers with the SCWRL algorithm[236,237], based on the steepest-descent methods. Hydrogen bonding networks were optimized and clashes were removed followed by a simulated annealing process to reach energy minimum. The YASARA View was downloaded to visualize the result and the file format obtained from YASARA was then converted into PBD format (Figure 4.6).

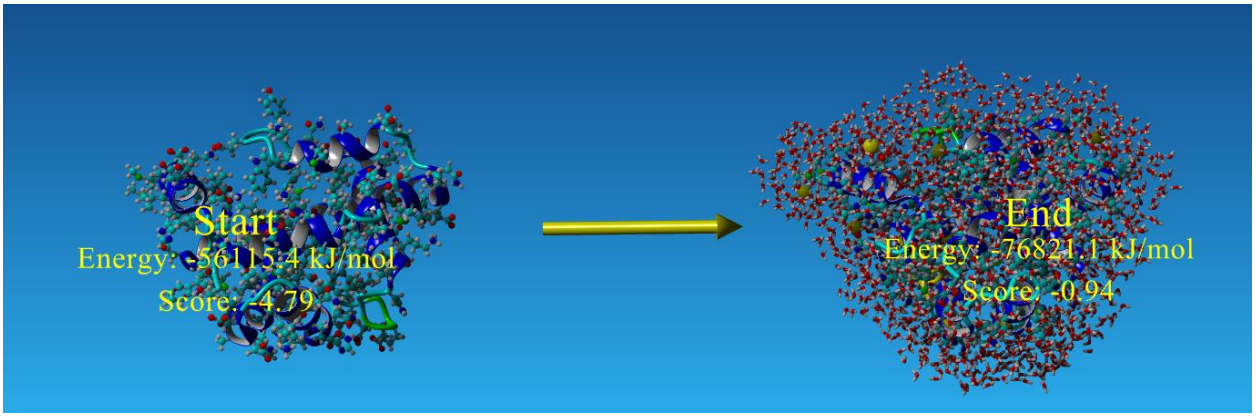


Figure 4.6: The 20 conformers with the lowest residual target function values obtained from the final cycle were used for energy minimization with the explicit solvent with the YASARA

module[211], using the YASARA force field.[212] It shows that the overall free energy and scoring function after energy minimization is much improved.

4.3.3 Ramachandran Plot

The Ramachandran plot is a structure validation tool used in structure determination. The quality of the structure can be assessed by the statistical distribution of (ϕ and ψ) dihedral angles. The Ramachandran plot visualizes energetically allowed and forbidden regions of the dihedral angles. The algorithm is based on the combinations of (ϕ and ψ) values in a structure and compares them with the commonly observed values in high-resolution crystal structures. Furthermore, PROCHECK[238] was used to check the stereochemical quality and overall residue-by residue geometry of the protein structure. PROCHECK is a program used to analyze the stereochemical quality of the models in the NMR ensemble. The input consists of solvent refined coordinates file of the protein structure. The program provides the summary of text files, postscript files, and geometry plots as an output. The output consists of a summary of the analysis of planarity of peptide bonds, bad non-bonded interactions, distortions of the geometry around the Cα atoms, energies of hydrogen bonds. The geometry plot consists of Ramachandran plots, torsion (ϕ and ψ) angle plots. It shows that the most favored region is 97.7% and the additional favored region 2.3% (Figure 4.7). PROCHECK also provides the atomic resolution of the NMR structure, which was reported to be 1.5 Å.

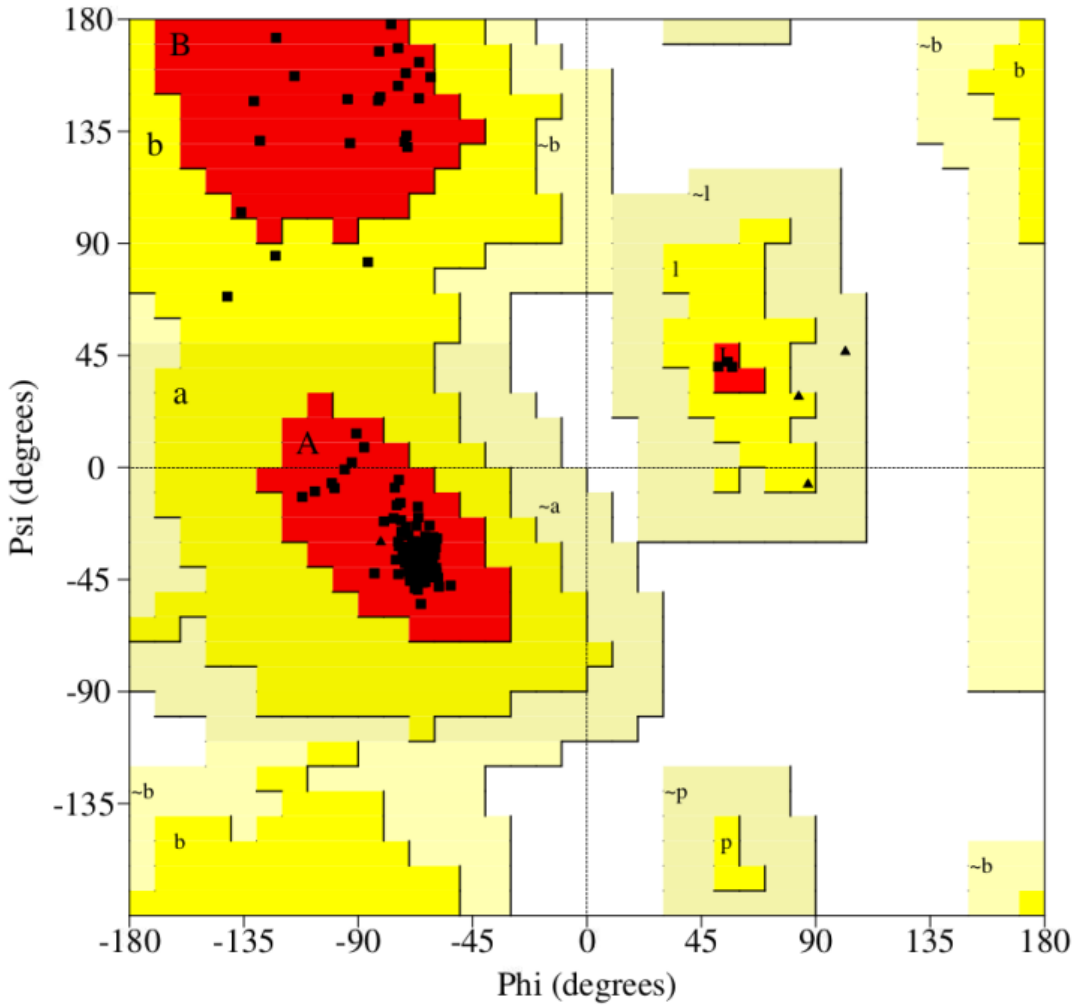


Figure 4.7: Ramachandran plot showing (ϕ and ψ) dihedral angles values. It shows that the most favored region is 97.7%, and the additional favored region is 2.3%. This plot and values were obtained from PROCHECK.

4.3.4 Effect of Delipidation

As previously reported, OfurPBP2 undergoes a reversible pH-dependent conformational change.[173] Both the circular dichroism (CD) and the NMR show a loss of tertiary structure at low pH. At neutral pH, the protein is folded into a compact globular conformation[173], however, a

conformational transition occurs between pH 5.0 and 4.5 as indicated by the 2D {^{1}H, ^{15}N} HSQC. The protein remains structured at pH 5.5 and above without major changes in the chemical shift. Peak broadening in a poorly dispersed {^{1}H, ^{15}N} HSQC spectrum indicates the presence of a molten-globule state. Molten globules exist under mild denaturing conditions, such as acidic pH, or when cofactor or ligand is removed.[161,239] At low pH, OfurPBP2 is likely in a molten globule state.

Interestingly, the delipidated (lipid-free) OfurPBP2, in which the endogenous ligand from the bacteria is removed[65], behaves similarly at pH 4.5. The overlay of {^{1}H, ^{15}N} HSQC spectra shows that the delipidated protein at pH 6.5 and undelipidated (lipid-bound) protein at pH 4.5 is similar. In this molten globule state, there results in a fluctuation of ensembles on the order of milliseconds to microseconds.[162] The loss of resolution and extreme line broadening in the {^{1}H, ^{15}N} HSQC spectra is likely due to the fluctuation of partially folded protein states on the millisecond to microsecond time scale.

4.3.5 The Overall Structure of OfurPBP2 at pH 6.5

The 144 amino acid residue OfurPBP2 structure at physiological pH consists of tightly packed globular arrangements of six α-helices with residues 2-14 (α1a), 16-22 (α1b), 27–37 (α2), 46-60 (α3), 70-80 (α4), 84-100 (α5), 107-124 (α6) (Figure 4.8). The C-terminus consists of the seventh helix with residues 131-143 (α7). The loops that interconnect the helices, are named L1, L2, L3, etc., and are used to name the loop. The protein has approximate dimensions of 41×38×46 Å, formed by a roughly conical arrangement of six α helices. Three helices, α3, α5, and α6, and loop 3 (L3), converge to form a binding pocket. The converging ends of the helices form the opening of the pocket, and the opposite end lies on the interface between the helices, α1, α2, and L2. The overall folding of helices on a three-dimensional structure is anchored and stabilized by three disulfide bonds and a large number of noncovalent interactions. These six cysteine residues

are conserved throughout the PBPs and form three interlocked disulfide bridges, which provide stability to the PBPs tertiary structure. The α1, α3, and α6 helices are anchored by two disulfide bonds between Cys19-Cys54 and Cys50-Cys108. Similarly, the third disulfide bond, Cys97-Cys117, connects α5 and α6, which provide structural rigidity to the protein structure. The helices are packed in a globular structure from residues 1-130, enclosing a large hydrophobic cavity inside. Stereoviews of the superposition of 20 lowest energy minimized structures and a ribbon diagram of the overall structure is shown in Figure 4.9. The N-terminal segment, consisting of residues 3-22, is helical but slightly distorted in the middle of the helix at Tyr16 as observed in AtraPBP1[85] and BmorPBP.[61] This arrangement is represented as α1a and α1b. These N-terminal peptides have propensities to form an amphipathic helix with the hydrophobic residues pointing towards the protein core, similar to AtraPBP1.[85]

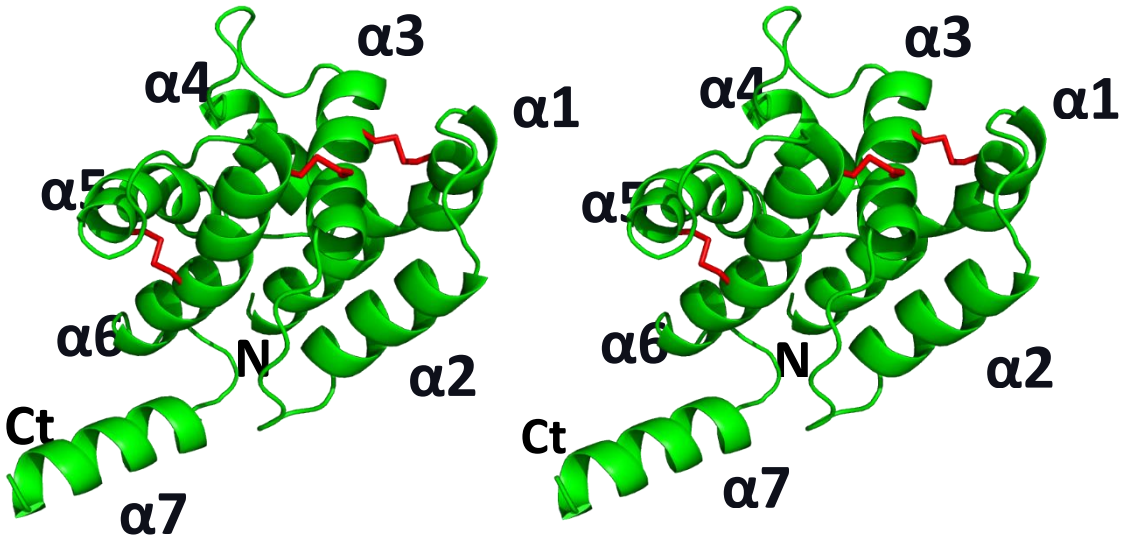


Figure 4.8: Stereoviews of the three-dimensional structures of OfurPBP2. Ribbon drawing of one of the OfurPBP2 structures. Helices, N and C termini, and disulfide bonds (red) are indicated.

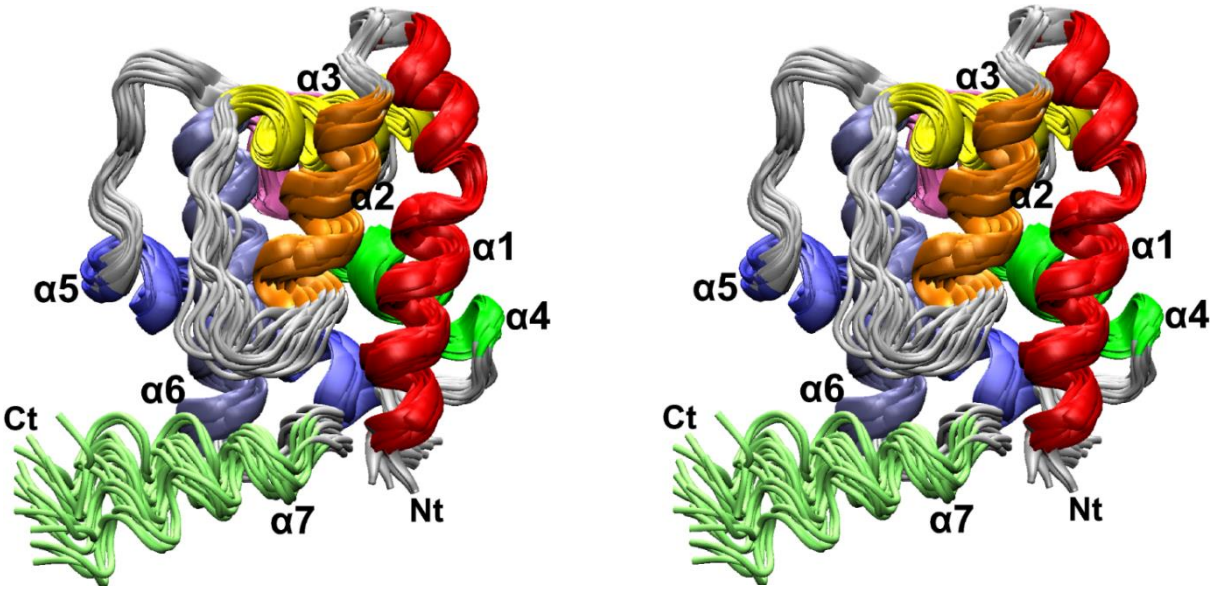


Figure 4.9: Stereoviews of a superposition of the three-dimensional structures of OfurPBP2. Superposition of the 20 energy-minimized and water-refined conformations of OfurPBP2. Backbone residues in the helical regions are shown in yellow and loop residues are shown in red.

The most interesting and noticeable feature of the OfurPBP2 structure is the C-terminus region. The C-terminal region from residues 131 to 143 forms an α-helix (α7), which is located outside the hydrophobic core. Although the helical region is well defined by sequentially neighboring amide protons dNN (i, i+1), there are no long-range NOEs (Figure 4.10) between protons of this helix with other residues. Thus, the C-terminal amino acid has fewer constraints, thus making it more flexible. It is hanging outside the core of the protein and solvent-exposed. The C-terminal helical is fluctuating and potentially associated with ligand binding and releasing function as reported for ApolPBP1[76] and LdisPBP2.[137]

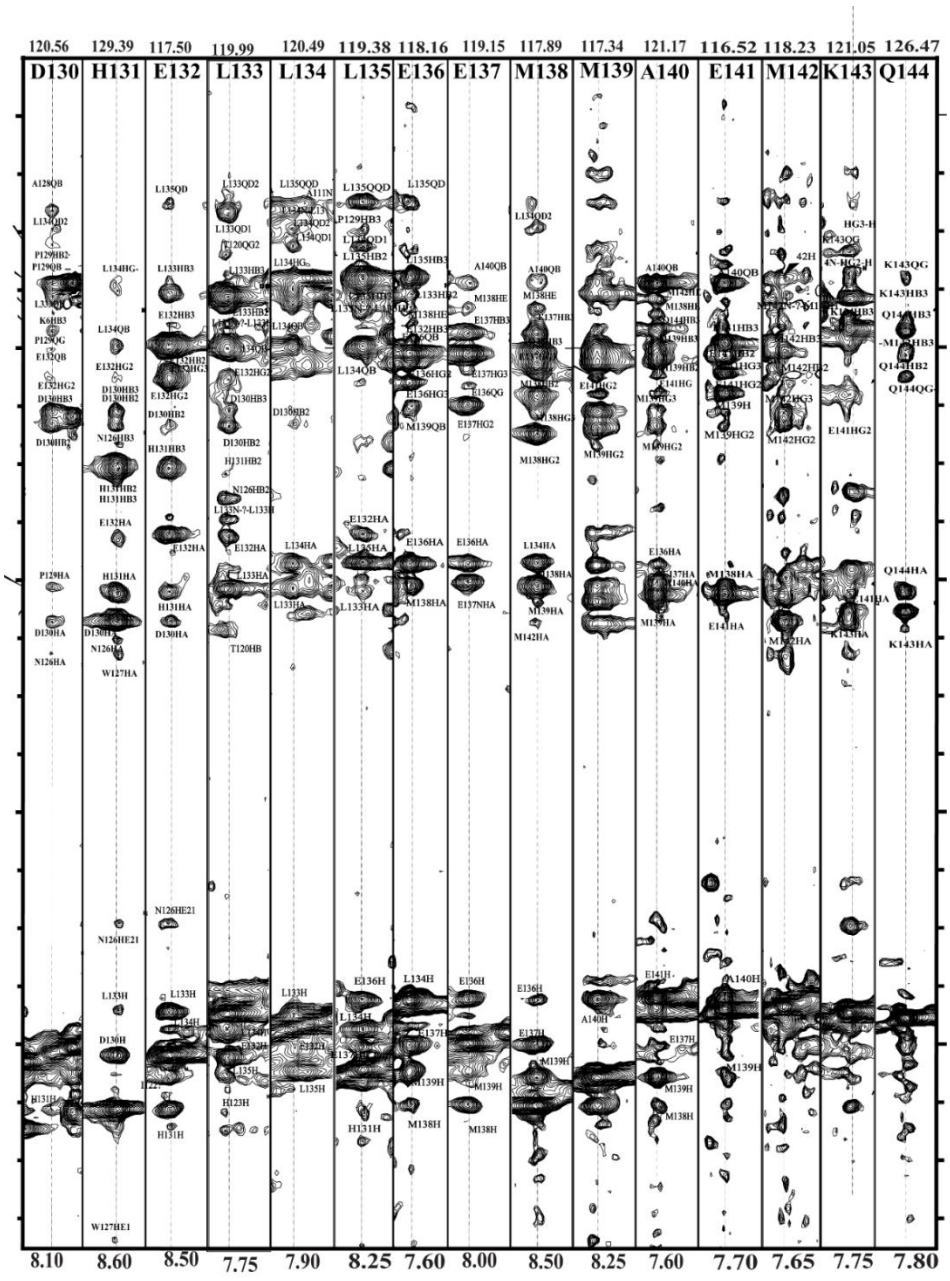


Figure 4.10: Strip plot showing ^{15}N HSQC-NOESY illustrating the NOE connection from Asp130 to Gln144.

The helices were packed closely at the crossing angles of 34° between (α1b–α2), 88°(α2–α3), 81°(α3–α4), 57°(α4–α5), 54° (α5–α6). The helix α1 is slightly distorted at 27° (α1a-α1b); this is likely due to the disulfide bridge between Cys19-Cys54, pulling the 1αb helix toward α3, resulting in the distortion. Residues Thr73, Val74, and Ala77 of α4 and Ala87, Leu90, and Val91 of α5 pack in a knobs-into-holes fashion, (Figure 4.11A), while residues Ile93, Val94 and Cys97 of α5 and Cys117, Phe118 and Glu121 of α6 are packed in a ridges-into-grooves arrangement (Figure 4.11B). These residues are well-conserved in lepidopteran PBPs, including BmorPBP.[61]

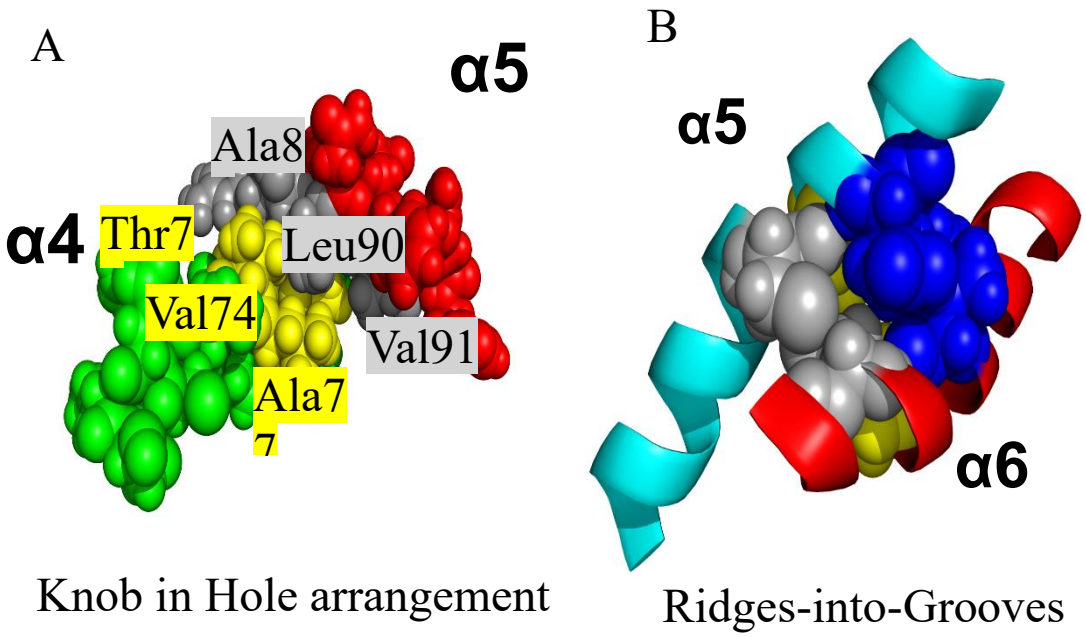


Figure 4.11: Structural features of OfurPBP2; A) Knob- in-hole arrangement between α5 and α4, B) Helices α5 and α6 pack in a ridges-into-grooves fashion with a 54° packing angle using residues Ile93, Val94 and Cys97 of α5 and Cys117, Phe118 and Glu121 of α6 to form a hydrophobic assembly in the contact region.

Helices α1 and α3 pack at an angle of 77° with small inter-helix contact through hydrophobic interactions between Tyr16, Try23, Cys50 of α1 and Cys54, Ala51, and Leu55 of α3. There is also an electrostatic interaction between Glu22 (α1) and Lys58 (α3). The packing angle between α2 and α3 is 88°, the residue Tyr16 and Cys19 from helix α1, Val33 from α2 and Ala48, Ala51, Ile52, and Leu55 from α3 formed a tiny hydrophobic core. Helices α3 and α6 cross at 84°, and the residues Gly49, Ile52, Leu53 form α3 forming hydrophobic interactions with Cys108 and Ala111 of α6. The side chain of Lys119 from α6 forms a cation-pi interaction with the aromatic ring of the Phe36, which is the last residue of the α2, responsible for maintaining the extra stability of the helices.

The primary sequence of OfurPBP2 contains 7 aspartates, 15 glutamates, 11 lysines, 2 arginines, and 6 histidine residues. Out of these charged residues, six are involved in the formation of a salt bridge: Glu22-Lys58, Arg46-Asp106, and Asp84-Lys78. The two salt bridges between Glu22-Lys58 and Arg46-Asp106 are common between OfurPBP2 and ApolPBP1[138]. Two polar amino acids are partially solvent accessible (His123, Lys119), while the remaining are found at the surface of the protein and are fully solvent-accessible which contributes to the solubility of the protein in the water. The high solubility of the OfurPBP2 is due to the exposure of charged residues on the surface of the protein.

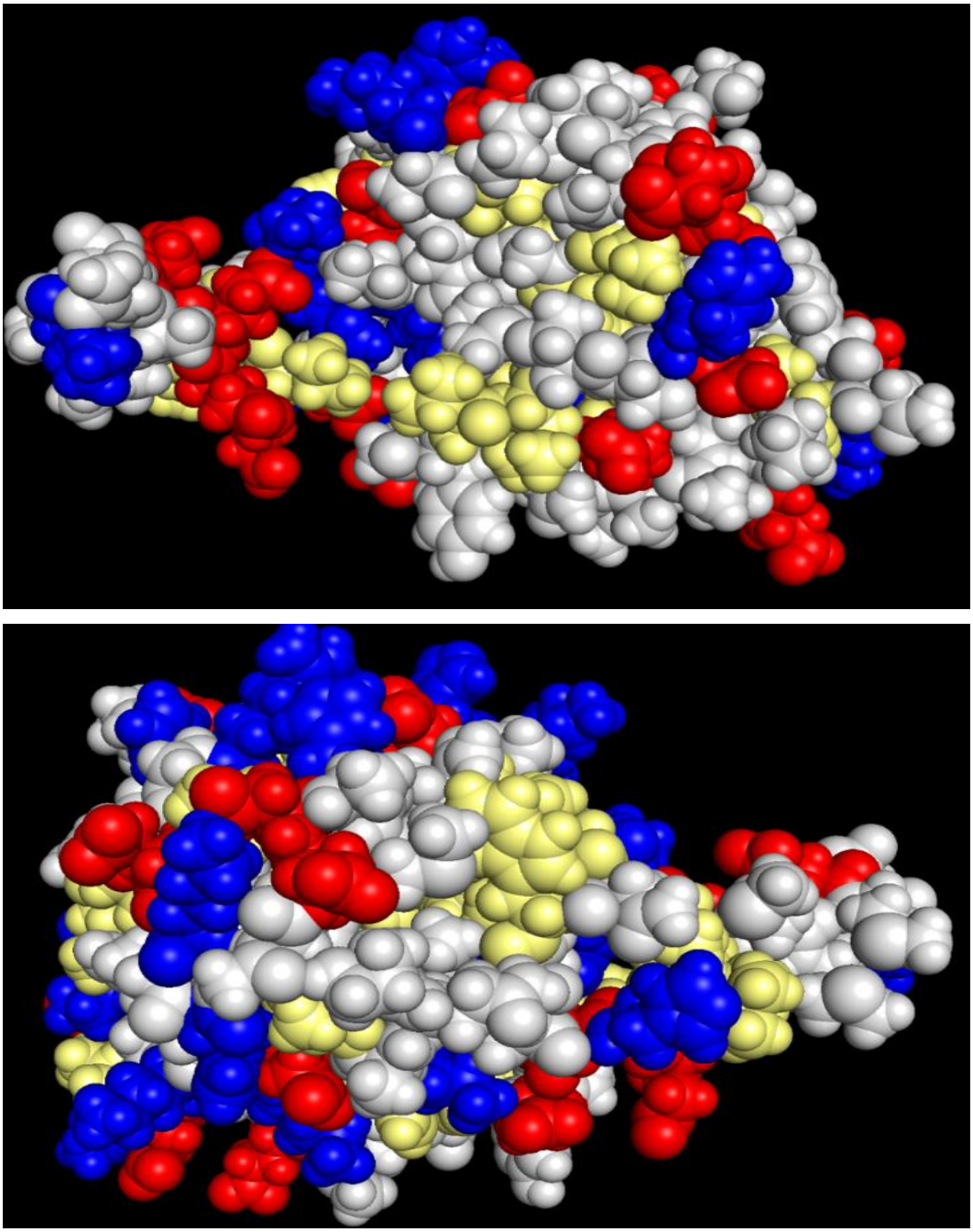


Figure 4.12: Space-filling representation of OfurPBP2. Acidic residues in red (Glu and Asp), basic residues in blue (Lys and Arg), and hydrophobic residues (Leu, Ile, Phe, Trp, Val) in yellow.

The histidine acts as a general acid-base by either donating or accepting a proton. There are 2 hydrogen bond pairs between His and Asp (Asp63-His69, Asp92-His88). The role of His-Asp pairs, linked by a hydrogen bond in many enzyme systems, appears to act as a charge relay system.[240]

The aromatic amino residues Phe12, Phe36, Trp37, Phe76, Phe118, and Trp127 line the hydrophobic cavity and are highly conserved among lepidopteran PBPs, while is replaced by aliphatic residue in GOBPs.[95] The aromatic ring of the Phe76 and Phe36 are orthogonal to Phe12 and Phe118, respectively. Likewise, Phe12 and Phe118 are sandwiched, forming the strongly conserved π−π interactions observed in the other lepidopteran PBPs. These residues form the wall of the pocket and are responsible for nonspecific binding. The Trp37 and Lys6 are solvent accessible and strictly conserved in Lepidopteran PBP. Both of these residues form the opening of one end of the hydrophobic cavity. All the seven helices on OfurPBP2 have C-capping hydrogen bond interactions. In addition, the α6 contains both C-capping and N-capping hydrogen bonds interaction. These capping interactions play a central role in the stabilization of the helices.

There is an extended turn, a β hairpin loop (Figure 4.13), consisting of residues Leu61-His69, that lies between α3 and α4, closely resembling a flap. The residues of the flap are projecting over the opening of the hydrophobic pocket, allowing ligand to access the pocket opening. This flap remains flexible and allows for hinge-like mobility, which may be responsible for the hydrophobic specificity. The mobile flaps contain three distinct regions; residues Leu61-Pro64 is the N-strand of the first half of the β-hairpin loop, residue Glu65 lies at the tip of the center turn region, and residue Gly66-His69, is the C-strand or the second half of the loop. The N-strand of the β-hairpin loop is antiparallel to the C-strand. Both the strands are held through four hydrogen bonds involving Gly66N-Asp63O, Asp63O-Gln65N, Asp63N-Thr67O, and His69N-Leu61O (Figure 4.13).

Furthermore, the side chain of Thr67 and the main chain NH of Asp63 interacts and enforces this conformation. These loop residues are involved in the opening of the flaps. The flap residues and with Arg70, His88, and His95 might play a major role in controlling the opening and closing of the binding pocket. These residues are conserved throughout the PBPs and GOBPs except Arg70, which is substituted by His70 in many well-studied PBPs of several lepidopteran species. Moreover, these flap residues in OfurPBP2 may have a potential role in ligand binding and/or release from the pocket. This loop/flap is relatively flexible and acts as a lid to the binding pocket. The side-chain residues of the loop/flap may control the opening of the pocket that is wide enough for a ligand to either enter or exit the pocket. Presumably, if this loop/flap were not in place, the resulting opening of the pocket would be inadequate for a pheromone to enter or exit the pocket similar to what has been reported for BmorPBP.[72]

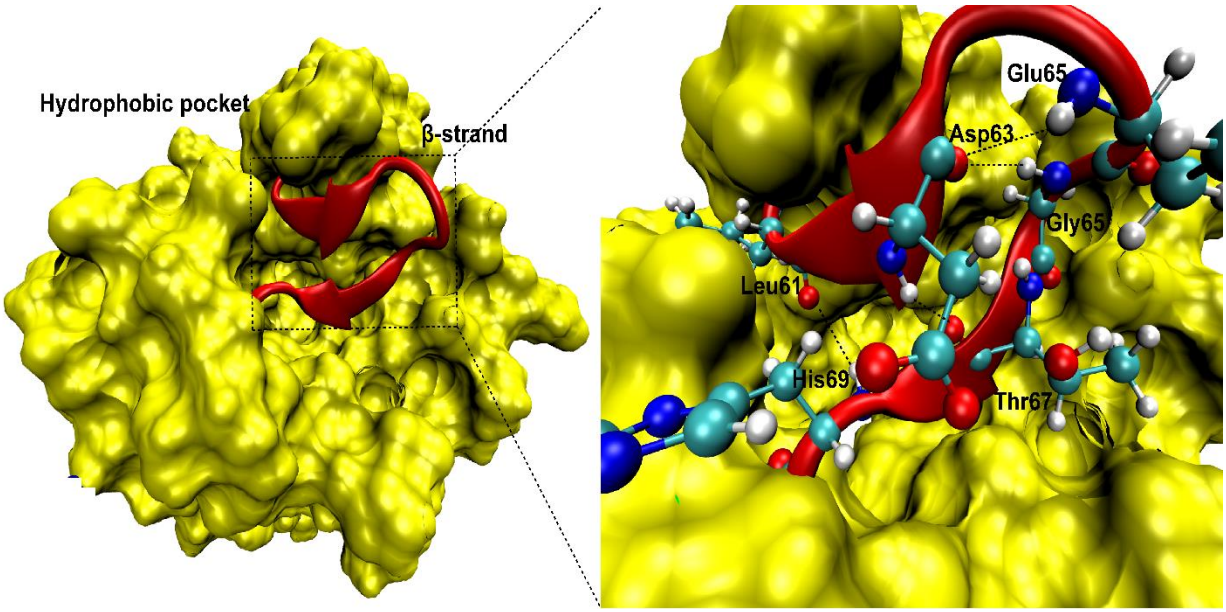


Figure 4.13: The β hairpin loop from residues from Leu61-His69, resembles a flap covering the hydrophobic pocket opening.

In many lepidopteran PBPs, including ApolPBP1, BmorPBP, Atra1PBP1, and LdisPBP2, residues His69, His70, and His95 are strictly conserved and act as a pH-dependent molecular switch, known as histidine gate, that in part regulate the characteristic conformational change associated with ligand-binding and release.

In OfurPBP2, His70 at the beginning of the α4 helix is replaced by Arg70. However, there is a His88 residue in the α5 helix of OfurPBP2 that is not present in the other well-studied lepidopteran PBPs containing the histidine gate mentioned above.[61,65,82,198] In OfurPBP2, Arg70 (in α4 helix) and His88, His95 (in α5 helix) are all located close to the base of the flap (Figure 4.14). The distance between Arg70 and His88 is less than 6 Å at pH 6.5, the positively charged Arg70 forms a cation-π (His-Arg+) interaction[241-243] with His88. However, at acidic pH, the protonated His88 would disrupt the cation-π interactions with Arg70. Furthermore, the repulsive forces between the positively charged residues in this region (α4 and α5 helices) may lead to a partial unfolding of the protein. The partial unfolding to a molten globule-like state as observed at low pH of 5.0 or below (Figure 2.14 in Chapter 2) may be the key to ligand release. Based on the OfurPBP2 structure at pH 6.5, we predict that Arg70 and His88 together may play the role of the histidine gate (His70-His95) reported for ApolPBP1[82], BmorPBP[61,198], AtraPBP1[85], and LdisPBP2.[89] Hence, at the physiological pH, we hypothesize that His69, His88, His95, and Arg70 together with β turned (residues 61-69) will regulate the width of the pocket opening and act as a lid, and thus may have a potential role in ligand binding and/or release.

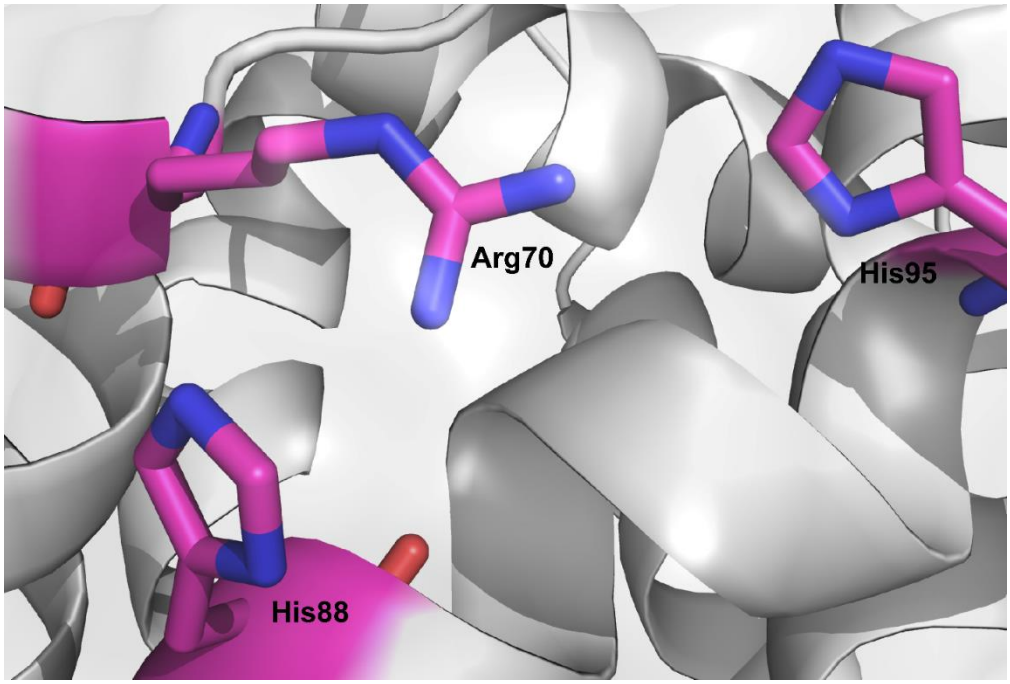


Figure 4.14: Cation-π interaction (Arg+ His), with His88, His 95, and Arg70.

The Binding Pocket of OfurPBP2

The structure of OfurPBP2 contains a large 304±32 $Å^3$ horse-shoe-shaped hydrophobic cavity that closely corresponds with the volume of 12-E/12-Z-tetradecyl acetate (290 $Å^3$) (Figure 4.15). The mouth of the hydrophobic cavity is elliptical with a diameter of 12 Å× 6 Å. The cavity opening is bordered mostly by hydrophobic residues consisting of Leu62, Gly66, Thr67, and Leu68 from L3, Met110, Ala111, Ile114 from α6, and Leu53, Ser56 from α3, and Val94 from α5. These hydrophobic side chains act as the lips of the cavity and can provide a basis for potential binding sites for hydrophobic ligands. The cavity is approximately 24.5 Å long. The other end of the cavity lies in the interface of loop 3, α6, and the C-terminus. The cavity is lined by the side chain of 29 residues: Met5, Lys6, Met8, Thr9, Phe12, and Ile13 of α1, Glu 32, Val33, Phe36, and Trp37 of α2, Ala 48, Ile52, Ser56, Leu61, and Leu62 of α3, Thr67 and Leu68 of L3, Asn72, Thr73, Phe76, and

Ala77 of α4, Leu90, and Val94 of α5, and Ala111, Leu112, Ile114, Ser115, Phe118, and Lys119 of L6. The pheromone of OfurPBP2 consists of a blend of E-12/Z-12-tetradecyl acetate (ratio of approximately 2:3)[244], accommodated inside the cavity in a bent conformation. Both pheromone molecules are stabilized inside the binding cavity by numerous hydrophobic interactions with the residues lining the pocket of OfurPBP2.

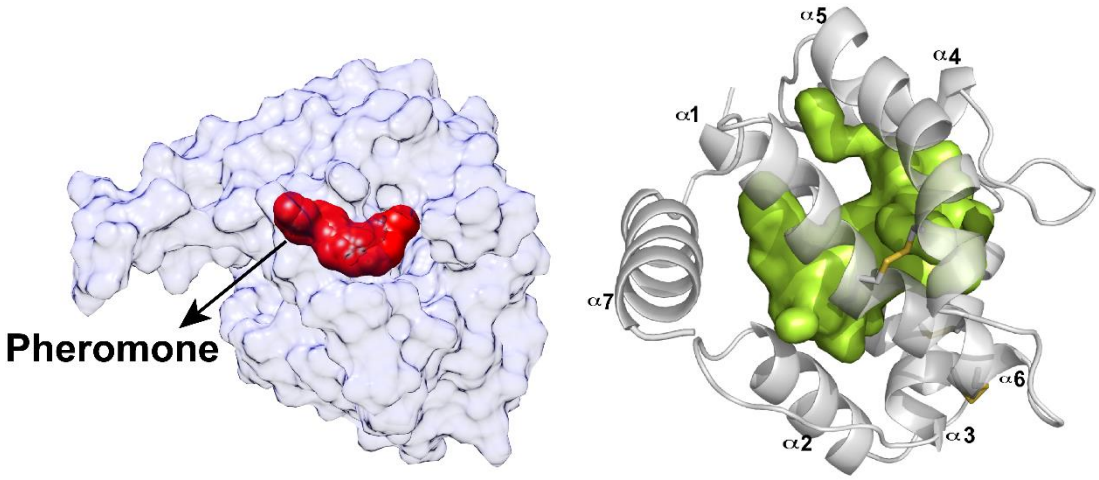


Figure 4.15: The binding cavity of OfurPBP2; the surface of the hydrophobic cavity is shown in red and green.

Docking is an important techinque for understanding the protein and ligand interactions[245]. In OfurPBP2, docking studies show both pheromones are accommodated inside a U-shaped pocket. The E-ligand interacts with Met5, Lys6, Met8, Thr9, Phe12, Phe36, Trp37, Phe76, Ser115, Phe118, Lys119, Ile122, His123, and Ala128 (Figure 4.16 A), while the Z-ligand interacts with Lys6, Met8, Thr9, Phe12, Phe36, Trp37, Phe76, Ser115, Phe118, Ile122, His123, and Ala128 (Figure 4.16B). Of the residues that interact with the pheromone, Phe12, Phe36, and Phe118 are strictly conserved throughout lepidopteran PBPs, including ApolPBP1[80] and BmorPBP.[61] The side chain of Thr9 forms a hydrogen bond with the acetate group of the ligand with an O-O the distance of 2.94 Å (for

E isomer) and 3.0 Å (for *Z* isomer) (Figure 4.16 and 4.17). The Thr9 in OfurPBP2 is replaced by the Ser in ApolPBP1, AtraPBP1, and BmorPBP.

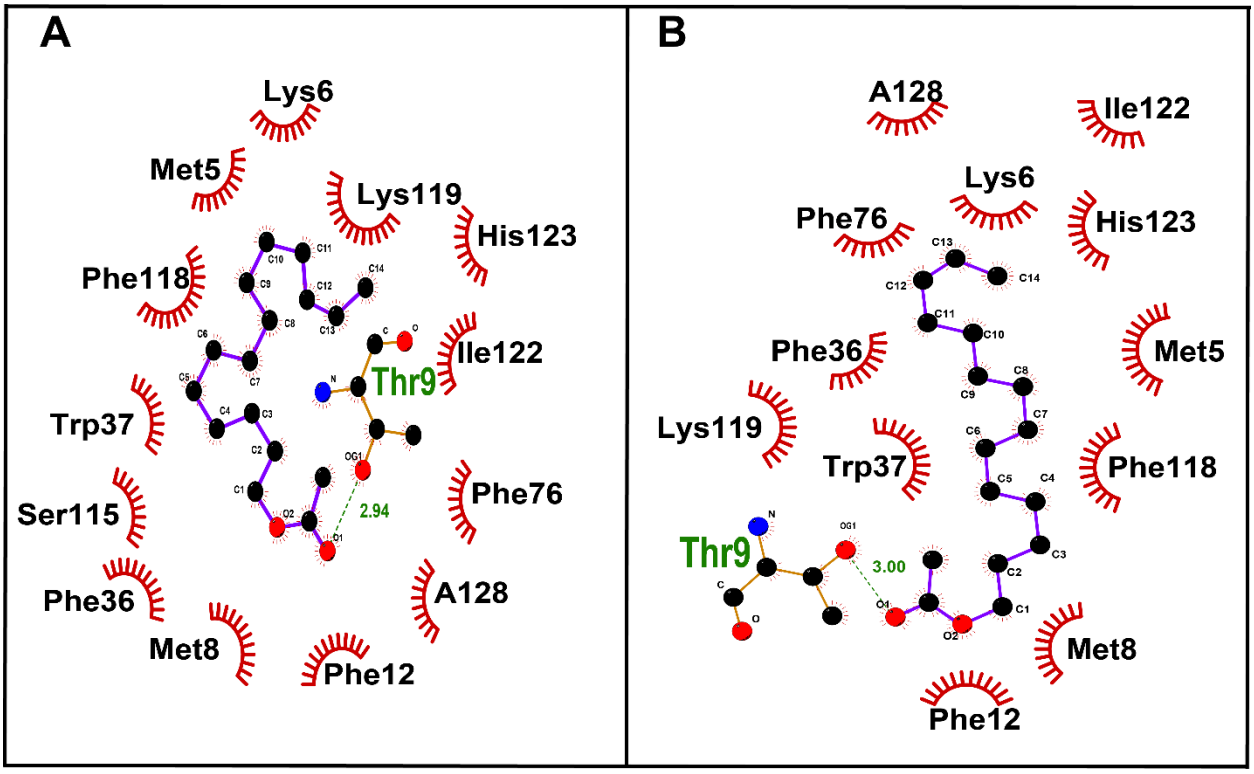


Figure 4.16: Ligplot showing the docking conformation of the pheromone with OfurPBP2 are showing the interaction with crucial residues in the hydrophobic pocket using AutoDock (A) *E*-12-tetradecenyl acetate pheromone (B) *Z*-12-tetradeceyl acetate pheromone. The oxygen atoms on the pheromones are represented as solid red circles. The hydrophobic interactions are shown as arcs with spokes radiating towards the ligand atoms they contact.

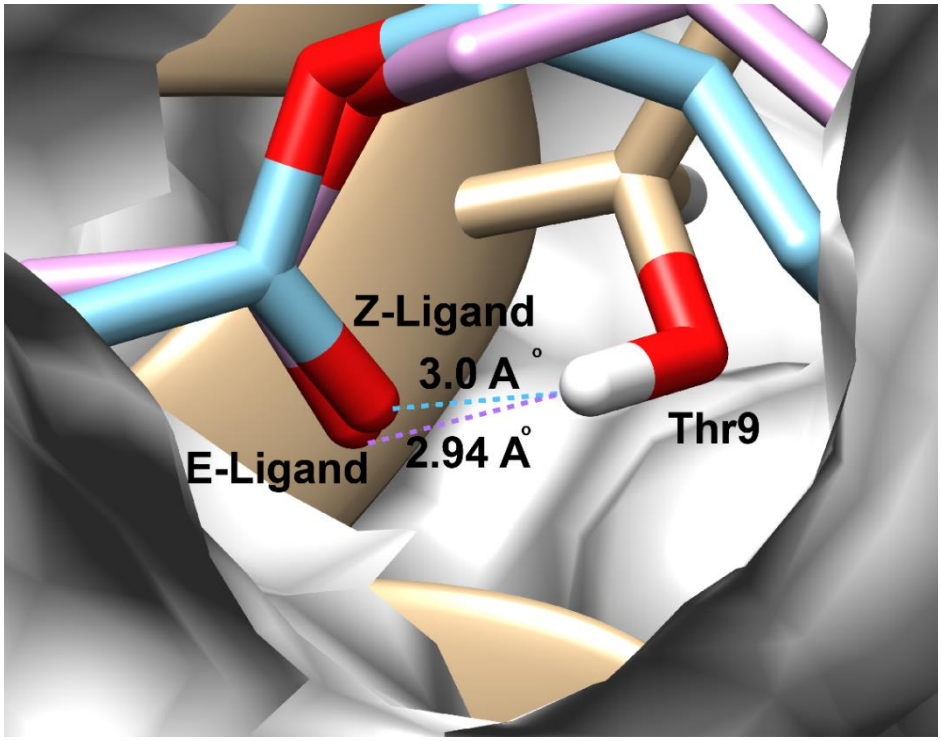


Figure 4.17: A) OfurPBP2 complexed with *E*-12 and *Z*-12 tetradecenyl acetate pheromone in the hydrophobic pocket. The oxygen atom in the pheromone (*E*-12/ *Z*-12 tetradeceyl acetate) molecule forming a hydrogen bond interaction with the side chain of Thr9.

4.3.6 The C-terminal Alpha Helix

The OfurPBP2 structure at physiological pH of 6.5, the C-terminus is composed of residues Pro129-Gln144. However, residues His131 to Lys143 form a well-structured amphipathic helix (α7). The flexible flap residues Leu61-His69 and Arg70-His88 clusters are situated at one end of the cavity, and the C-terminal amphipathic helix is located at the other end as a gate. The C-terminal helix (α7) is18.7 Å long. The helix α7 is composed of five charged residues that include four negatively charged residues (Glu132, Glu136, Glu137, and Glu141) and one positively charged residue at pH 6.5 (Lys143). However, an additional negatively charged (Asp130) is located at the base of helix α7. The charged residues lie on the same face of the helix, giving it an amphipathic character. The charged residues of the C-terminus tail may also play a role in protein switch to a molten globule form at acidic pH. At pH 4.5, His131 will also be protonated, giving rise to a total of seven charged residues in the C-terminus. The pH-titration and ligand-binding studies suggest that OfurPBP2 undergoes both pH and ligand-induced conformational change.

4.3.7 The Structure Comparison and Significance

The NMR derived structure of OfurPBP2 at pH 6.5 shows an overall fold similar to that of BmorPBP[61] ApolPBP1[80], AtraPBP1[85], and GmOBP2[95] except for the C-terminus. The rmsd values listed in Table 4.3 show the comparison with previously studied pheromone-binding proteins in the literature. The superimpose structures are shown in Figure 4.18. Interestingly, OfurPBP2 has a high degree of structural similarity with a general odorant-binding protein 2 from Bombyx mori (BmorGOBP2). Although there is only a 29% sequence identity between OfurPBP2 and BmorGOBP2, they have a long C-terminal amphipathic α-helix unlike the other well-studied PBPs mentioned above.[61,63,80,83,88] However, a significant difference is observed in the orientation of the C-terminal helix. In BmorGOBP2, the C-terminus is oriented roughly orthogonal to the N-terminal

helix, whereas, in OfurPBP2, the orientation of the C-terminus is revered. In addition, residues from Glu27-Trp37 form a single α-helix in OfurPBP2, but the same region is composed of a loop (Asp27-Asp30) and a helix (Glu31-Trp37) in BmorGOBP2, and the residues from Phe33 to His35 bulge out of the helical axis.

Table 4.3. Structural comparison with the neighborhood of OfurPBP2

Protein	(R.m.s.d. (Å)	% Identity	Protein length	Binding cavity vol. ($Å^3$)	PDB ID	Interacting residue	Reference
Silkworm moth, Bombyx mori pheromone-binding protein (BmorPBP)	1.95	57.75	137	171	1DQE	Ser56	Ref [61]
Malaria mosquito, Anopheles gambiae odorant-binding protein 1 monomer (AgamOBP1)	3.88	14.78	125	27	2ERB		Ref [97]
Cockroach, Leucophaea maderae pheromone-binding protein (LmaPBP)	3.9	18.97	119	85	1ORG	Tyr5 Tyr75, Lys89 Phe110	Ref [90]
Honey bee, Apis mellifera antennal-specific protein 1 (AmelASP1)	5.5	15.65	119	128	2H8V		Ref [92]
Honey bee, Apis mellifera antennal-specific protein 2 (AmelASP2)	4.57	13.56	123	157	1TUJ		Ref [91]
Silkworm moth, Antherea polyphenum pheromone-binding protein (ApolPBP1)	4.39	52.11	142	282	1QWV	Asn53	Ref [80]
Amyelois transitella pheromone-binding protein (AtraPBP1)	1.86	53.52	142	156	4INX	Arg107, Met61, Gly66	Ref [85]
Silkworm moth, Bombyx mori	2.68	29.08	142	n/a	2WC5	Arg110	Ref [95]

ordrant -binding protein (BmorGOBP2)							
Silkworm moth, Bombyx mori pheromone-binding protein (BmorPBP)	2.366	57.75	137	272	1LS8		Ref [63]
Fruit fly, Drosophila melanogaster odorant-binding protein(LUSH)			124	114	1OOH	Thr57, Ser52 and Thr48	Ref [102]

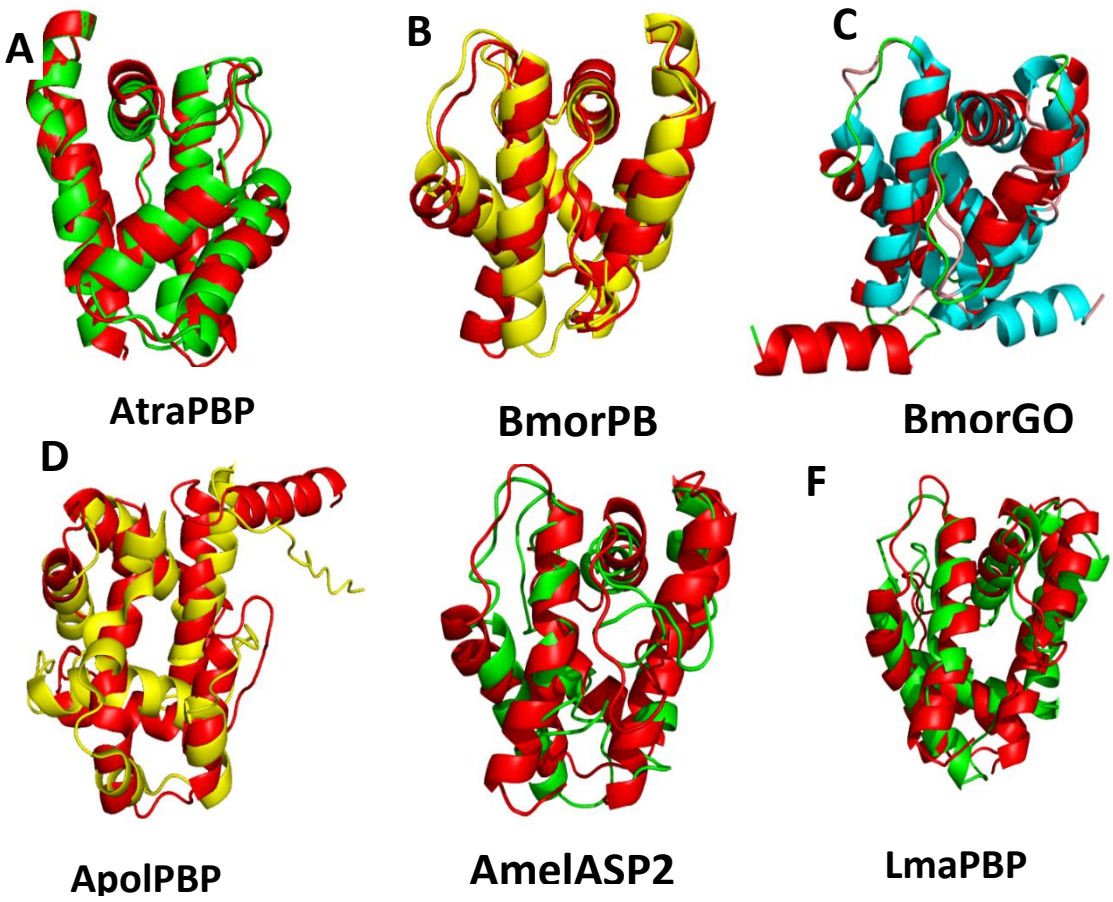


Figure 4.18: Superimposition of the NMR structure of OfurPBP2 at pH 6.5 (red) with A) AtraPBP1, B) BmorPBP, C) BmorGOBP2, D) ApolPBP1, E) Honey bee ASP2, and F) LmaPBP

The well-conserved residue Tyr41, within the PBPs, is replaced by Phe in GOBP2. In GOBP2, Phe41 is more deeply buried inside the protein core than the Tyr41 in BmorPBP1 and OfurPBP2. Furthermore, helix α2 (Glu27-Trp37) of OfurPBP2 is roughly parallel to helix α1a, while the same region in GOBP2 has a shorter helix (Asp31-Trp37) that is orientated at a 50˚angle from the N-terminal helix.

Regardless of the presence of pheromone, the C-terminus is always outside of the binding pocket in GOBP2. The ligand-binding and release mechanism are not always dependent on the C-terminus, as in BmorGOBP2, where the C-terminus is always an α-helix that does not occupy the binding cavity and also does not participate in ligand binding.[95]

The pheromone molecules in lepidopteran generally consist of an oxygen-containing head group and a long hydrocarbon chain, which differ substantially among species both in terms of functional group and carbon chain length. The head group of the pheromone is alcohol for *Bombyx mori*[61], an aldehyde for *Amyelois transitella*[85], and acetate for both *Anthereae polyphemus*[80,82] and *Ostrinia furnacalis*[114]; however, *A. polyphemus* has a 16-carbon chain pheromone as opposed to the 14-chain for *O. furnacalis*. Although most hydrophobic residues that stabilize the pheromone in the pocket are conserved, the key residue that hooks the pheromone by forming hydrogen bonds varies among the species; these are Ser56 in BmorPBP[61], Asn53 in ApolPBP1[80], Arg110 in BmorGOBP2[95], and Arg107, Met61, and Glu98 in AtraPBP1[85] (Table 4.3). Similarly, based on our docking studies, Thr9 forms a hydrogen bond with the acetate head group of both E and Z pheromones. These key residues are not strictly conserved in well-studied PBPs. The variation of hydrogen bond-forming residues potentially plays a crucial role in the specificity and selectivity of the corresponding pheromones in their PBPs.

The side-chains residues at 8 and 9 positions may serve a crucial role in the determination of pheromones chain length. The primary sequence of ApolPBP1 contains Leu8 and Ser9. However, in OfurPBP2 and ApolPBP3 the Met8 and Thr9 are found. The pheromone of ApolPBP1 contains a 16-carbon chain, while the pheromone of OfurPBP2 and ApolPBP3 contains a 14-carbon chain. PBPs with longer chain length (i.e. a 16-carbon) pheromone, contain Leu8 and Ser9 while Met8 and Thr9 are found in 14-carbon pheromones. This observation is consistent with a previous

suggestion that Met8 and Thr9 replace the less bulky Leu8 and Ser9 to reduce the volume of the pocket, making it favorable for shorter chain length pheromones. The OfurPBP2 data provide further evidence to support the earlier proposed pheromone selection mechanism.[80]

Two distinctively different models have been proposed for the mechanisms that transfer the molecular signal to ORs. The pheromone binding protein from BmorPBP[58-64,67,69,71,168,197,198], ApolPBP1[65,66,76,80,82,196,246], and AtraPBP1[83-85] bind and release the pheromone through a pH-dependent conformational change, where the C-terminal region is unstructured at pH 6.5. However, at acidic pH, the extended C-terminus switches to an α-helix and is inserted inside the binding pocket to compete with the pheromone and releases it out of the pocket.[59,61-63,65,67,76,196,197] These proteins have both pH-induced and ligand-induced conformational changes. Contrary to the pH-dependent conformational change, LUSH (odorant-binding protein of the fruit fly *Drosophila melanogaster*) undergoes a conformational change due to the consequence of ligand binding, which occurs at pH 7.0[102] but not with pH.[247] In the absence of a ligand, LUSH exists in a partially molten globule/unstructured state.[101] The binding of a ligand causes a conformational switch and shifts to a compact, folded, and active conformation state, which increases overall protein stability that can trigger the signal activation. The general odorant-binding protein structure of *Bombyx mori* (BmorGOBP2) shows a significant deviation in the C-terminal conformation from BmorPBP1, AtraPBP1, and ApolPBP1, showing that ligand binding and releasing mechanisms are independent of the C-terminus. Although BmorGOBP2 has a comparably-sized C-terminus, it forms an amphipathic α-helix that packs across the top of the N-terminal helix, and neither covers the binding site nor participates in ligand binding but rather remains outside of the pocket.[95] There are no conformational switches on BmorGOBP2 structures when bound with the sex pheromone components and their analogs.

In well-characterized lepidopteran PBPs[65,82,61,198,85,89], the histidine gate, which consists of His70 and His95, plays a critical role in pheromone release. However, in OfurPBP2, His70 is replaced by arginine. Furthermore, the histidine residue present in OfurPBP2 at position 88 is absent in other structurally characterized PBPs mentioned above. There are some lepidopteran OBPs such as *Helicoverpa armigera* OBP15[248] and *Papilio xuthus* GOBP1[249], where His70 substituted with an arginine. Likewise, in *Maruca vitrata* PBP2[250] and *Conogethes punctiferalis* PBP1[251], where a lysine replaces His70. Furthermore, Arg70 and His88 are well conserved within the *Ostinia* PBPs, including OfurPBP3, *O. nubilalis* PBP2[136], and PBP3[136] (OnubPBP2 and OnubPBP3), and *O. latipennis* PBP-A.[252] In the structure of OfurPBP2, the His88 is at an approximate distance of 4.3 Å from Arg70, indicating a cation-π interaction.[242,243,253] We anticipate this structure is in the open or PBP^B conformation where the Arg70-His88 gate is closed. The His88 is closer to Arg70 in space (4.3 Å) as opposed to His95, which is approximately 9.3 Å away. However, based on the approximate distance between His70 and His95 is 8.1 Å in ApolPBP1[80], and 5.1 Å in AtraPBP1[85], we suggest that His95 may also participate in the cation-π interaction with Arg70 but it would be weaker than His88. Additionally, His95 is conserved among the *Ostrinia*[136,252], suggesting that it has a role in the conformational change of the protein. Thus, the protonation of His88 and His95 at low pH will result in the opening of the Arg-His gate due to electrostatic repulsion. As Arg70 is part of helix α4, we hypothesize the gate opening actuates movement in loop L3 and plays a key role in ligand release from the binding pocket of OfurPBP2.

OfurPBP2 shares 97 % sequence identity with OnPBP2 and has a strictly conserved C-terminal residue. However, there are 6 charged residues in the C-terminus in *Ostrinia nubilalis* pheromone binding protein 2 (OnPBP2) instead of 7 in OfurPBP2. Because of high sequence similarity, the structure of OnPBP2 will likely have a similar structure as OfurPBP2. Likewise, OfurPBP3 and OnPBP3 share 70 % and 73 % sequence identity with OfurPBP2 with 4 and 5 charged residues and

2 and 3 aromatic residues in the C-terminal region. The variation of charged residues in the C-terminal region impacts the structure and orientation of the C-terminus. The nature of the C-terminus, either helical or random-coil, is an interesting possibility among the sibling PBPs. The phylogenetic tree diagram of *Ostrinia* PBPs shows that OfurPBP2 and OnPBP2 are more closely related proteins with the most common ancestor.

The question of whether the flexible C-terminus is responsible for ligand releasing cannot be answered with only this NMR structure. Due to the dynamics of the C-terminal region, there is a subtle conformational fluctuation, which may have a role in opening and closing the path for ligand binding. Although OfurPBP2 and other Lepidopteran PBPs are highly conserved, their amino acid sequences in the C-terminal portion are quite divergent. These observations support our hypothesis that the C-terminal helical segment of OfurPBP2 is fluctuating and associated with the ligand.

4.3.10. Molecular Dynamics Simulations

The flexibility of a protein was characterized by running MD simulations of 200 ns because it was long enough to observe large flexibility in protein loops and terminal regions. A root means square fluctuation (RMSF) analysis was carried out on the protein to reveal the flexibility and local motion (Figure 4.16 A and B). The conventional Cα RMSF supports the validity of using the first principal component of the trajectory as a measurement of an atom's flexibility. The RMSF profile confirmed that OfurPBP2 adopts a compact module in the region 1–130. Four different loops; L1(Lys21-Ala29), L2 (Lys38-Thr44), L3 (Gly66-Gly71), and L4 (Ser100-Asp106), display a large amount of structural fluctuation confirming the flexibility with the RMSF larger than 0.4 Å (Figure 4.19 C and D). The C-terminal region has a high RMSF value of more than 1. This finding suggests that the loops and the terminal fragments are sufficiently flexible, whereas the helical scaffolds are more rigid. The flexibility of the flaps/lid has a physical significance, which involves an opening motion

of the pocket (Figure 4.19). The backbone RMSD shows that protein has reached stability before 200 ns.

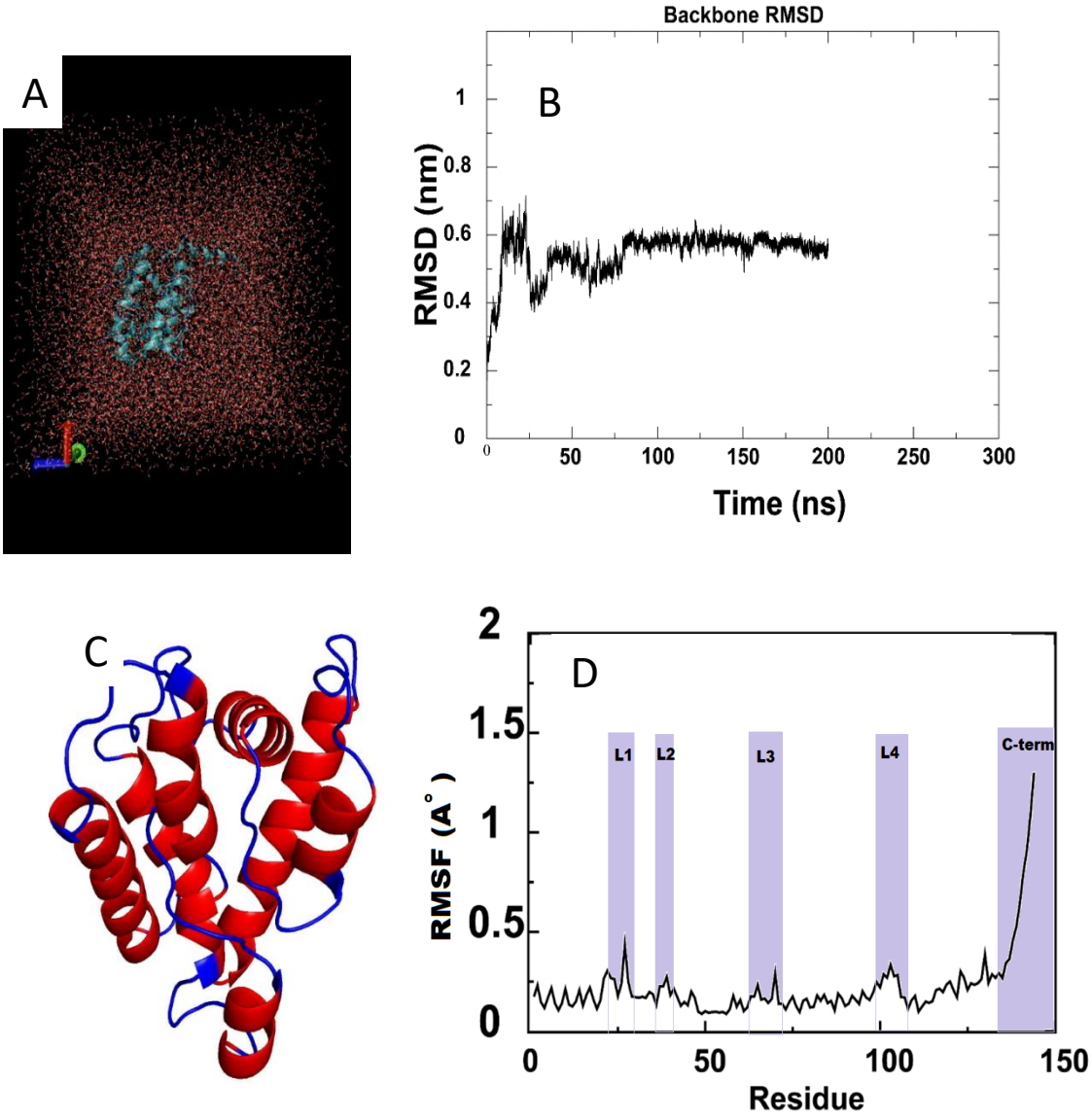


Figure 4.19: MD simulations, A) Preparation of the MD system where protein is surrounded with water in a cubic box. B). The root means square deviation of the protein showing stability at 200 ns. C) The protein structure showing higher RMSF values at the flexible loop region is shown in blue color. D) The plot Root Mean Square fluctuation (RMSF) values against protein residues.

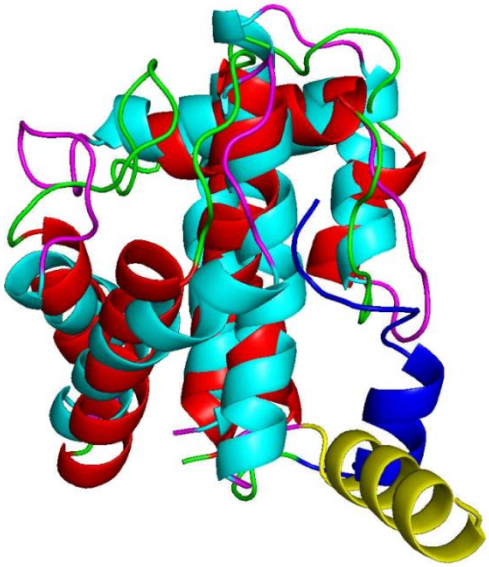


Figure 4.20: Overlay of OfurPBP2 model (red) with 150 ns of the MD trajectory (blue). The purple color and green color show the loops, and the yellow color and dark blue color show the C-terminal region on the respective models.

4.3.11. Dynamics of the Bound Complex

Molecular dynamics (MD) is a computational technique that simulates the dynamic behavior of molecular systems as a function of time, treating all the entities in the simulation box, ligand, protein, along with waters (if explicit) as flexible. To obtain the ligand poses in the pocket, the protein-ligand interaction dynamics were characterized. The 150 ns long MD simulations were conducted. The mechanism of ligand uptake and release may occur within the millisecond timescale. To understand the putative pathway for the pheromone entrance and exit, we need to execute the detailed simulation studies. It is shown already that OfurPBP2 undergoes a pH-dependent conformational transition, which is also potentially correlated with the ligand uptake and release mechanism. From a structural point of view, we suggest that the opening and closing of the pocket is regulated by the β-loop region formed by residues Leu61-His69, which resembles the

flaps and acts as a lid. One speculation is that the ligand might pass through this lid. Another potential opening of the pocket lies at an interface of N and C-terminal chains. The helical C-terminus is flexible. During the C-terminus motion, there is always the potential for the opening of the pocket that might allow the ligand to enter or exit. Regarding an interaction fingerprint analysis between the ligands and the protein, snapshots of the ligand and residues within 5 Å of the ligand mass center were extracted from the final 150 ns trajectory.

a) E- 12 -tetradecenyl acetate

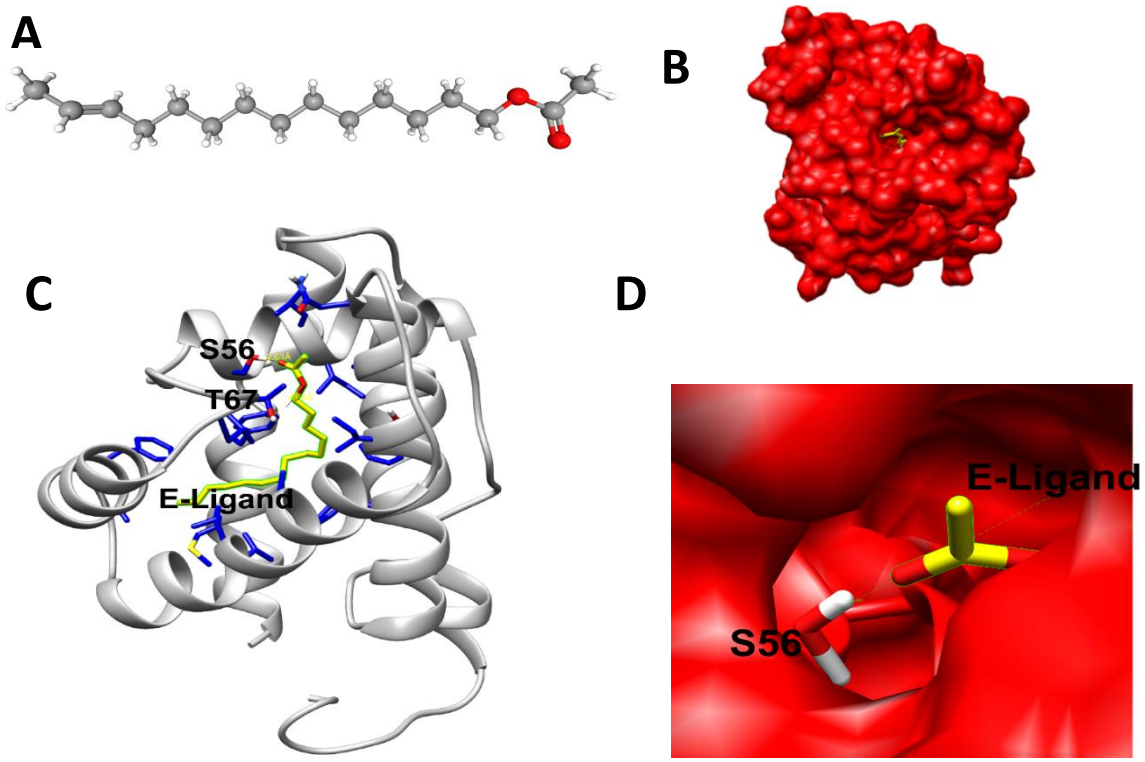


Figure 4.21: A) Structure of *E*-12-tetradecenyl acetate, B) The pheromone is tightly buried in the hydrophobic cavity yellow) C) The critical amino acid residues responsible for interaction in the pocket. D) The hydrogen-bonding interactions between Ser56 and the head group of the pheromone.

The residues taking part in the hydrophobic interaction are Met8, Phe12, Phe36, Ile52, Leu60, Leu68, Phe76, Ile114, Phe118, Ala77, Leu90, Val91, Val94, Asn107. The polar residues, Ser56, Thr67, Thr73, Ser115 also interact with the ligand. In addition to the significant hydrophobic interactions, the hydrogen-bonding interactions between the hydrogen atom of Ser56 and Thr67 with the carbonyl oxygen atom of the acetate head group were formed inside the hydrophobic pocket with an average distance of 2.63Å (Figure 4.21).

b). Z-12 -tetradecenyl acetate

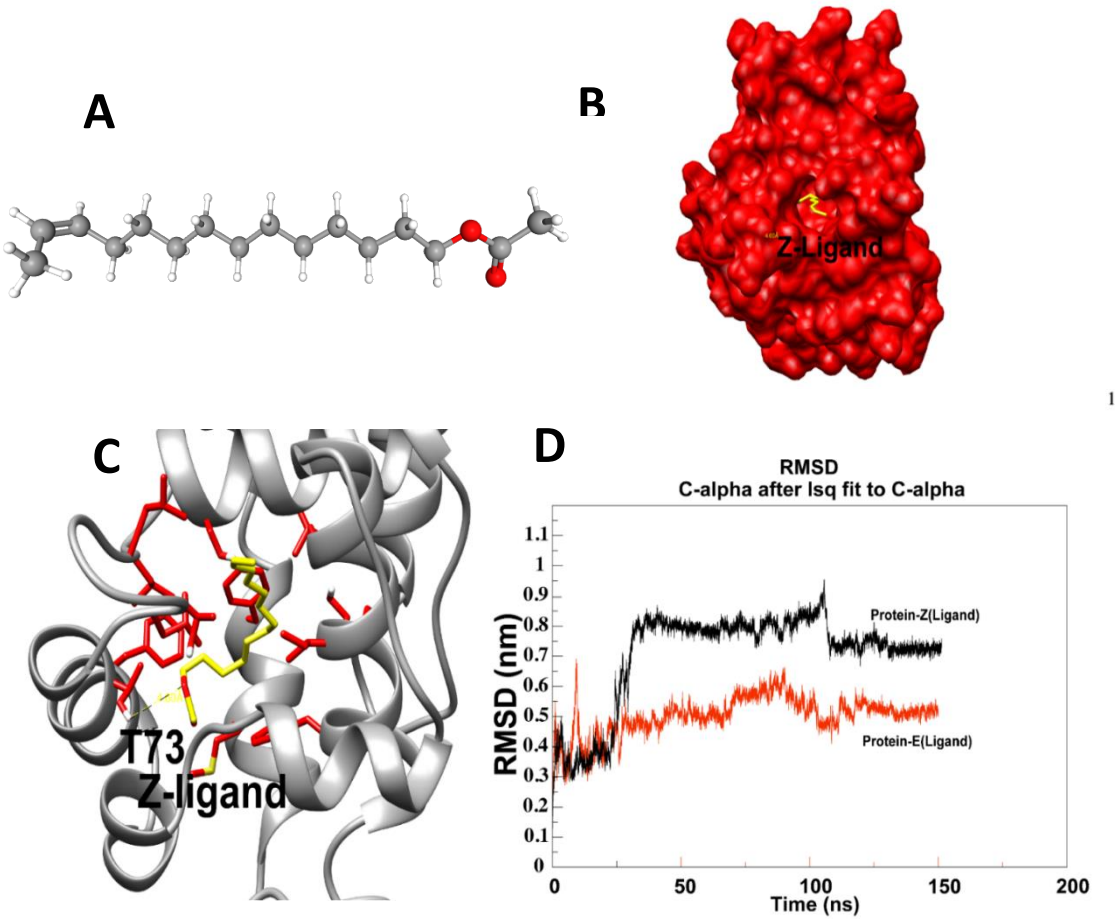


Figure 4.22: A) Structure of *Z*-12 -tetradecenyl acetate B)The pheromone is tightly buried in the hydrophobic cavity yellow) C) The critical amino acid residues responsible for interaction in the pocket. D) The root mean square deviation of the protein showing stability at 150 ns.

Similarly, *Z*-12 -tetradecenyl acetate is also tightly buried in the hydrophobic cavity. The critical amino acids are Phe12, Phe36, Ile52, Leu60, Leu68, Phe76, Ile114, and Phe118 (hydrophobic AA). Along with these hydrophobic residues, the ligand-binding pocket contains Ser56, Thr67, Thr73, Ser115 (polar AA). The residue Thr73 forms H-bonds with the Z isomer.

Both the pheromones (E and Z) were tightly buried in the hydrophobic cavity. The slight conformational fluctuations were observed, which may be due to the orientation of pheromone molecules. The residues: Phe12, Phe36, Ile52, Leu60, Leu68, Phe76, Ile114, Phe118, Ser56, Thr67, Thr73, and Ser115 of OfurPBP2 are common residues that interact with both E- and Z-isomers. The trajectories for the Z ligand have shown a rotational motion within the pocket. The orientation of the polar head group is the opposite of the E ligand. The C-terminus and the β-loop (residues 61–69) are more flexible, which would be the potential path for the pheromone to enter and exit in the pocket. The complex rmsd of the *Z*-ligand/protein complex was higher and showed larger fluctuation than that of the *E*-ligand/protein complex. Both complexes start reaching a plateau at around 40 ns. This is indicative of a stable simulation. From this point on, the rmsd remained constant (approximately 0.75 nm for *Z*-ligand and 0.45 nm for *E*-ligand). For the *Z*-ligand, the rmsd is larger (Figure 4.22D). The structure needs to fluctuate more to accommodate the Z –ligand. At around 100 ns, there is some fluctuation in the rmsd. The fluctuation is greater on the Z-ligand complex, which might be due to the rotational motion of the *Z*-ligand in the pocket. The protein/*E*-ligand complex reached dynamic equilibrium, indicating no significant fluctuation after 100 ns (Figure 4.22 D).

4.3.12. Structural Fluctuations During MD Simulation

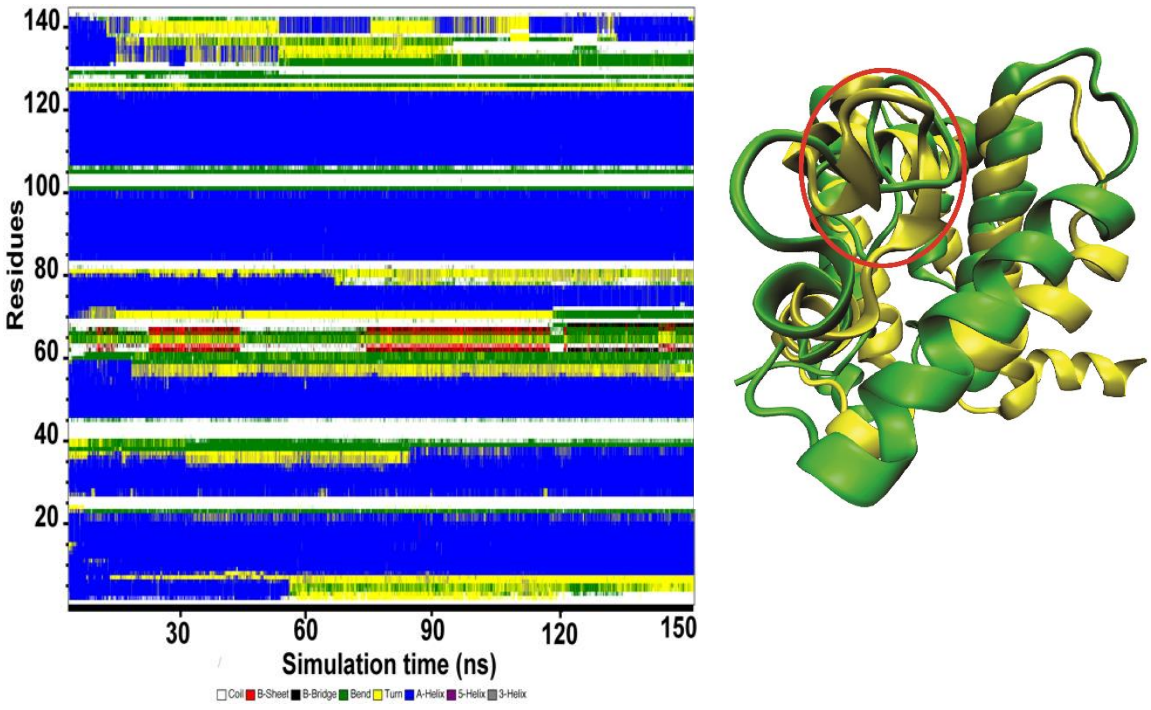


Figure 4.23. Time evolution of the secondary structural elements, based on DSSP plots in apo-OfurPBP2 during 150 ns MD simulations using the GROMOS force fields. The comparison of initial structure (yellow) with final structure (green) after 150 ns MD simulations.

The secondary structure on 150 μs MD simulations were monitored using DSSP (Figure 4.23). DSSP is a program that is based on the use of hydrogen bonds and geometric pattern recognition for secondary structure. The α-helix of the N-terminal segment was retained up to around 55 ns, whereas this part of the N-terminal segment, α1a residue from 1-8, appears to be converted into a turn bend. The residues 34−42 appear to have a higher helix-forming propensity capable of existing in the α- helical form after 80ns. Throughout the 150 ns trajectory, the residue segment between 60-70 partially interconverted into a turn, beta-sheet, and bend. At the end of the simulation, there is no longer appearance of β-sheet. In the apo-OfurPBP2, the interconversion of this segment's

secondary feature might bring flexibility to the protein, which may also involve controlling the pocket's opening. Likewise, the C-terminal helix was rapidly (within 20 ns) interconverted to turns and helix and coil. Most of the flexibility of the protein is due to the C-terminus tail. The conventional Cα RMSF supports the validity of using the first principal component of the trajectory as a measurement of an atom's flexibility. The RMSF profile confirmed that OfurPBP2 adopts a compact module in the region 1–130. Four different loops; L1(21-29), L2 (38-44) and L3 (66-71), and L4 (100-106), display a large amount of structural fluctuation, confirming the flexibility with the RMSF larger than 0.4 Å. The C-terminal region has a high RMSF value of more than 1. This finding suggests that the loops and the terminal fragments are sufficiently flexible, whereas the helical scaffolds are rigid. The flexibility of the flaps residues has a physical significance, which involves an opening motion of the pocket

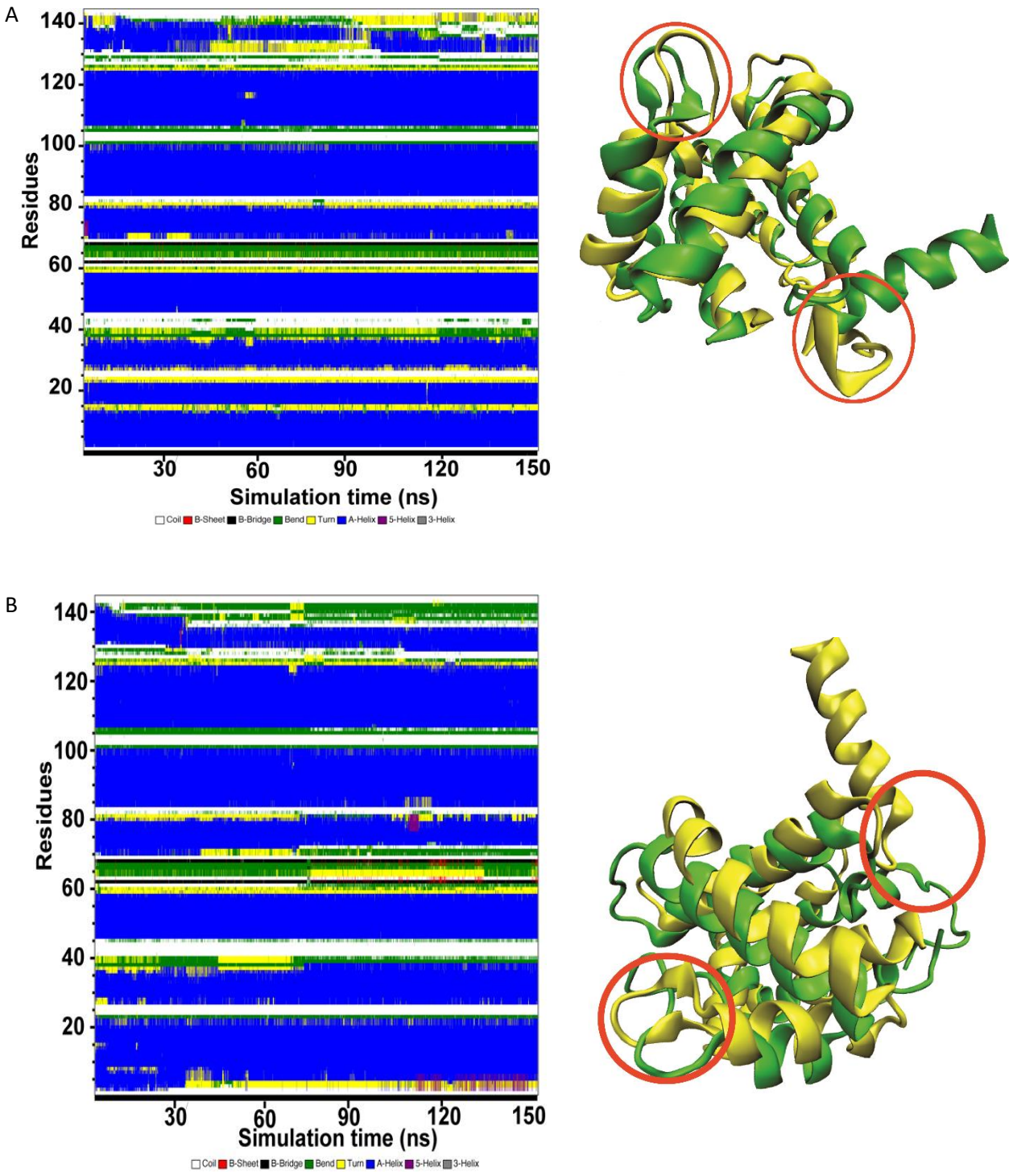


Figure 4.24. Time evolution of the secondary structural elements, based on DSSP plots in OfurPBP2 during 150 ns MD simulations using the GROMOS force fields A) Protein-E ligand B)

Protein-Z ligand. The comparison of initial structure (yellow) with final structure (green) after 150 ns MD simulations.The red circle in beta-loop and C-terminal is showing the most flexibile region.

The secondary structural element was classified using DSSP classification of secondary structure elements for each amino acid in the course of simulation time (Figure 4.24 A and B). The segment of residues 60–70 remains as beta-turn (red-yellow) conformation throughout the simulation in the E-ligand complex. While comparing with the Z-ligand complex, the beta-sheet region frequently interconverted into bend-turn-bend. It remained as bend conformation up to 70 ns and converted into turn conformation at 130ns and again retained bend conformation. Taken together, the results suggest that beta-turn has a certain flexibility, and with the slight displacements on the position, which would be the potential path for the ligand to enter in the pocket.

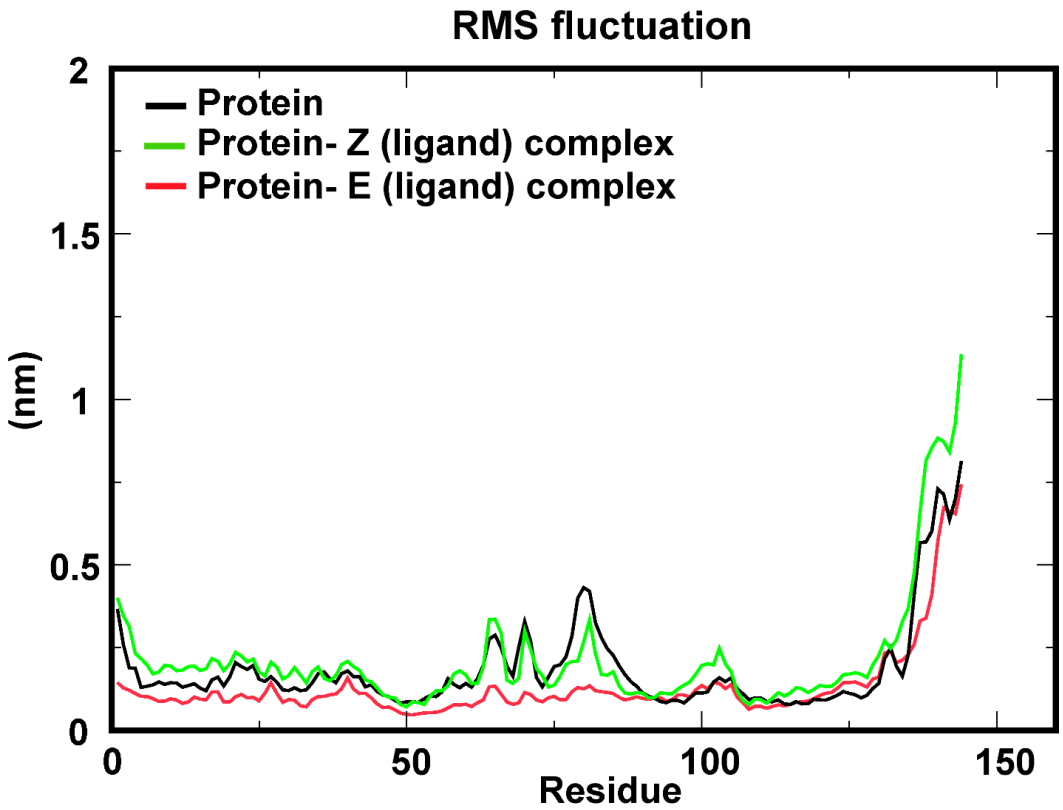


Figure 4.25: Backbone RMSF for OfurPBP2 (black), protein-Z ligand (green), and protein-E ligand (red) as a function of residue number during the MD simulations.

In order to probe how ligand interaction affects the dynamics of the backbone atoms, the root means square fluctuation (RMSF) values were calculated for backbone atoms at each time point of the trajectories for apo and bound complexes (Figure 4.25). Higher RMSF values indicate greater flexibility during the MD simulation. A plot of the RMSF shows that the N-terminal of OfurPBP2, N-terminal residues up to 10, becomes increasingly mobile, where this part of the helix is displaced up to 0.4nm. Surprisingly, the mobility of the protein has no significant impact on the Z ligand-protein complex. The degree of flexibility remains similar to that of apoprotein. Interestingly, the protein remains more stable on the E ligand complex on the N terminus, the loops, and the β-strands. But there was still a higher degree of mobility on the C-terminus, suggesting the C-terminus of the protein always remains mobile in both apo or holo state. MD simulation shows that there is

an increase in backbone dynamics for the apo-OfurPBP2 and Z ligand complex. The fluctuations are larger in the beta loop adjusted to the pocket opening. The larger fluctuations on residue Pro129 and Glu144 had larger flexibility. The larger RMSF values indicate increased random motions of these residues. Thus, RMSF and RMSD data support the interaction of the protein with E ligand stabilize the structure of OfurPBP2, as it has lesser flexibility. The protein residues' critical region from 60-70, which form beta-turn, is more flexible in apo-OfurPBP2 and Z ligand complex. The interaction of the E ligand makes the structure more compact and more rigid. This shows that the C-terminal region has similar apo and complex protein behavior, which shows that the C-terminal region may not directly involve a ligand-binding mechanism.

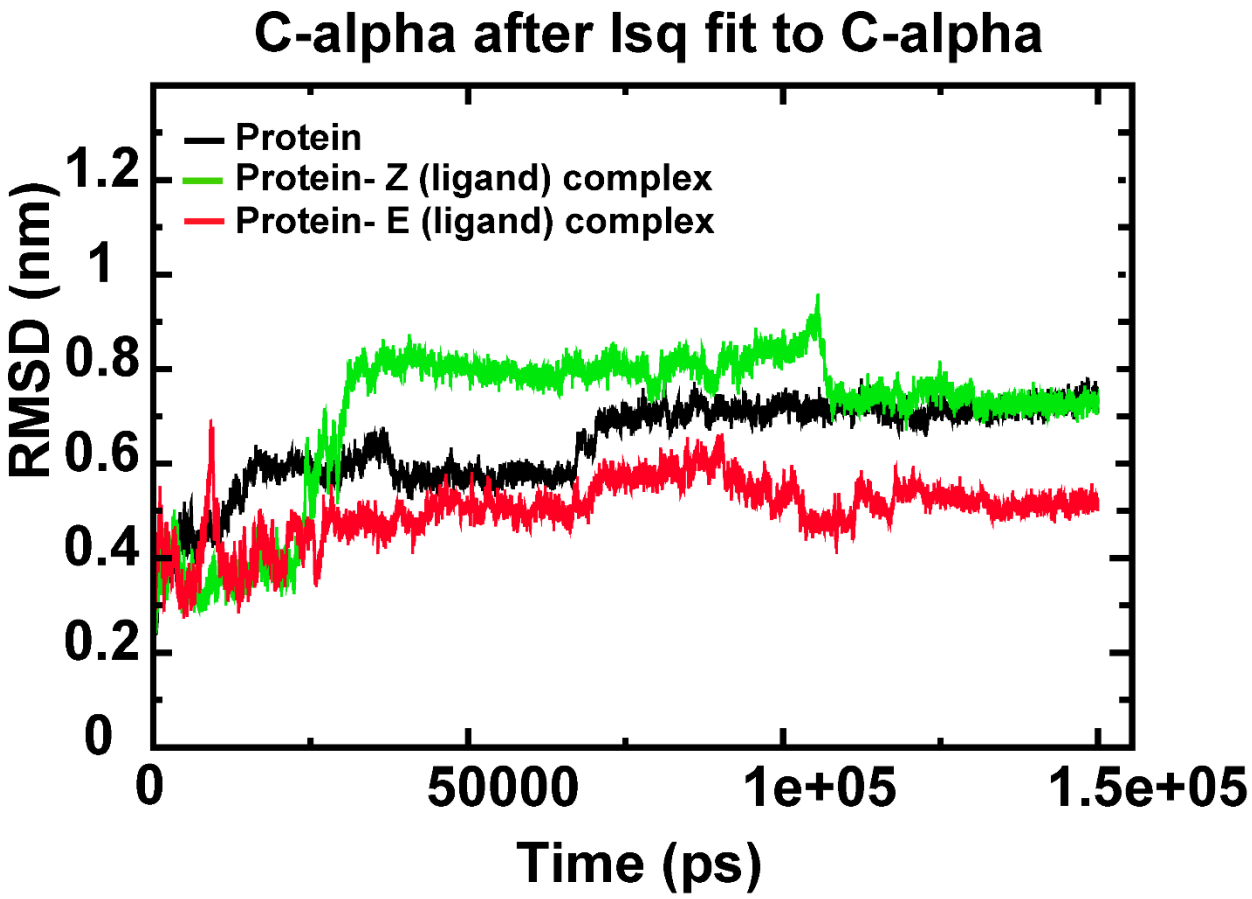


Figure 4.26: The root mean square deviation of the apo-protein and complex reaching stability at 150 ns MD simulation.

The larger RMSD values of the structures indicate that the overall topologies of the structures have changed across the trajectory. RMSD remains within 0.6 nm for the first 30 ns of the simulation, and rapid unfolding begins for the Z ligand-protein complex. RMSD reached to the plateau for a specific time. Both the complexes and apoprotein have reached a plateau at around 40 ns. This is indicative of a stable simulation. From this point on, the rmsd remained constant (approximately 0.75 nm for *Z*-ligand and 0.45 nm for *E*-ligand). For the *Z*-ligand, the rmsd is larger (Figure 4.26). The structure needs to fluctuate more to accommodate the Z–ligand. At around 100 ns, there is

some fluctuation in the rmsd. The fluctuation is greater on the Z-ligand complex, which might be due to the rotational motion of the *Z*-ligand in the pocket. Both the complexes and apoprotein have reached dynamic equilibrium, indicating no significant fluctuation after 100 ns (Figure 4.26). RMSD analysis shows that the motion E-ligand -protein complex has low RMSD values.

4.3.13. Protein-Ligand Interaction Study by {^{1}H, ^{15}N} HSQC NMR

NMR is a useful technique for monitoring the structure-activity relationship (SAR) in protein-protein or protein-ligand interactions. The two-dimensional {^{1}H, ^{15}N} HSQC spectrum is the fingerprint region of a protein. The HSQC is extremely sensitive to environmental changes, such as pH, temperature, substrate binding, etc. Any perturbation in the chemical shift or resonance from the original position indicates a change in the conformation of the protein. This change can be local, involving a few residues, or it may be an overall conformational change involving most of the residues in the protein. To investigate the effect of the pheromone on the conformational change of delipidated OfurPBP2, pheromone titration studies were conducted. The pheromone was titrated in increasing concentration (0:0 to 1:10 molar ratio). The fingerprint region of the protein in the 2D {^{1}H, ^{15}N}-HSQC spectra were monitored. Analysis of the HSQC spectra of the titration studies indicated that drastic changes in the chemical shift positions occur for the complexes. The overlays of the HSQC spectra collected in the presence of different concentrations of ligands are shown in Figure 4.27.

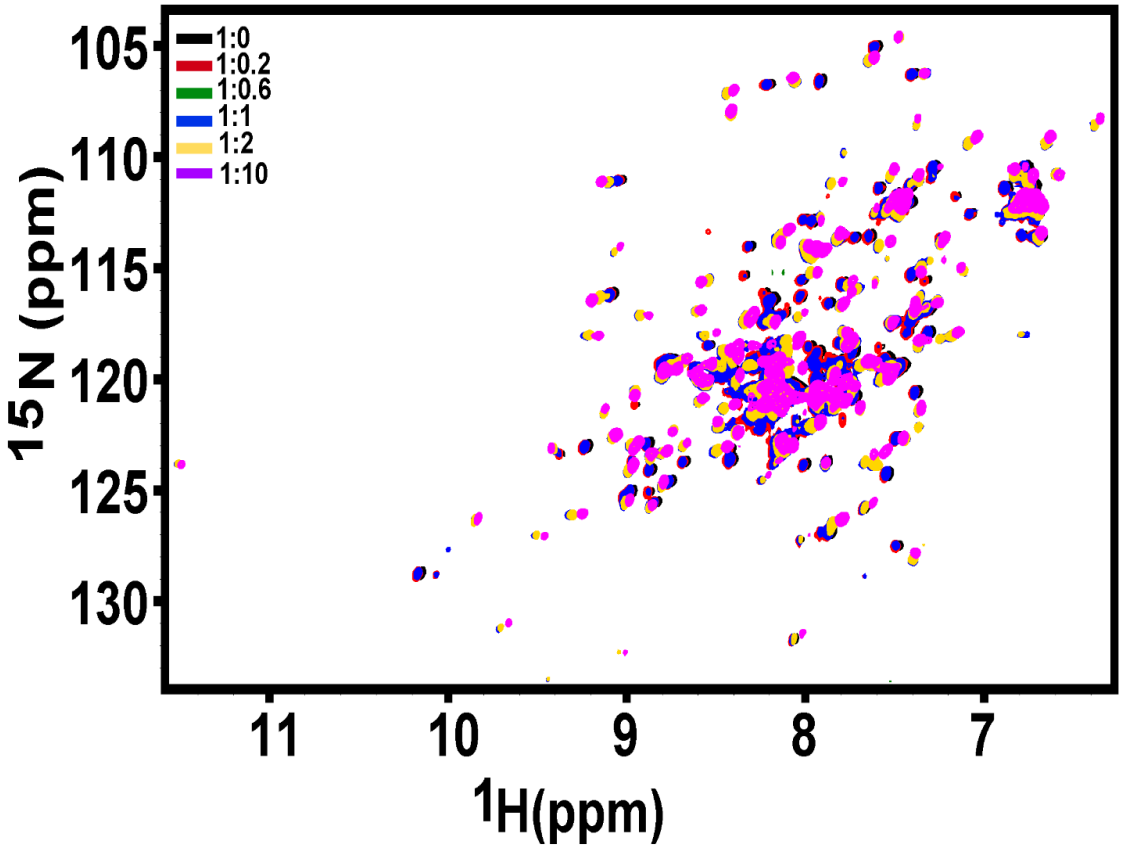


Figure 4.27: Overlay of 2D {^{1}H,^{15}N}-HSQC spectra of OfurPBP2 collected at the different concentrations of Z-pheromone.

The delipidated OfurPBP2 at pH 6.5 showed a single set of peaks corresponding to an open conformation. As soon as 0.2% of pheromone was added, two different sets of peaks appeared. One peak from a major conformation and the other was from a minor conformation corresponding from bound and free protein, respectively (Figure 4.28).

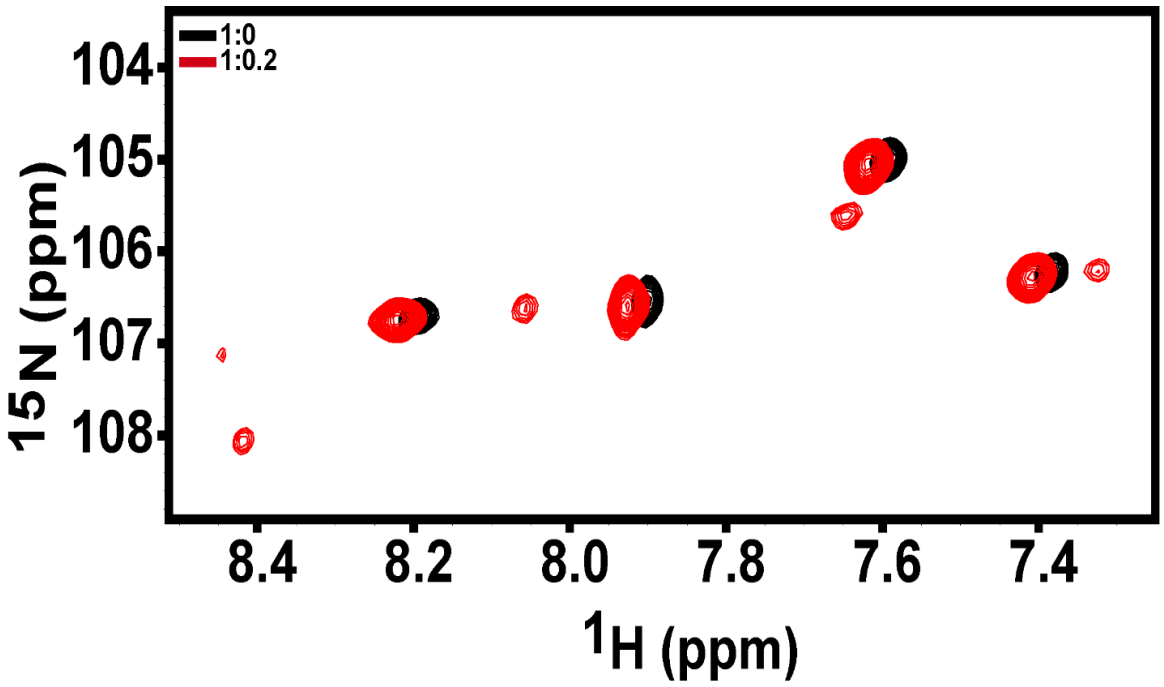


Figure 4.28: Expanded region of the two-dimensional {^{1}H, ^{15}N} HSQC spectra of OfurPBP2 titration with Z-pheromone. Protein: ligand ratios are indicated in the figure.

As the ligand concentration increased, the intensity of the resonances belonging to the free conformation of the protein gradually decreased while those belonging to the bound conformation began to appear and slowly increased in intensity. In the ratio of 1:1, the intensity of the free protein was roughly equal to the intensity of the bound protein (Figure 4.29).

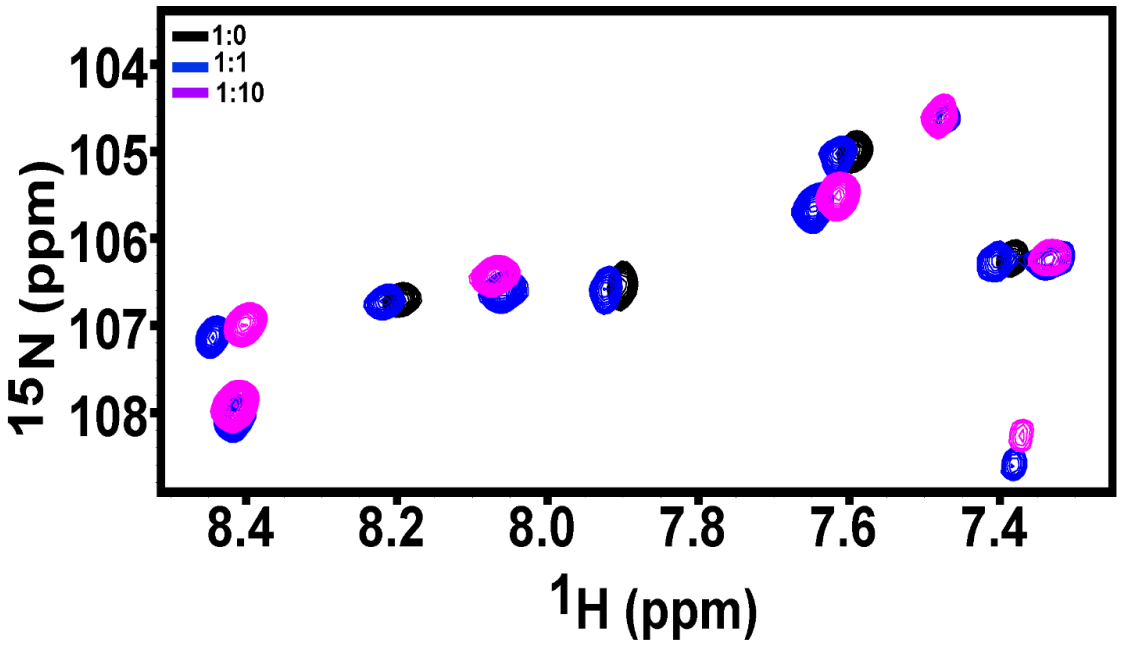


Figure 4.29: Expanded region of the two-dimensional {^{1}H, ^{15}N} HSQC spectra of OfurPBP2 titration with Z-pheromone. Protein: ligand ratios are indicated in the figure, at a 1:1 ratio (blue peaks) showing the mixture of conformations.

A mixture of conformation exists, which is approximately 50% free and 50% bound protein. At 1:1 ratio. At a 1:10 ratio, the resonance from the free protein was converted entirely into the bound form (Figure 4.29). Interaction of OfurPBP2 with pheromones was seen to be a slow exchange on the NMR time scale as two different sets of peaks were observed for free and bound states. This indicated very high affinities of ligands toward the protein, which are characteristic features for the formation of a tighter complex.

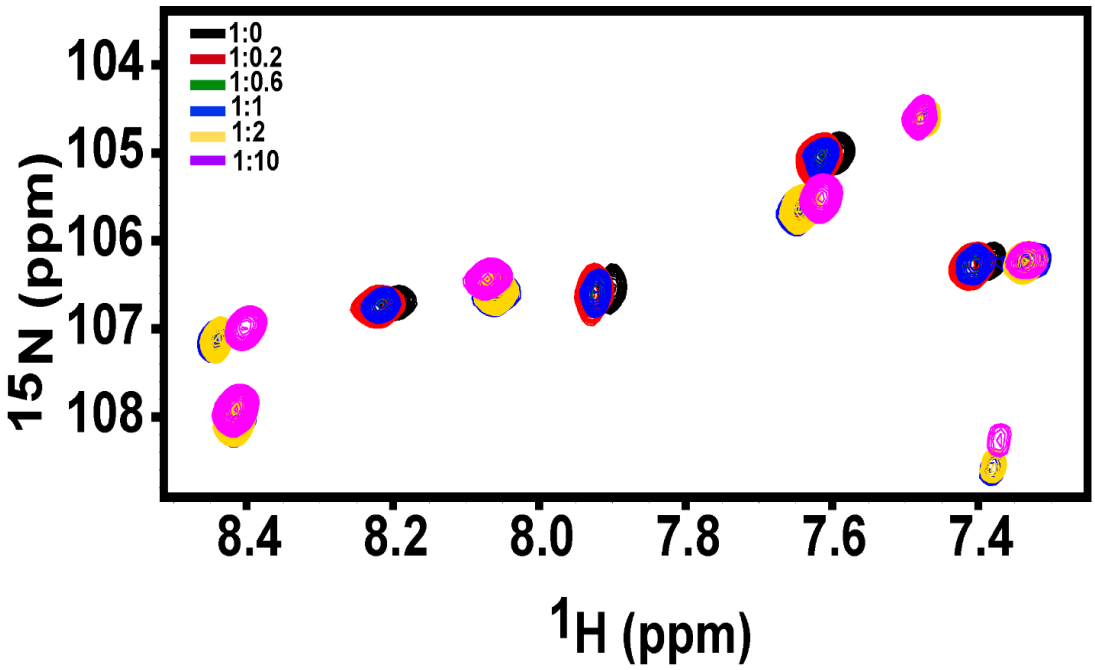


Figure 4.30: Expanded region of the two-dimensional {^{1}H, ^{15}N} HSQC spectra OfurPBP2 titration with Z-pheromone. Protein: ligand ratios are indicated in the figure.

Similarly, titration studies were carried for *E*-12-tetradecenyl acetate (*E*-12-14:OAc) pheromone (Figure 4.29 and 4.30). The behavior was similar to that of *Z*-12-tetradecenyl acetate (*Z*-12-14:OAc) (Figure 4.30). Thus, the pheromone binds with high affinity as evidenced by the presence of peaks from the free and the bound states present in slow exchange on the NMR time scale when the ligand is present in sub-stoichiometric quantities.

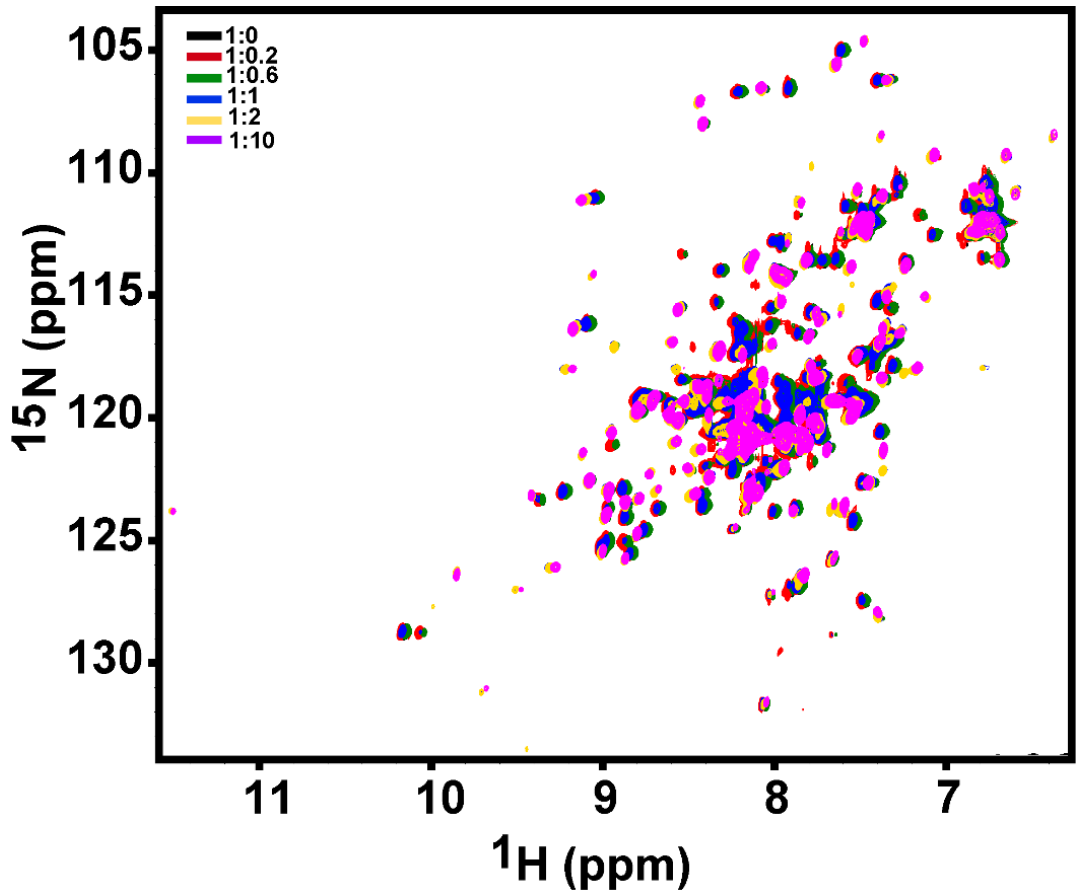


Figure 4.31: Overlay of 2D {^{1}H,^{15}N}-HSQC spectra of OfurPBP2 collected at the different E-pheromone concentrations.

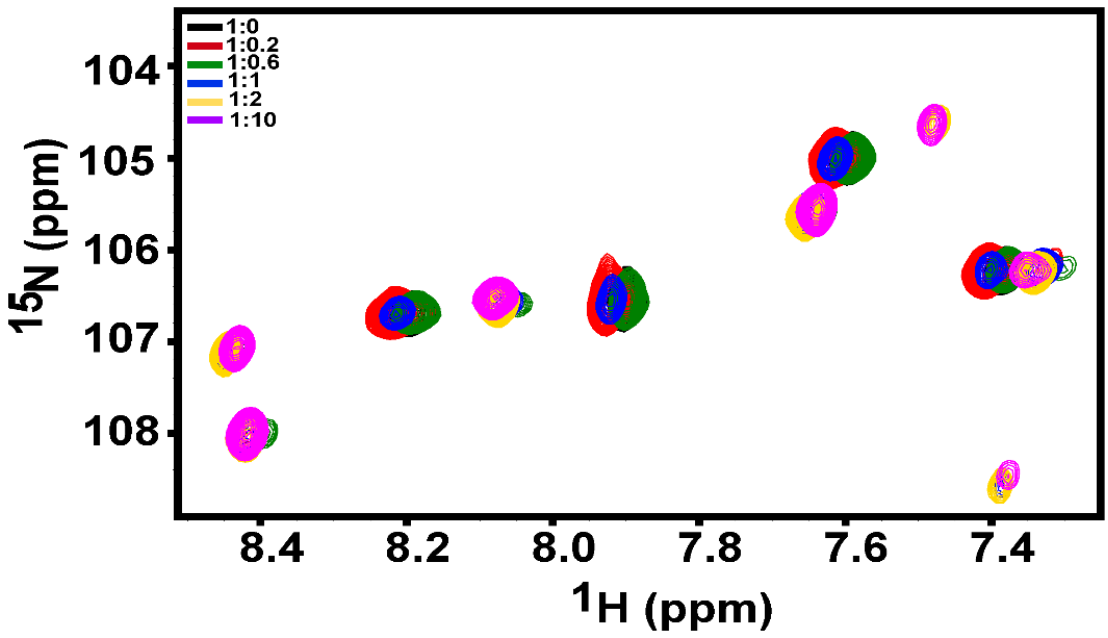


Figure 4.32: Expanded region of the two-dimensional {^{1}H, ^{15}N} HSQC spectra of OfurPBP2 titration with E-pheromone. Protein: ligand ratios are indicated in the figure.

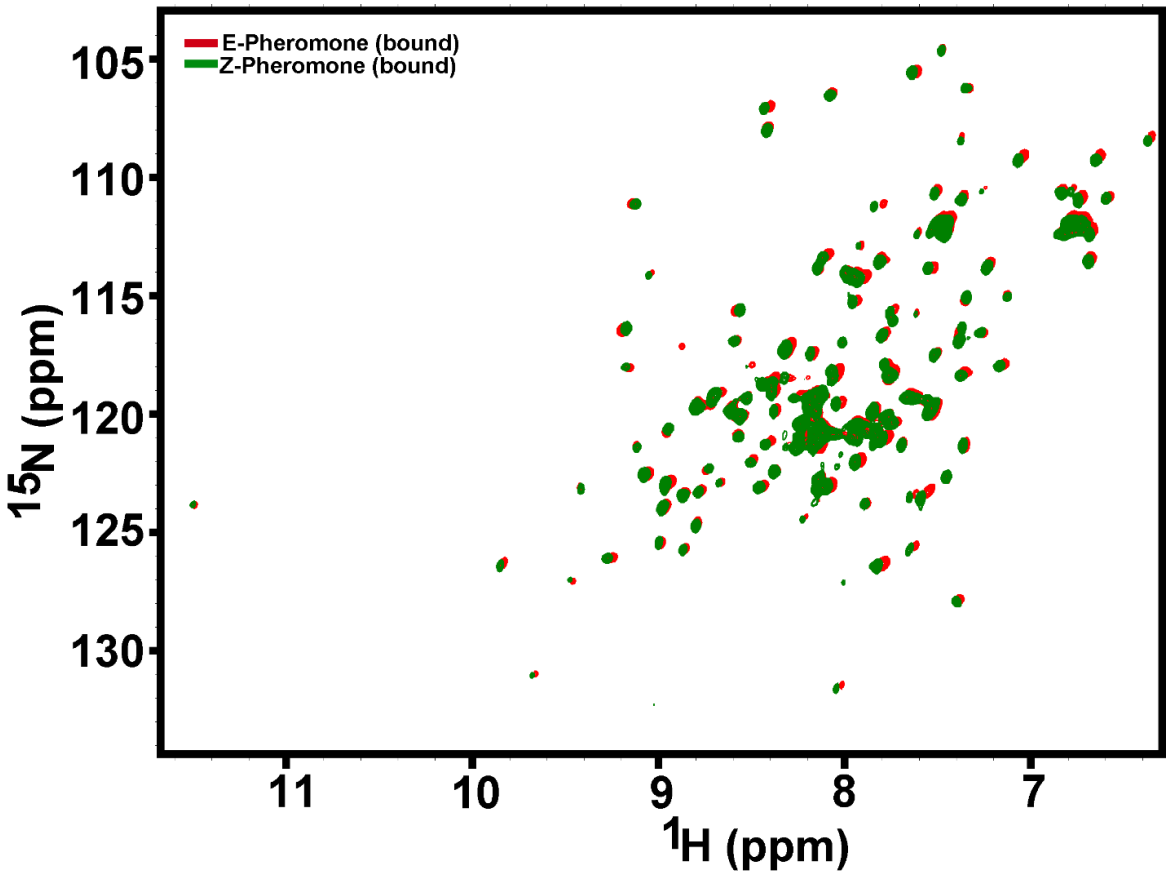


Figure 4.33: Overlay of 2D {^{1}H,^{15}N}-HSQC spectra of OfurPBP2 with 1:10 ratio of protein and E-pheromone (red), 1:10 ratio of protein, and Z-pheromone(green).

The position of the peaks after binding with both Z and Z pheromone were exact matches on the top of the undelipidated protein. This shows that protein has reached saturation. The effects of both pheromones studied on the conformation of delipidated OfurPBP2 were exactly the same (Figure 4.31).

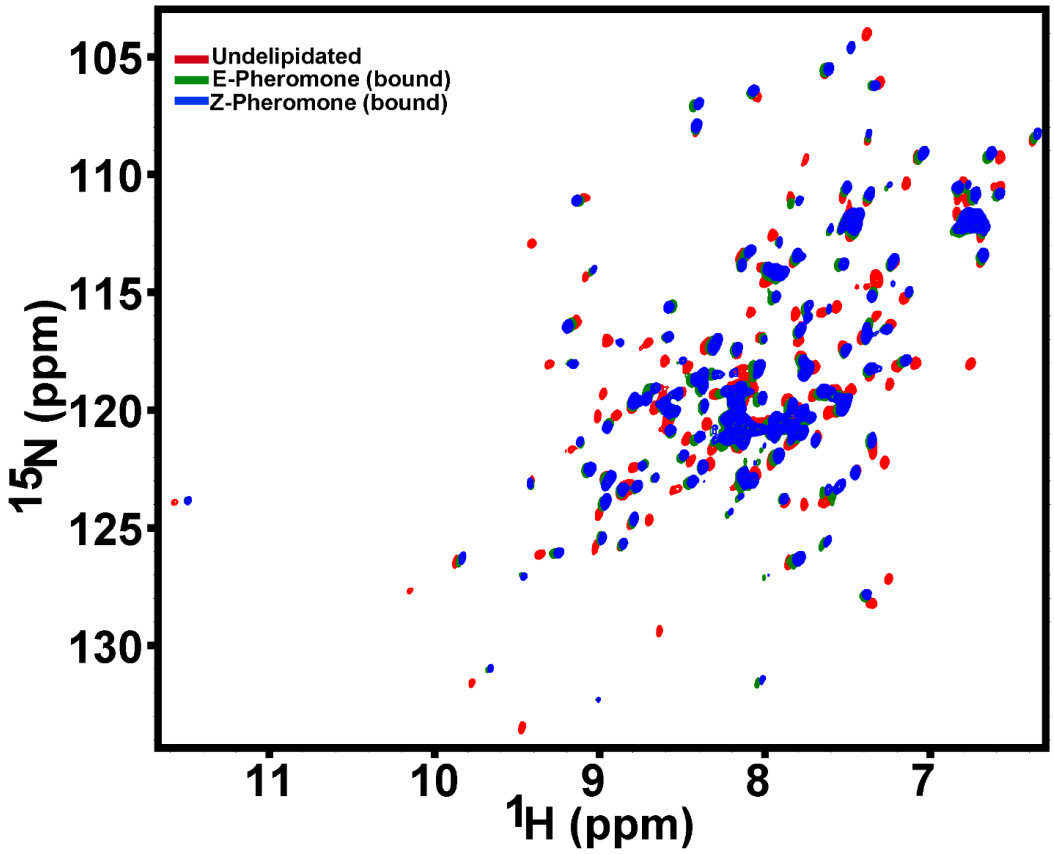


Figure 4.34: Overlay of 2D {^{1}H,^{15}N}-HSQC spectra of OfurPBP2 with Undelipidated protein, E-pheromone bound, and Z-pheromone bound.

The change in the peak intensity, as shown by considering a single peak, is displayed in Figure 4.35.

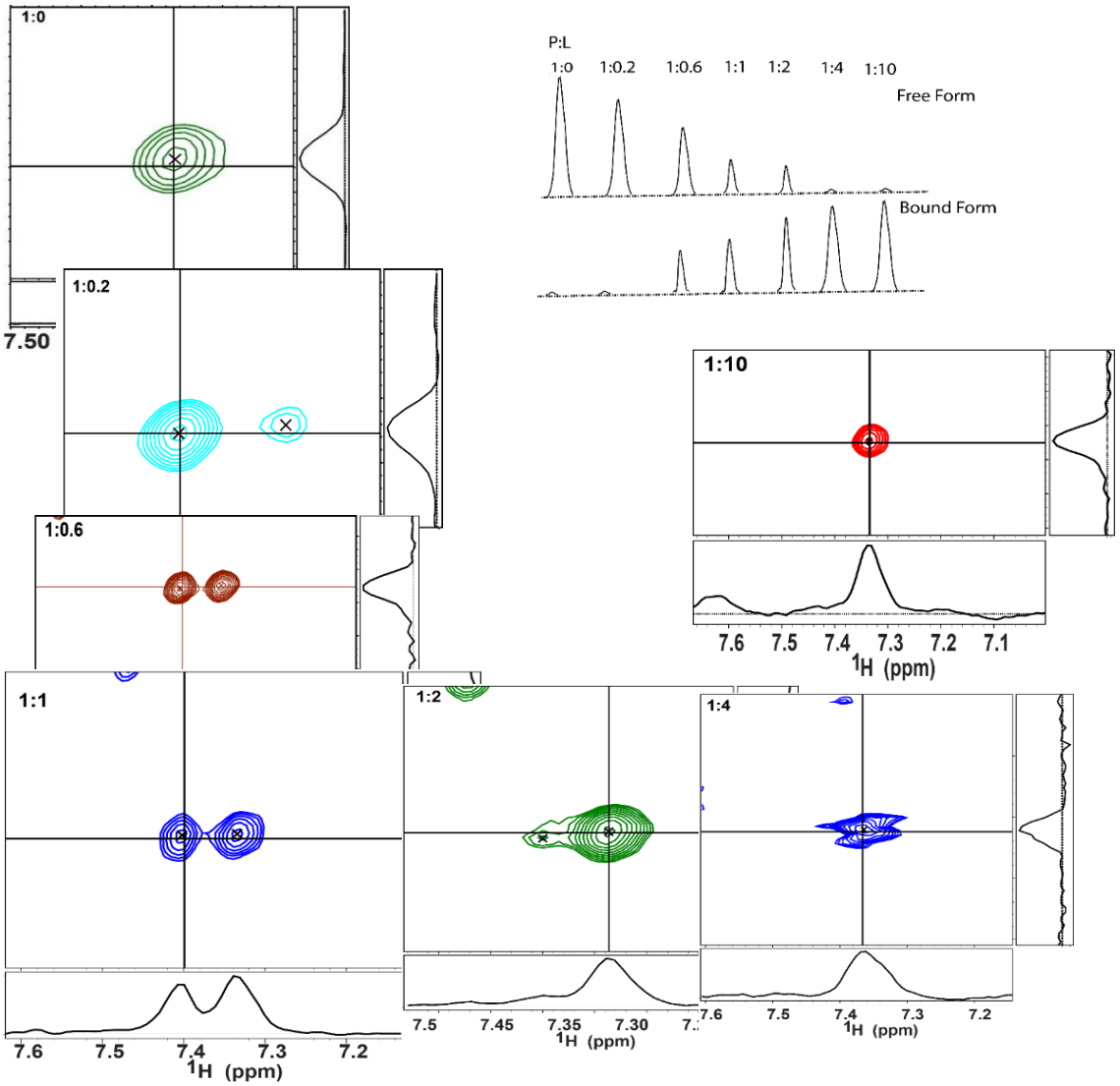


Figure 4.35: One-dimensional slices from the ^{1}H axis for the glycine peak taken showing the intensity of the peak corresponded to a free form slowly decreasing. The intensity of the other peak corresponded to the bound form gradually increasing.

4.3.14. Interaction of Ligand and Protein Fast Exchange

Similarly, the chemical shift of some residues shifted gradually with an increase in the ligand concentration, which indicates that a few residues from the loops are in fast exchange on the NMR time scale. The resonances of several residues located in the β hairpin loop showed severe broadening, indicating conformational exchange phenomena. We attempted to determine the dissociation constants from the NMR titration data by considering the fast exchange peak only. Although NMR spectroscopy has been used efficiently to determine dissociation constants in the micro-molar to the milli-molar range, strong binding with a dissociation constant in the nano-molar range cannot be determined accurately due to an inherent sensitivity issue. The average Kd values were calculated by considering 10 residues and were found to be 340 µM and 510 µM for the E and Z ligand, respectively (Figure 4.36).

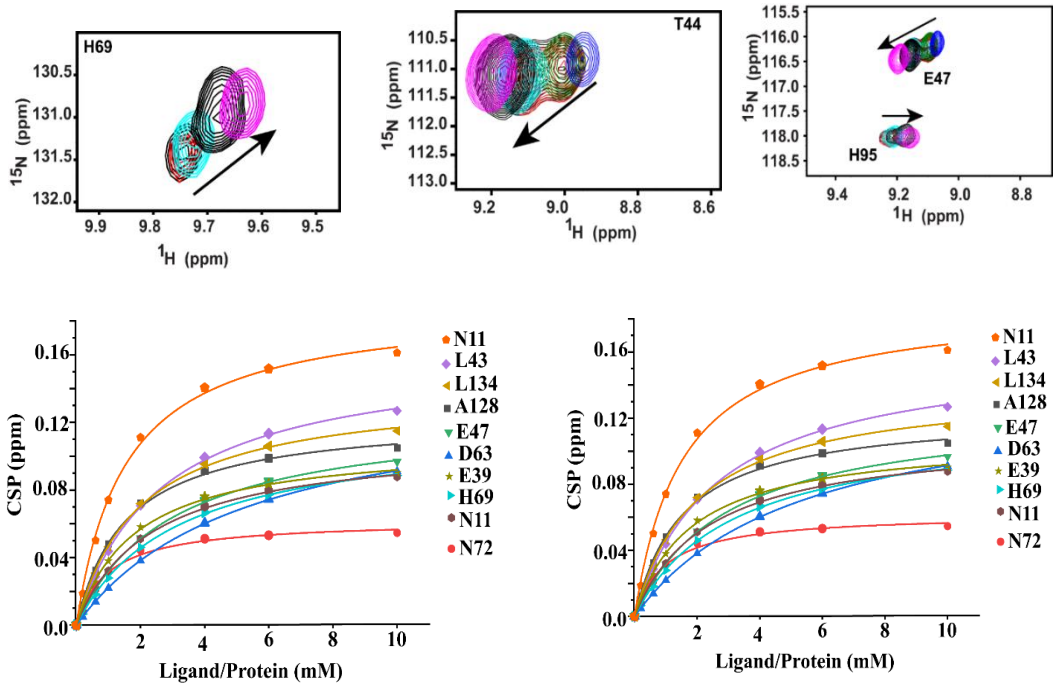
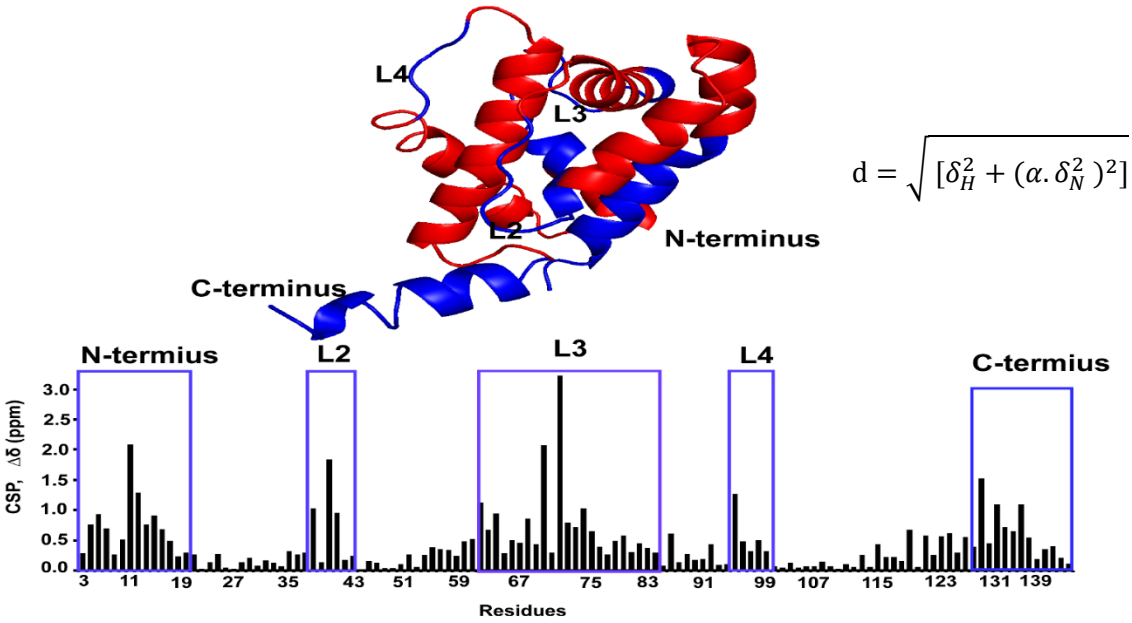


Figure 4.36: Resonance corresponding to fast exchange was shown for His69, The44 Gln47, and His95. The chemical shift perturbation was used to calculate the dissociation constant from Mnova for *E*-ligand (left) and *Z*-ligand (right).

4.3.15. Chemical Shift Perturbation (CSP)

The chemical shift perturbation was calculated between undelipidated and delipidated OfurPBP2. The most considerable chemical shift changes also occurred in the long stretch of loops (L2, L3/flaps, and L4). The chemical shift perturbations for the C- and N-terminus were large. The chemical shift perturbation map for free and bound protein indicated that many residues were affected from different regions of the protein. The residues with large chemical shift changes (greater than 0.5 ppm) of their backbone amide signals were found to cluster in the several areas of the protein, in particular, Ala3−Lys21, Gln39−Leu43, Asp60−Asp84, Val94−Val101, and Pro129-Gln144 (Figure 4.37). The resonances of several residues from the delipidated protein mainly located in the N/C-terminus, β hairpin loop, and loops showed severe broadening, and with occasional peak disappearance, indicating the presence of conformational exchange phenomena. The movement of a particular signal in chemical shift perturbation experiments does not always suggest that the corresponding residue is close to the binding interface. Conformational changes also lead to differences in resonance frequencies. These peak shifts provide information about allosteric changes in the protein upon the binding of a ligand.



$$d = \sqrt{[\delta_H^2 + (\alpha.\delta_N^2)^2]}$$

Figure: 4.37: The chemical shift perturbation showing the largest deviation in the blue in the cartoon diagram of the protein and the blue box in the CSP plot.

Conclusion

Structural studies are imperative to gain a comprehensive understanding of the functional properties of OBPs, including PBPs. The NMR structure of OfurPBP2 at physiological pH has shown that the ligand is occupied in the binding cavity and the C-terminal helix is outside the cavity. The most interesting and noticeable feature in the OfurPBP2 structure is the C-terminal tail. This finding was striking, compared with other PBPs where their C-terminus is a random coil. The NMR structure of OfurPBP2 provides valuable insights into the structural basis of ligand specificity and the binding and release mechanisms. This work contributes to understanding invertebrate chemical communication through a detailed study of pheromone olfaction. Such an understanding would guide efforts to develop pheromone-mimetic semiochemicals to control insect agricultural pests and disease vectors with wide-reaching impacts.

CHAPTER V

STRUCTURAL CHARACTERIZATION OF PROTEIN BY SMALL ANGLE X-RAY SCATTERING (SAXS)

5.1 Introduction

The small-angle X-ray scattering (SAXS) has emerged as a key complementary technique in structural biology. SAXS can be used to obtain macromolecular structures in solution directly, regardless of their size. The data obtained from the SAXS experiment provide information on the molecular weight, oligomeric state, as well as detection of conformational changes due to the effect of pH, temperature, ligand, or mutation on the protein.[254] Several parameters, like the radius of gyration (Rg), forward scattering I(0), porod invariant (Q), porod volume (VP), and the volume of correlation (Vc) can be obtained from scattering data. These parameters give information to interpret the folding states, aggregation, overall shape, conformations of the molecules, size, and distribution of mass within the molecule in the solution state. The three-dimensional low-resolution structural model is generated from the scattering profile between 10-20 Å. In a typical SAXS experiment, the sample is placed in a quartz capillary and exposed to the X-ray beam. The scattered intensity is recorded and radially averaged to obtain the scattering curve. The scattering of X-ray photons occurs due to the interaction with electrons in the sample. When the protein sample is irradiated with X-ray photons, the electrons on the sample scatter photons in all directions with the

same wavelength λ as the incident wave called elastic scattering. The scattering vector "q" is defined as (4π sin θ)/λ. The detector then records the beam scattered by the solution. The scattering is anisotropic and two-dimensional (2D) detector images are reduced to one-dimensional (1D) scattering profiles (Figure 5.1). These profiles, after background subtraction, are used to analyze and acquire structural information including the three-dimensional models. The scattering vector "q" is defined as (4π sin θ)/λ. The detector then records the beam scattered by the solution. The scattering is anisotropic and two-dimensional (2D) detector images are reduced to one-dimensional (1D) scattering profiles (Figure 5.1). These profiles, after background subtraction, are used to analyze and acquire structural information, including the three-dimensional models.

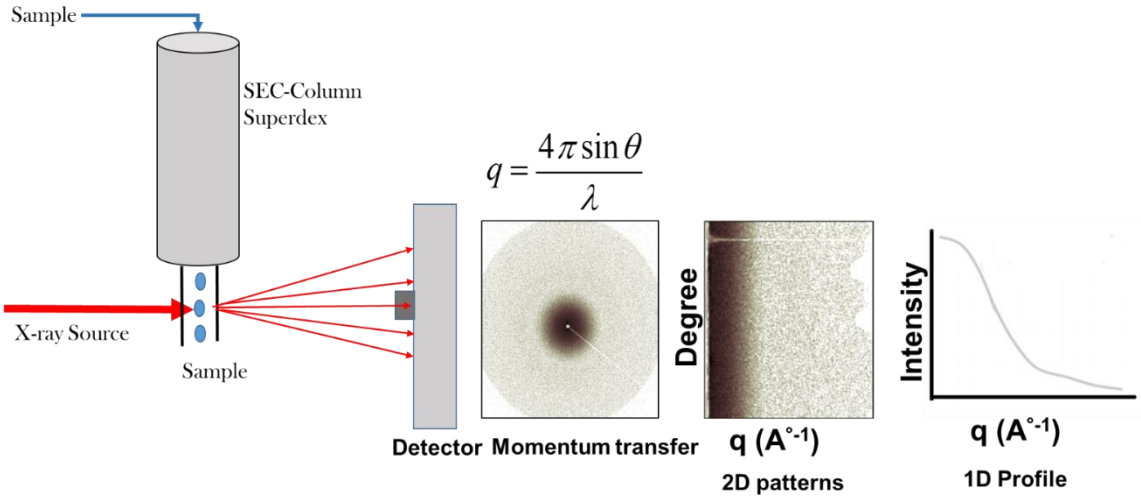


Figure 5.1: A schematic of a SAXS experiment. An X-ray source irradiates a sample. Interactions between the X-rays and the atom of the sample cause a portion of the incident beam to scatter with a certain angle with the intensities, I(s), which is recorded by a 2D detector, where, s = 4π sinθ/λ (λ is the incident radiation wavelength and 2θ the scattering angle). Figure adapted from reference 213.[255]

In this chapter, the detailed investigation of OfurPBP2 is discussed. Size-exclusion chromatography-coupled with SAXS (SEC-SAXS) was used, which is an effective technique to obtain highly pure, stable, and homogenous samples. Size exclusion chromatography (SEC) purification removes aggregates and precipitates from the protein sample. From SAXS data, low-resolution structural properties of OfurPBP2 were obtained.

5.2 Materials and Methods

5.2.1. Sample Preparation for SAXS Experiments

The protein samples were expressed and purified as discussed in earlier chapters. Data were collected at a high radiation synchrotron source. The purity of the sample was assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Size exclusion chromatography was used to monitor a monodisperse protein sample. The major problems that are usually encountered during the SAXS data collection are aggregation and radiation damage. Radiation damage changes the SAXS scattering curve. To reduce radiation damage, 5% glycerol was added to the sample since it acts as a good radical scavenger. The SAXS experimental data were subtracted from the noise that was obtained from the buffer as a blank. Data were collected for the exact matching buffer. 50 mM HEPES buffer pH 6.5 with 5% glycerol and 50 mM acetate buffer with 5% glycerol for pH 4.5 was used.

5.2.2. Size Exclusion Chromatography-Small Angle X-ray Scattering measurements (SEC-SAXS)

Small Angle X-ray Scattering (SAXS) experiments were carried on the BioCAT beamline 18-ID at the APS of Argonne National Laboratory. The beamline source was 3.3 Undulator, Si (111)

monochromatic with incident X-ray flux at 2 × 1013 photons s^{-1} at 12 keV. A MAR165 CCD detector (Rayonix Inc., Evanston, IL, USA) was used to detect the scattered patterns for in-line SEC-SAXS. The sample-to-detector distance of 3.7 m with a q range 0.004-0.33 $Å^{-1}$ was used. The FPLC was equipped with a UV-monitor capable of measuring UV absorbance wavelengths (280 nm). The sample runs through a size exclusion column to separate potential aggregates or different oligomeric states if present immediately before exposure to the X-ray beam[256]. The SEC-SAXS separates the protein potential aggregates, oligomers, or breakdown products. The continuous flow of samples reduces the radiation damage. The default flow rate (0.7 mL/min) was used for SEC-SAXS experiments, and the experiment took around 40 min. For SEC-SAXS, 0.5 mL protein samples at 20 mg/mL were loaded onto the Superdex-75 columns (GE Healthcare). Columns were connected in-line with the flow cell for SAXS data collection and 1s exposures taken at 5 s intervals during elution. The UV absorbance at 280 nm was monitored during chromatography for the separation of monomeric species.

a. Data Processing and Normalization

When protein samples were eluted from the SEC column, the SAXS data were collected across the eluting peak. Approximately 459 images were collected, including buffer region, which were then averaged, normalized, processed, and analyzed by using the ATSAS package and Scattered software. The sample images from 110 to 210 were taken as blanks for the background subtraction from the sample images from 265 to 330. The scattered intensity plot was thus obtained. During the elution, the concentration of protein at each point is not known. The SAXS data were collected along with the buffer. The buffer was subtracted from the data. i.e., solution minus solvent. The curve obtained after the subtraction of the buffer is known as the scattering profile. The visual inspection of the scattering profile helps to identify the quality of the data.

b. Data Processing

Normalization, buffer subtraction, and data reduction were performed with the ATSAS[257, 258] RAW [259,260], and SCATTER.[261] The general scattering profile provides basic information on the quality of the data. Guinier plots are the first quality control method used to determine data quality. The parameters, like the radius of gyration (Rg) and zero angle scattering, I(0), were calculated from Guinier analysis with the program PRIMUS.[262] The compactness of the structure was analyzed using the Kratky plot. The pairwise distribution function P(r), provides I(0), Rg, and the maximum dimension (D_{max}) of the particles.

c. Guinier Analysis Plots

Guinier analysis was used to evaluate the scattering data at a very small scattering angle. The plot of the natural logarithm of scattering intensity, I(q), with the square of the amplitude of the scattering vectors is called the Guinier plot. The Guinier plot follows the Guinier law given in equation 5.1. The lowest portion provides the information on the radius of gyration (Rg). The Rg is the average electron density-weighted squared distance of the atomic distance from the center of mass. The Guinier plot [plot of the log I(q) versus q^2] gives a straight line from which values of Rg and I(0) can be obtained.

$$\mathbf{Ln[I(q)]} = -\frac{\mathbf{q2Rg2}}{\mathbf{3}} + \mathbf{ln[I(0)]} \qquad (5.1)$$

where, I(q)= scattering intensity, Rg = radius of gyration, I(0) = forward scattering.

The Guinier region reveals an important behavior of the protein samples, such as the presence of aggregation and inter-particle interference. The lack of linearity in the Guinier region indicates inter-particle interference, aggregation, or radiation damage. The forward scattering, I(0), is an intensity measure at zero angles (q = 0). The zero angle scattering intensity cannot be measured experimentally. This can be estimated by extrapolation of scattering intensity at zero angle and is unaffected by the shape of the particles.[263] The forward scattering, I(0), is used to obtain an apparent molecular weight of the particles. For globular proteins, the product of scattering vector magnitude (s) and radius of gyration (Rg) value should not exceed 1.3. For the large particles, we can lower q_{min} to obtain the Guinier region for Rg determination. The value of q_{max} will depend on the shape of the molecule, i.e., for spherical particles $q_{max} < 1.3*Rg$ and for elongated particles $q_{max} < 0.8*Rg$.[264]

d. The Kratky Plot

The plot of I(q)*q2 versus q is called the Kratky plot. This plot is sensitive to the morphology and compactness of the particles. The Kratky plot is used to differentiate between folded and unfolded states of the macromolecules. The compact folded globular macromolecules exhibit a bell-shaped curve (parabola) with a prominent peak at low q angle[265], and the extended unfolded macromolecules have plateaued at a higher q region. Porod invariant "Q" is the integral of the area of the Kratky curve. It is a concentration-independent value and is proportional to molecular mass.

The Kratky plot is used for the detection of flexibility in the proteins. The scattering intensity of elongated/unfolded particles decays slower (proportionally to q^{-1}), which could be observed in the Kratky plot as a plateau. The flexible shape of scattering data in the Kratky plot is the indication of flexibility.[266] The well folded globular protein follows Porod's law, which states that the scattering intensity of compact, globular particles decay proportionally to q^4, which could be observed as a

bell-shaped curve in the Kratky plot. For the well-folded particles, there is a sharp peak at low q values, and then then it returns to near zero (Figure 5.3). Completely unfolded proteins will have a sharp rise at low q values with a plateau. Partially unfolded proteins will have a broader peak at low q values than folded proteins but will remain at a higher level at high q values and not return to zero (Figure 5.2). The Kratky plot is used to identify the flexible multi-domain proteins. The information obtained from a Kratky plot is used to identify the expected conformations and dynamics of the protein in the solution.[265]

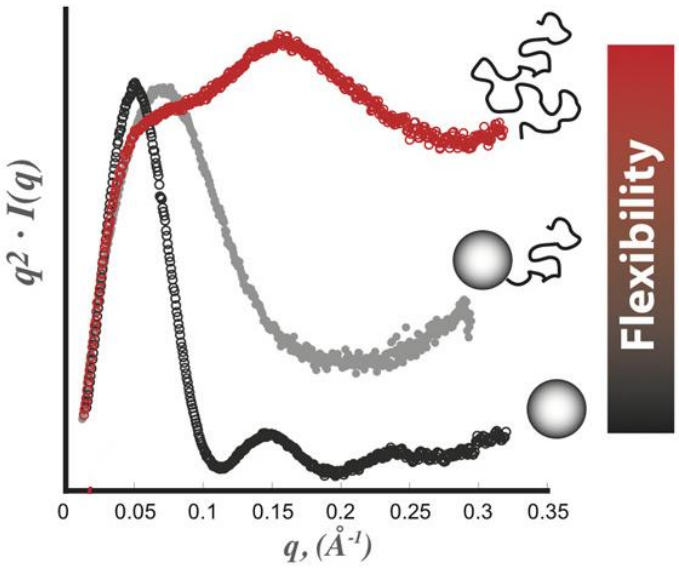


Figure 5.2: Kratky plot of scattering data illustrating changes in the behavior of the curve for folded (sphere), partially folded (sphere-random coil), and completely unfolded particles (random coil). Figure adapted from.[267]

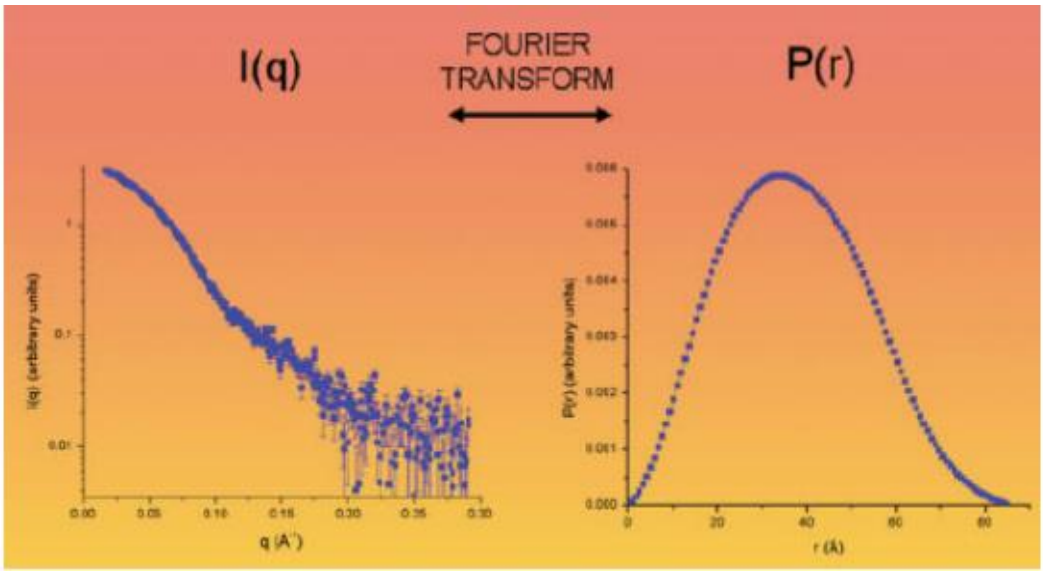
e. Pair-wise Distribution Function P(r)

Pairwise distribution function P(r) is the Fourier transform of the scattering curve I(q) into real space. The P(r) distribution was obtained from program GNOM.[268] It is related to the frequencies of the distances within the particles. P(r) distribution provides the distribution of distances between pairs of particles within a given volume (Figure 5.3). The overall structure of a protein can be

determined by the shape of the P(r) distribution.[269] P(r) function is equal to 0 at r = 0 and $r \geq D_{max}$, where D_{max} is the maximum linear dimension of the particles. The P(r) function also indicates the data quality. Sometimes unfolded proteins may not reach zero even at r = 0. It is difficult to determine D_{max} for the extended structures. The Fourier transformation was applied to the data by using equation 5.2 to obtain information in real space.

$$\mathbf{P(r)} = \frac{\boldsymbol{r2}}{\mathbf{2\pi r}} \int_{\mathbf{0}}^{\infty} \mathbf{q2} * \mathbf{I(q)} \frac{\mathbf{sin(qr)}}{\mathbf{qr}} \ \mathbf{dq} \qquad \textbf{(5.2)}$$

A



B

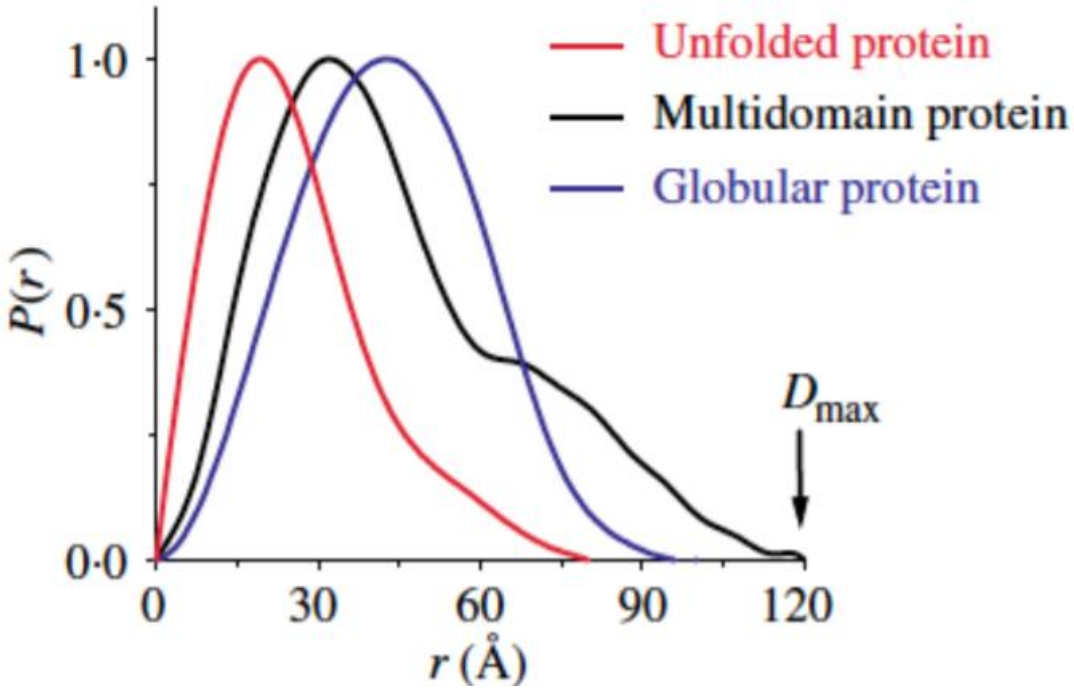


Figure 5.3: A) Fourier transformation of a scattering curve, B) The P(r) functions for flexible (unfolded), multi-domain, and globular proteins.[270]

The P(r) distribution also calculates Rg and I(0). The Rg obtained from the pairwise function is more accurate than the Guinier plot because it includes the entire data range. When samples are aggregated, the P(r) distribution will not smoothly approach zero. The high D_{max} (7-8*Rg) can provide some useful information, but the P(r) distribution curve may have several peaks at a high r-value.[271] For aggregated samples, distribution generally does not reach zero.[269] The distribution function shows that the compact globular particle has a smooth, symmetric bell-shaped P(r) with a non-negative value and reached zero at the maximum dimension, whereas unfolded particles have an elongated distribution curve.[272]

f. Calculation of Ab-initio Models

The 3D structure can be generated from the one-dimensional scattering profiles through ab-initio modeling using programs DAMMIN[273] and DAMMIF.[274] DAMMIN is based on the volume and Rg. DAMMIF is based on simulated annealing to construct a compact interconnected model yielding a scattering pattern that fits the experimental data. The program DAMAVER[275] is used to align ab-initio low-resolution models and build an average model. The accuracy of the averaged model is based on the quality of the data. The model obtained can be superimposed with the NMR or X-ray crystal structure using CRYSOL[276] to determine the structural differences.

g. Molecular Weight Calculations

The molecular weight can be estimated from the Porod volume, where MW = Vporod ($Å^3$)/1.6.[277] Secondly, the molecular weight was obtained using the program SAXSMoW

(http://www.if.sc.usp.br/~saxs/). Another way to calculate the molecular weight is by using a standard value of bovine serum albumin (BSA), using equations 5.3 and 5.4.

$$\text{MW of Sample} = \text{Vporod (Sample)}*\text{Mw(BSA)}]/\text{V porod (BSA)}^{263} \quad (5.3)$$

$$\text{MW of Sample} = \text{MW BSA} *\text{I (0) sample} /\text{I (0) BSA}^{263} \quad (5.4)$$

where Vporod of BSA was taken as 118 kDa and MW of BSA was taken as 66.2 kDa.

The molecular weight (MW) determination helps to identify the oligomeric state of the protein. MW can also be calculated by taking water as a standard. This procedure allows one to measure the MW particle by placing the scattering curve on an absolute scale (in units of cm^{-1}) instead of an arbitrary scale using equation 5.5.

$$\text{MW particle} = \text{I(0) particle}*\text{NA}/ \{\text{particle C } (\Delta \rho.\nu)^2\}^{277} \quad (5.5)$$

where, NA is Avogadro's number, C is the concentration (in g/cm^3) is the particle concentration, $\Delta\rho$ is the contrast (in e/cm^3), and ν is the partial specific volume (in cm^3/g).

5.3. **Results and Discussion**

5.3.1. SEC-Small Angle X-ray Scattering (SAXS) Analysis of OfurPBP2 at 6.5

Size exclusion chromatography small-angle x-ray scattering (SEC-SAXS) was performed in HEPES buffer at pH 6.5. SEC-SAXS removes the aggregation or degradation product. The pure OfurPBP2 was injected into the size exclusion column, and data were collected as the protein was eluted at each point of the peak (Figure. 5.4). The signal plot, shown as a dotted line across the peak, demonstrated that R_g of the protein is independent of protein concentration, suggesting that

scattering is also independent of concentration.[145] (Figure 5.4). In SEC–SAXS data sets, the subtraction of the buffer depends on the regions selected for the buffer. One of the major issues that can be resolved from SEC–SAXS is to identify oligomeric species, as the sample is continuously scattered during elution. During the elution, the concentration of protein at each point is not known. The Radius of Gyration over the datasets frames of elution peak is stable which is shown by the red flat line in Figure 5.4. Ofur PBP2 recombinant protein was monodisperse as SEC–SAXS shows the single peak i.e. monodisperse. Red dots showing the Radius of Gyration (Rg). The Radius of Gyration is constant over the image number showing Rg is concentration independence.

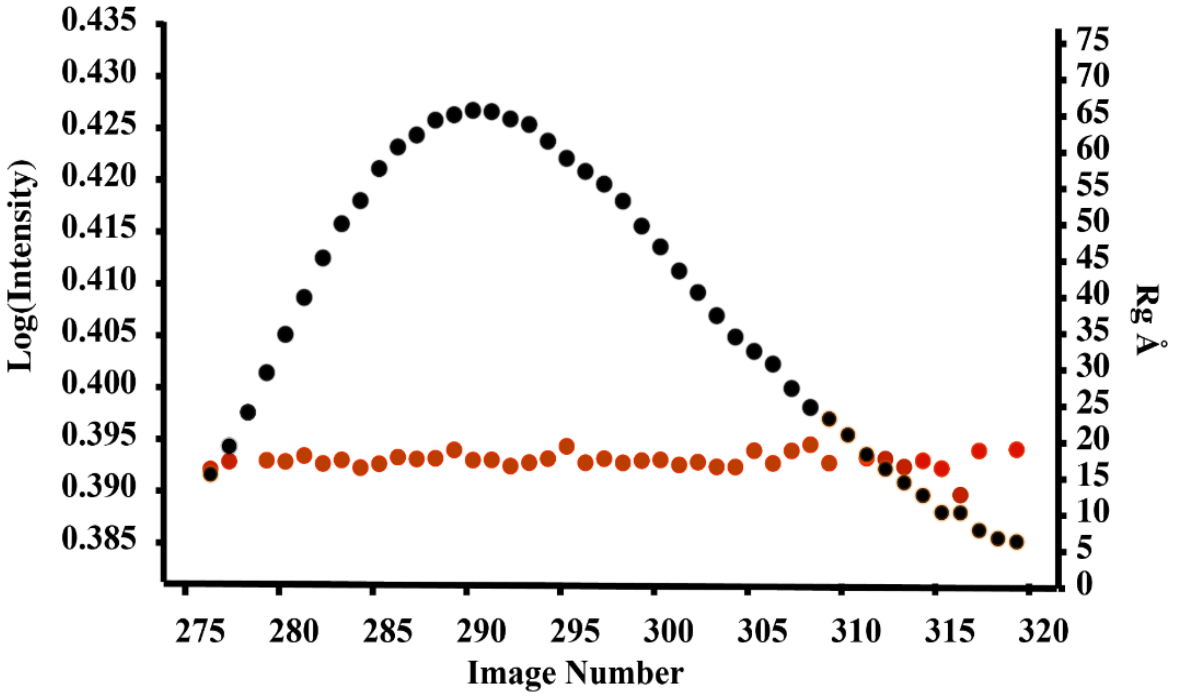


Figure 5.4: Log intensity versus sample image number. Diagram of the SEC-SAXS mode, the sample is loaded on top of a gel filtration column before being eluted by the buffer. As the protein elutes, a UV spectrophotometer monitors through protein absorbance at 280 nm shortly before the eluted solution reaches the SAXS measuring cell. The elution volumes 273 mL to 325 mL were considered for the SAXS measurement. SEC-SAXS shows the single peak, i.e., monodisperse. The

signal plot, shown as a dotted line across the peak, suggests that R_g is independent of protein concentration.

The scattering intensity of the sample is higher than that of the buffer, and thus the buffer needs to be subtracted (Figure 5.5A). The resultant curve obtained after the buffer subtraction is called the intensity plot. The SAXS intensity plots of log[I{q}] versus 's' or 'q' demonstrated that the protein solution was homogeneous and monodisperse without any aggregation or inter-particle interaction (Figure 5.5B).

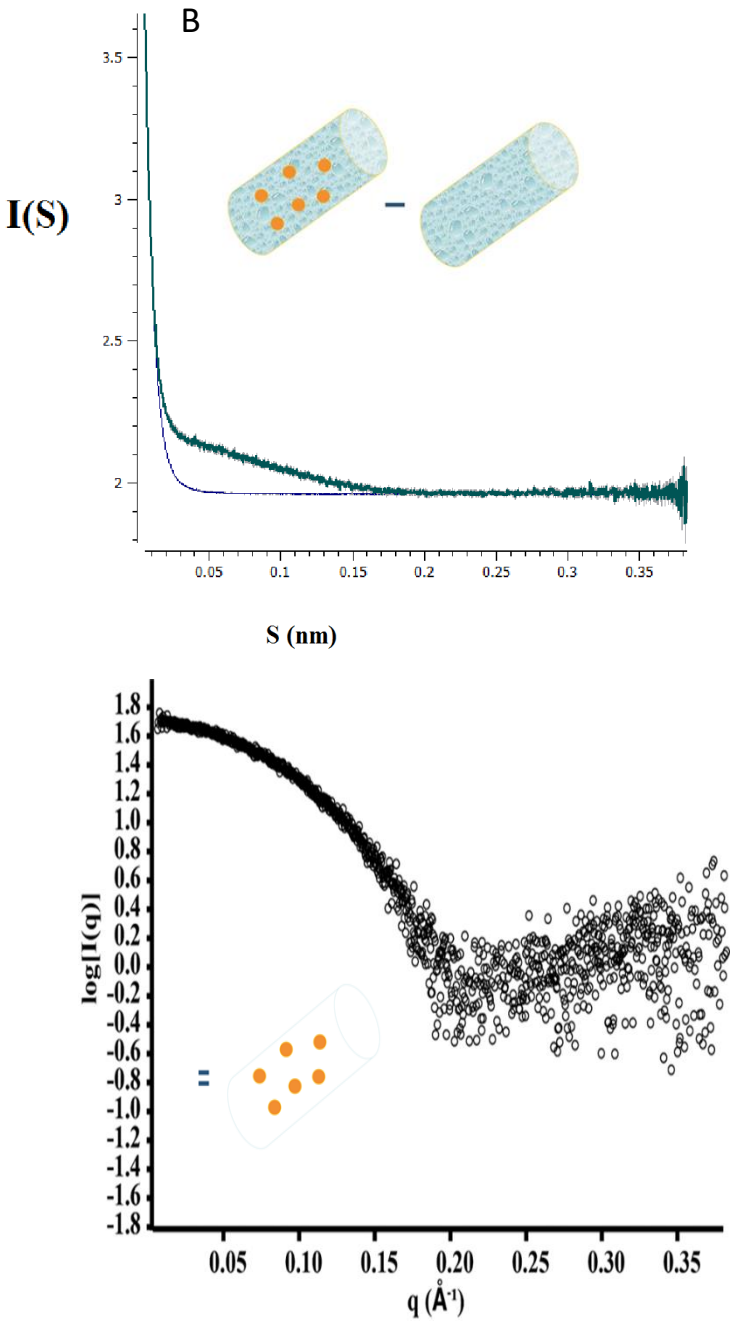


Figure 5.5: A) Buffer subtraction B) Scattering plot indicates that the protein in solution is homogeneous and monodispersed.

The overall size of a protein is obtained from the Guinier plot, which estimates the radius of gyration, Rg, and the forward scattered intensity, I(0). A Guinier plot is constructed by plotting ln I(q) against q^2. The Guinier plot is linear at a low q range, indicating a monodisperse solution without any non-specific aggregation during data collection. An upward curve represents the non-linear trend towards low q, due to aggregation. Similarly, the downward turn on the Guinier curve represents the inter-particle repulsion (Figure 5.6).

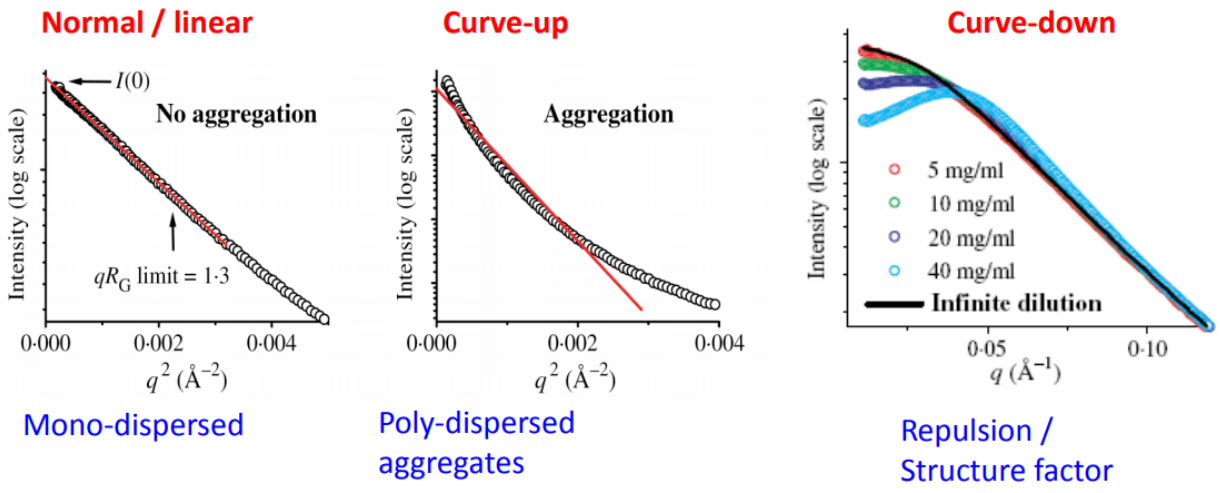


Figure 5.6: Guinier plot showing three possible scenarios: straight line represents monodispersed particles, upturn shape at low q related to aggregation, radiation damage, inter-particle interactions, and high polydispersity of the sample protein, downturn shape at a low angle suggests inter-particle repulsion in the sample.[272]

The Guinier plots are sensitive for detecting aggregation and radiation damage. Figure 5.7 is a Guinier plot obtained for OfurPBP2. From the slope of the Guinier plot, we can get the value of the radius of gyration (Rg) and I(0). The Rg is an important parameter that can explain the overall size of the molecule, while I(0) depends on the molecular weight times the concentration. The Guinier analysis confirms the absence of non-specific aggregation in OfurPBP2 samples with an Rg of 16.96 Å.

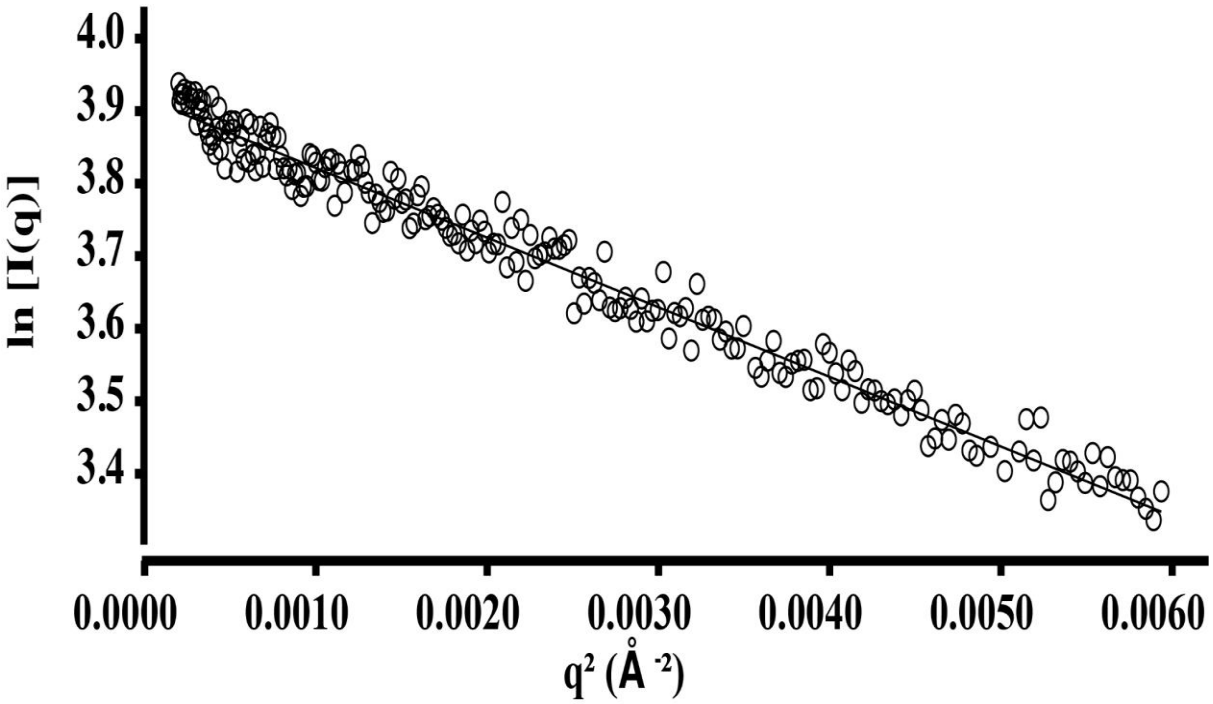


Figure 5.7: The linear low-q regions of the scattering curves used for the Guinier analysis confirm the absence of non-specific aggregation in OfurPBP2 samples with an Rg of 16.96 Å.

Kratky plot is an excellent tool for evaluating the folding behavior of the sample. The Kratky plot [q2 I (q) as a function of q, which can be calculated directly from the scattering curve shown in Figure 5.8. The Kratky plot generated a bell-shaped curve with a well-defined maximum (Figure 5.8), which unequivocally established that OfurPBP2 is a homogeneous, well folded, compact globular protein as it follows Porod's law. The homogenous and well folded globular compact particles follow porod law, which means the particle's intensity decay is proportional to s^{-4} at a higher angle.

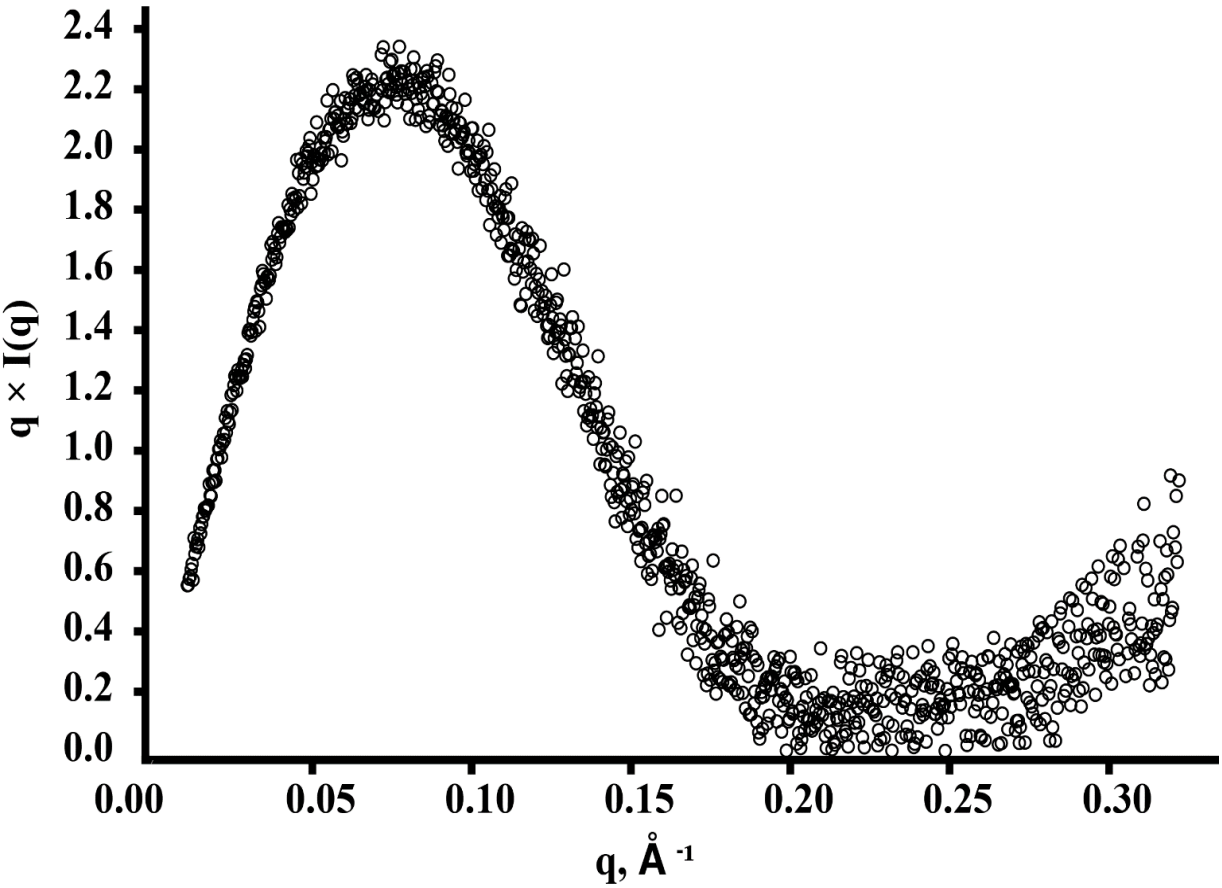


Figure 5.8: Kratky plot suggests that the protein is well-folded and has a globular shape.

SEC-SAXS is employed to identify oligomeric species, as a sample is continuously scattered during elution. During the elution, the concentration of protein at each point is not known. The Porod volume (V_p) obtained for globular particles can be used to estimate the mass. While for highly flexible or disordered proteins, the estimation of mass from this approach is invalid.[278] Typically, for a globular protein V_p (in nm^3), is 1.5 to 2 times the molecular weight in kDa.[269] The excluded volume of the hydrated particle, also known as the Porod volume for the globular protein, is 26114 $Å^3$, which is estimated using the volume interface of SCÅTTER.

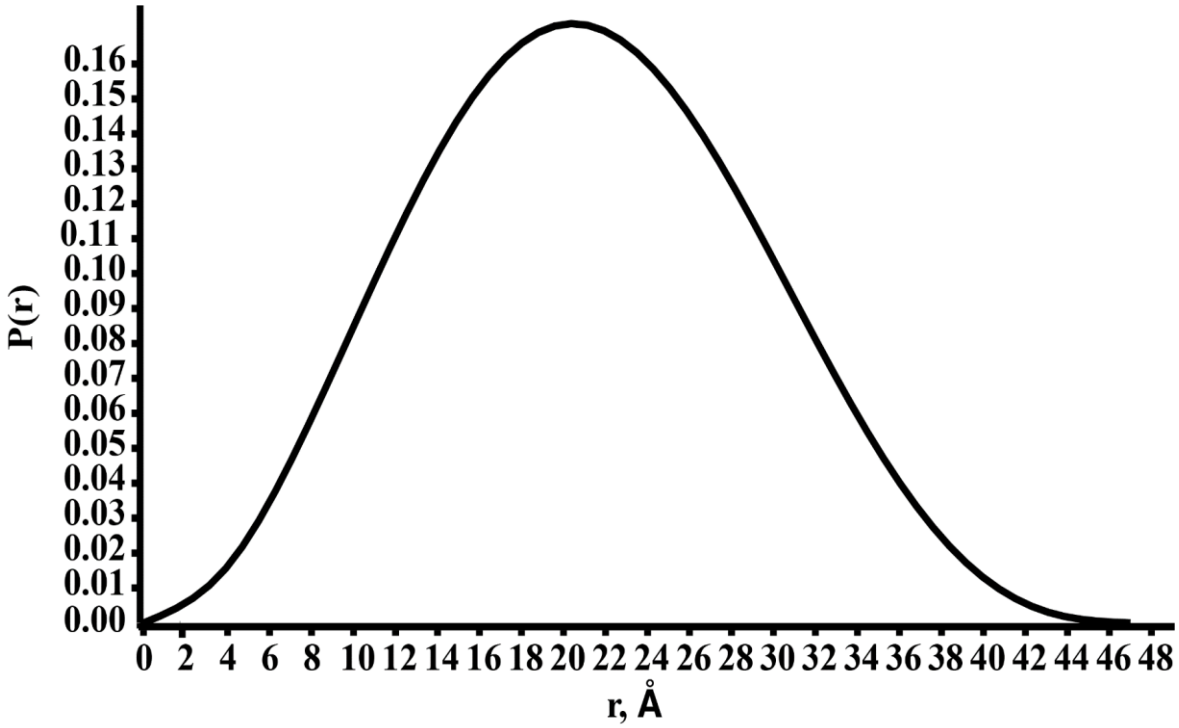


Figure 5.9: The normalized pairwise distribution function shows that OfuPBP2 has a globular shape.

The P(r)-distribution is determined using an Indirect Fourier Transform (IFT) method. Pair-wise distribution tells about the inter-particles interaction effect. The distribution function shows the globular compact particle has a smooth symmetric bell-shaped p(r) and non-negative and reaches zero at maximum dimension, whereas unfolded particles have an extended tail. The data shows that the macromolecular particle is globular and compact with the radius of Gyration is 16 Å. The Rg determined from this P(r) is called real space Rg. The Rg value obtained from both real space and reciprocal space is in good agreement which indicates that data obtained were well measured. Unfolded proteins will not end with zero at r = 0 in P(r) functions, and non-zero values at r= D_{max} indicate aggregation or improper background subtraction. The calculation of D_{max} is prone to error also sometimes difficult for extended structures and globular structures with disordered /unstructured N- and C-termini in proteins. The D_{max} is the maximum inter-particles distance obtained from The P(r) function is 47.0 Å shown in Figure 5.9. The P(r) function gives a symptom of homogeneity. The radius of gyration was determined from Guinier analysis as 16.96 Å, suggesting OfurPBP2 is a globular protein, which was further confirmed by the symmetrical bell-shaped curve for the pairwise distribution function P(r) (Figure.5.9).

To estimate the molecular weight, the Porod volume approach[279] was used where MW= Vporod ($Å^3$)/1.7 The molecular weight obtained from this approach was 15361 Da. The molecular weight was also obtained using the program SAXSMoW[280], which gives a molecular weight value of 15165 Da. Lastly, BSA as a standard value was used to calculate the molecular weight of OfurPBP2 by using the formula, MW of Sample = Vporod (Sample)*Mw (BSA)]/Vporod (BSA).[281] Vporod of BSA was taken as 118, and MW of BSA was taken as 66.2 kDa[281], which produces the value of 14650 Da. All such methods provide a molecular weight close to a monomeric protein.

The ab initio molecular models were reconstructed by DAMMIF[282] and were fitted to the predicted model (Figure 5.10). The low-resolution envelope model suggested that the molecule in solution behaved like a globular-shaped particle. The theoretical small-angle scattering curves were back-calculated from the predicted model and compared to the experimental scattering plots. The theoretical Rg (16.90 Å) obtained was comparable to the experimental Rg (16.96 Å). Similarly, the theoretical maximum particle dimension (D_{max}) of the model was 47 Å, which was closed to the experimental D_{max} of 49 Å, indicating a good fit and a similar shape.

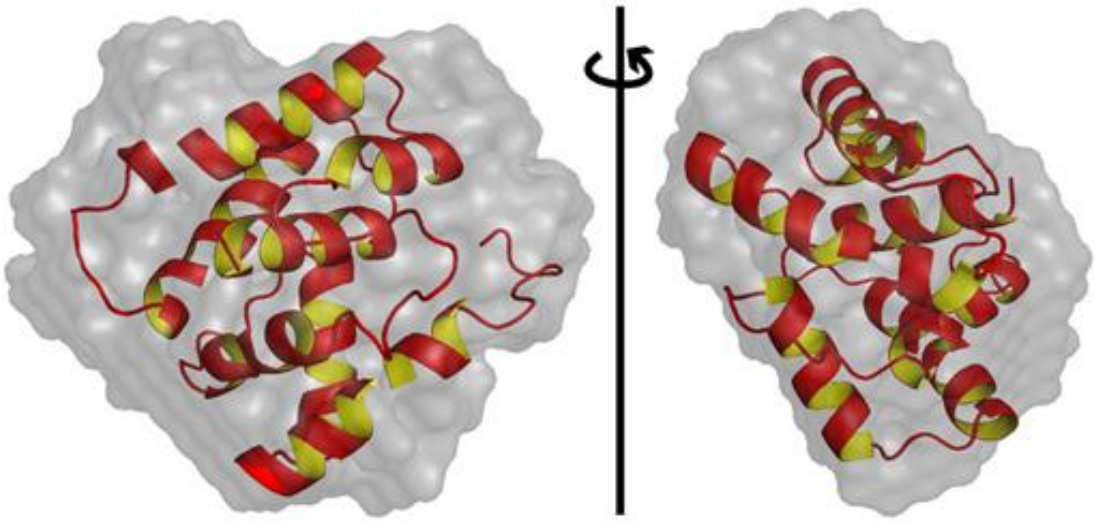


Figure 5.10: Superposition of the SAXS ab initio envelope (grey) with the OfurPBP2 model. The right-hand view is rotated 90 degrees (side view).

5.3.2. Conclusion

We performed small-angle X-ray scattering experiments and calculated a low-resolution molecular envelope. The Guinier plot provided the information on sample aggregation, homogeneity, and radius of gyration (Rg) to validate our result. The monomeric particles were independent of the protein concentration, and the protein does not form aggregates and no had radiation damage. The scattering data showed a linear correlation in the Guinier region for the scattering vector range, suggesting that the molecules were free of significant intermolecular interference and aggregation. Thus produced scattering curves show no evidence of aggregation or interparticle repulsion after buffer subtraction. There is no significant deviation from linearity which suggests that there are no aggregations and repulsion effect. However, some of the curves show slight deviations from linearity (curves not shown) which may be due to buffer scaling rather than inter-particle effects. Guinier plot gives the radius of gyration, Rg, and the forward scattering intensity I(0). Rg is a measure of the effective size of the sample. Rg can be determined in two ways by using the Guinier approximation for the low-resolution scattering ($qRg < 1.3$ for globular scatters) is 16.96, which was determined from the slope of the linear fit. The intercept gives the forward scattering I(0) is 55.26. Non-linearity in the Guinier plot is an indicator of a lack of monodispersed and/or the presence of attractive or repulsive interactions between scatters. Secondly, Rg is determined from and from the pair-distribution function P(r), which is a histogram of all-electron distance in scattering particles. Sometimes there present some aggregates, at a very small angle, systematic removal of data points at the very low angle portion in GNOM analysis diminished the effect of aggregations. Furthermore, the radii of gyration determined at different protein concentrations were very close. There was no observation of concentration-dependent behavior. The characteristic curves of the Kratky plot qualitatively provided information on the degree of a partially folding and globular protein. The unfolded protein showed a plateau in the larger q region, and there was a lack of a well bell-shaped curve. However, for the globular protein, the Kratky plot had a well bell-

shaped curve since the symmetrical bell-shaped Kratky plot was a clear indication of a globular protein. SEC-SAXS is carried on beamline as a final online purification step to minimize the aggregation due to instability and shipping issues and helps to obtain good quality SAXS data. In SEC–SAXS data sets, subtraction of the buffer depending on the regions selected for the buffer. One of the major issues that can be resolved from SEC–SAXS is to identify oligomeric species, as a sample is continuously scattered during elution. During the elution, the concentration of protein at each point is not known. The ab initio modeling was carried and various Bead models were obtained and averaged which corresponds to the hollow sphere structure with chi-square 0.93. The mean value of NSD 0.612, Standard deviation of 0.015. Despite various advantages, precise protein concentration is difficult to obtain. Due to this, the mass determination based on the forward scattering is not possible. But, the Porod volume obtained for globular particles can provide an alternative mass that is less precise. Whereas for disordered proteins, mass estimation is completely invalid. In conclusion, online size-exclusion chromatography is an important biochemical purification method that can be coupled with SAXS.

Table 5.1: SAXS Data Collection Parameters and scattering derived parameters of *Ostrinia furnacalis* pheromone binding protein 2

Instrument	The BioCat Beamline 18ID (Argonne National Laboratory)
Wavelength (A °)	1.03
q range ($A^{\circ -1}$)	0.004-0.33
SEC instrument	Superdex75(GE Health Care)
Exposure time	Continuous 1 s data-frame measurements of SEC elution
Sample concentration (before SEC)	20 mg/ml
Sample to Detector Distance	1.5 m
Structural Parameters	
I (0) from guiner anlysis	47
Rg (A°)	16.96
q_{max} ($A^{\circ -1}$)	0.286
q Rg ($A^{\circ -1}$)	0.1796-1.2990
P(r) analysis	
I (0) (A°)	50
R(g) (A°)	16.00
D_{max} (A°)	47
Chi square (total estimate from GNOM)	0.83
Porod Volume ($A^{\circ -3}$)	26114
MW mass estimated (Porod volume) (Da)	15361
MW form SAXSMoW (Da)	15165
MW from BSA as standard (Da)	14650

MW from Sequence (Da)	16109
Primary data reduction	SCATTER
Data processing	ATSAS
Ab initio analysis	DAMMIF
Validation and averaging	DAMAVER
Three-dimensional graphical representation	PyMol

Acknowledgment

We gratefully acknowledge Professor Blaine Mooers, Director of the Laboratory of Biomolecular Structure and Function from the University of Oklahoma Health Sciences Center for helping to collect the SEC-SAXS at Advanced Photon Source on Argonne National Laboratory.

REFERENCES

1. Gerlach, W. & Stern, O. Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld. *Zeitschrift für Physik* **9**, 349-352 (1922).
2. Kellogg, J.M.B., Rabi, I.I., Ramsey, N.F. & Zacharias, J.R. The Magnetic Moments of the Proton and the Deuteron. The Radiofrequency Spectrum of ${\mathrm{H}}_{2}$ in Various Magnetic Fields. *Physical Review* **56**, 728-743 (1939).
3. Bloch, F. Nuclear Induction. *Physical Review* **70**, 460-474 (1946).
4. Purcell, E.M., Torrey, H.C. & Pound, R.V. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review* **69**, 37-38 (1946).
5. Overhauser, A.W. Polarization of Nuclei in Metals. *Physical Review* **92**, 411-415 (1953).
6. Aue, W.P., Bartholdi, E. & Ernst, R.R. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *The Journal of Chemical Physics* **64**, 2229-2246 (1976).
7. Nagayama, K., Wuthrich, K., Bachmann, P. & Ernst, R.R. Two-dimensional J-resolved 1H n.m.r. spectroscopy for studies of biological macromolecules. *Biochem Biophys Res Commun* **78**, 99-105 (1977).
8. Wagner, G. & Wüthrich, K. Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra: Basic pancreatic trypsin inhibitor. *Journal of Molecular Biology* **155**, 347-366 (1982).
9. Williamson, M.P., Havel, T.F. & Wuthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J Mol Biol* **182**, 295-315 (1985).
10. Wüthrich, K. The way to NMR structures of proteins. *Nature Structural Biology* **8**, 923 (2001).
11. Bax, A. Multidimensional nuclear magnetic resonance methods for protein studies. *Current Opinion in Structural Biology* **4**, 738-744 (1994).
12. Ikura, M., Kay, L.E. & Bax, A. A novel approach for sequential assignment of proton, carbon-13, and nitrogen-15 spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29**, 4659-4667 (1990).
13. Clore, G.M. & Gronenborn, A.M. Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252**, 1390 (1991).
14. Cavanagh, J., Fairbrother, W.J., Palmer, A.G., Rance, M. & Skelton, N.J. CHAPTER 1 - CLASSICAL NMR SPECTROSCOPY. in *Protein NMR Spectroscopy (Second Edition)* (eds. Cavanagh, J., Fairbrother, W.J., Palmer, A.G., Rance, M. & Skelton, N.J.) 1-28 (Academic Press, Burlington, 2007).
15. Traficante, D.D. Relaxation. Can T2, be longer than T1? *Concepts in Magnetic Resonance* **3**, 171-177 (1991).
16. Harris Robin, K., Becker Edwin, D., Cabral de Menezes Sonia, M., Goodfellow, R. & Granger, P. NMR nomenclature. Nuclear spin properties and conventions for chemical

shifts(IUPAC Recommendations 2001). in *Pure and Applied Chemistry* Vol. 73 1795 (2001).
17. Prestegard, J.H., Bougault, C.M. & Kishore, A.I. Residual Dipolar Couplings in Structure Determination of Biomolecules. *Chemical Reviews* **104**, 3519-3540 (2004).
18. Rule, G. & Hitchens, K. *Fundamentals of Protein NMR Spectroscopy*, (2005).
19. Mittermaier, A.K. & Kay, L.E. Observing biological dynamics at atomic resolution using NMR. *Trends in Biochemical Sciences* **34**, 601-611 (2009).
20. Bax, A. Two-dimensional nuclear magnetic resonance in liquids. (1981).
21. Bodenhausen, G. & Ruben, D.J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* **69**, 185-189 (1980).
22. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **99**, 201-207 (1992).
23. Kay, L.E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **89**, 496-514 (1990).
24. Clore, G.M., Bax, A., Driscoll, P.C., Wingfield, P.T. & Gronenborn, A.M. Assignment of the side-chain 1H and 13C resonances of interleukin-1 beta using double- and triple-resonance heteronuclear three-dimensional NMR spectroscopy. *Biochemistry* **29**, 8172-84 (1990).
25. Lundström, P. Nuclear Overhauser Effect. in *Encyclopedia of Biophysics* (ed. Roberts, G.C.K.) 1753-1759 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).
26. Quinn, C.M., Wang, M. & Polenova, T. NMR of Macromolecular Assemblies and Machines at 1 GHz and Beyond: New Transformative Opportunities for Molecular Structural Biology. *Methods in molecular biology (Clifton, N.J.)* **1688**, 1-35 (2018).
27. Tikole, S., Jaravine, V., Orekhov, V.Y. & Güntert, P. Effects of NMR Spectral Resolution on Protein Structure Calculation. *PLOS ONE* **8**, e68567 (2013).
28. Ishima, R. Protein-Inhibitor Interaction Studies Using NMR. *Applications of NMR spectroscopy* **1**, 143-181 (2015).
29. Becker, W., Bhattiprolu, K.C., Gubensäk, N. & Zangger, K. Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. *Chemphyschem : a European journal of chemical physics and physical chemistry* **19**, 895-906 (2018).
30. Günther, U., Mittag, T. & Schaffhausen, B. Probing Src homology 2 domain ligand interactions by differential line broadening. *Biochemistry* **41**, 11658-69 (2002).
31. Kleckner, I.R. & Foster, M.P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et biophysica acta* **1814**, 942-968 (2011).
32. Craik, D.J. & Wilce, J.A. Studies of Protein-Ligand Interactions by NMR. in *Protein NMR Techniques* (ed. Reid, D.G.) 195-232 (Humana Press, Totowa, NJ, 1997).
33. Karlson, P. & Luscher, M. Pheromones': a new term for a class of biologically active substances. *Nature* **183**, 55-6 (1959).
34. Elshafie, H. semiochemicals and their potential role in pest management. 3-22 (2017).
35. Regnier, F.E. & Law, J.H. Insect pheromones. *J Lipid Res* **9**, 541-51 (1968).
36. Greenfield, M.D. Moth Sex Pheromones: An Evolutionary Perspective. *The Florida Entomologist* **64**, 4-17 (1981).
37. Bossert, W.H. & Wilson, E.O. The analysis of olfactory communication among animals. *Journal of Theoretical Biology* **5**, 443-469 (1963).
38. Vogt, R.G. & Riddiford, L.M. Pheromone binding and inactivation by moth antennae. *Nature* **293**, 161-163 (1981).

39. Gyorgyi, T.K., Roby-Shemkovitz, A.J. & Lerner, M.R. Characterization and cDNA cloning of the pheromone-binding protein from the tobacco hornworm, Manduca sexta: a tissue-specific developmentally regulated protein. *Proc Natl Acad Sci U S A* **85**, 9851-5 (1988).
40. Vogt, R.G. 12 - The Molecular Basis of Pheromone Reception: Its Influence on Behavior. in *Pheromone Biochemistry* (eds. Prestwich, G.D. & Blomquist, G.J.) 385-431 (Academic Press, 1987).
41. Xiu, W.M. & Dong, S.L. Molecular characterization of two pheromone binding proteins and quantitative analysis of their expression in the beet armyworm, Spodoptera exigua Hubner. *J Chem Ecol* **33**, 947-61 (2007).
42. Tian, Z. & Zhang, Y. Molecular characterization and functional analysis of pheromone binding protein 1 from Cydia pomonella (L.). *Insect Mol Biol* **25**, 769-777 (2016).
43. Shields, V.D.C. Ultrastructure of Insect Sensilla. in *Encyclopedia of Entomology* 2408-2420 (Springer Netherlands, Dordrecht, 2005).
44. Kaissling, K.E. Frontiers in Neuroscience.Pheromone Reception in Insects: The Example of Silk Moths. in *Neurobiology of Chemical Communication* (ed. Mucignat-Caretta, C.) (CRC Press/Taylor & Francis(c) 2014 by Taylor & Francis Group, LLC., Boca Raton (FL), 2014).
45. Sakurai, T., Namiki, S. & Kanzaki, R. Molecular and neural mechanisms of sex pheromone reception and processing in the silkmoth Bombyx mori. *Frontiers in Physiology* **5**(2014).
46. Vogt, R.G., Kohne, A.C., Dubnau, J.T. & Prestwich, G.D. Expression of pheromone binding proteins during antennal development in the gypsy moth Lymantria dispar. *The Journal of Neuroscience* **9**, 3332 (1989).
47. Flower, D.R. The lipocalin protein family: structure and function. *Biochemical Journal* **318**, 1-14 (1996).
48. Pelosi, P., Zhu, J. & Knoll, W. Odorant-Binding Proteins as Sensing Elements for Odour Monitoring. *Sensors (Basel, Switzerland)* **18**, 3248 (2018).
49. Smita, M., Joshua, R.R. & Rabi, K.P. Chemical Communication: A Visit with Insects. *Current Chemical Biology* **2**, 83-96 (2008).
50. Gong, Y. et al. Ligand-interaction kinetics of the pheromone- binding protein from the gypsy moth, L. dispar: insights into the mechanism of binding and release. *Chem Biol* **16**, 162-72 (2009).
51. Kowcun, A., Honson, N. & Plettner, E. Olfaction in the gypsy moth, Lymantria dispar: effect of pH, ionic strength, and reductants on pheromone transport by pheromone-binding proteins. *J Biol Chem* **276**, 44770-6 (2001).
52. Leal, W.S. Odorant Reception in Insects: Roles of Receptors, Binding Proteins, and Degrading Enzymes. *Annual Review of Entomology* **58**, 373-391 (2013).
53. Hekmat-Scafe, D.S., Scafe, C.R., McKinney, A.J. & Tanouye, M.A. Genome-wide analysis of the odorant-binding protein gene family in Drosophila melanogaster. *Genome research* **12**, 1357-1369 (2002).
54. Zhou, J.-J., Huang, W., Zhang, G.-A., Pickett, J.A. & Field, L.M. "Plus-C" odorant-binding protein genes in two Drosophila species and the malaria mosquito Anopheles gambiae. *Gene* **327**, 117-129 (2004).
55. Venthur, H., Mutis, A., Zhou, J.-J. & Quiroz, A. Ligand binding and homology modelling of insect odorant-binding proteins. *Physiological Entomology* **39**, 183-198 (2014).

56. Xu, P.X., Zwiebel, L.J. & Smith, D.P. Identification of a distinct family of genes encoding atypical odorant-binding proteins in the malaria vector mosquito, Anopheles gambiae. *Insect Mol Biol* **12**, 549-60 (2003).
57. Lagarde, A. et al. The Crystal Structure of Odorant Binding Protein 7 from Anopheles gambiae Exhibits an Outstanding Adaptability of Its Binding Site. *Journal of Molecular Biology* **414**, 401-412 (2011).
58. Wojtasek, H. & Leal, W.S. Conformational change in the pheromone-binding protein from Bombyx mori induced by pH and by interaction with membranes. *J Biol Chem* **274**, 30950-6 (1999).
59. Damberger, F. et al. NMR characterization of a pH-dependent equilibrium between two folded solution conformations of the pheromone-binding protein from Bombyx mori. *Protein Sci* **9**, 1038-41 (2000).
60. Leal, W.S. Duality monomer-dimer of the pheromone-binding protein from Bombyx mori. *Biochem Biophys Res Commun* **268**, 521-9 (2000).
61. Sandler, B.H., Nikonova, L., Leal, W.S. & Clardy, J. Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein-bombykol complex. *Chem Biol* **7**, 143-51 (2000).
62. Horst, R. et al. NMR structure reveals intramolecular regulation mechanism for pheromone binding and release. *Proc Natl Acad Sci U S A* **98**, 14374-9 (2001).
63. Lee, D. et al. NMR structure of the unliganded Bombyx mori pheromone-binding protein at physiological pH. *FEBS Lett* **531**, 314-8 (2002).
64. Lautenschlager, C., Leal, W.S. & Clardy, J. Coil-to-helix transition and ligand release of Bombyx mori pheromone-binding protein. *Biochem Biophys Res Commun* **335**, 1044-50 (2005).
65. Katre, U.V., Mazumder, S., Prusti, R.K. & Mohanty, S. Ligand binding turns moth pheromone-binding protein into a pH sensor: effect on the Antheraea polyphemus PBP1 conformation. *J Biol Chem* **284**, 32167-77 (2009).
66. Mazumder, S., Chaudhary, B.P., Dahal, S.R., Al-Danoon, O. & Mohanty, S. Pheromone Perception: Mechanism of the Reversible Coil–Helix Transition in Antheraea polyphemus Pheromone-Binding Protein 1. *Biochemistry* **58**, 4530-4542 (2019).
67. Leal, W.S. et al. Kinetics and molecular properties of pheromone binding and release. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5386 (2005).
68. Regnier, F.E. & Law, J.H. Insect pheromones. *Journal of lipid research* **9**, 541-551 (1968).
69. Lautenschlager, C., Leal, W.S. & Clardy, J. Bombyx mori pheromone-binding protein binding nonpheromone ligands: implications for pheromone recognition. *Structure* **15**, 1148-54 (2007).
70. Ishida, Y. & Leal, W.S. Rapid inactivation of a moth pheromone. *Proc Natl Acad Sci U S A* **102**, 14075-9 (2005).
71. Michel, E. et al. Dynamic conformational equilibria in the physiological function of the Bombyx mori pheromone-binding protein. *J Mol Biol* **408**, 922-31 (2011).
72. Sandler, B.H., Nikonova, L., Leal, W.S. & Clardy, J. Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein–bombykol complex. *Chemistry & Biology* **7**, 143-151 (2000).
73. Lautenschlager, C., Leal, W.S. & Clardy, J. Bombyx mori Pheromone-Binding Protein Binding Nonpheromone Ligands: Implications for Pheromone Recognition. *Structure* **15**, 1148-1154 (2007).

74. Damberger, F.F., Michel, E., Ishida, Y., Leal, W.S. & Wüthrich, K. Pheromone discrimination by a pH-tuned polymorphism of the Bombyx mori pheromone-binding protein. *Proceedings of the National Academy of Sciences* **110**, 18680 (2013).
75. van der Goot, F.G., González-Mañas, J.M., Lakey, J.H. & Pattus, F. A 'molten-globule' membrane-insertion intermediate of the pore-forming domain of colicin A. *Nature* **354**, 408-410 (1991).
76. Katre, U.V., Mazumder, S. & Mohanty, S. Structural insights into the ligand binding and releasing mechanism of Antheraea polyphemus pheromone-binding protein 1: role of the C-terminal tail. *Biochemistry* **52**, 1037-44 (2013).
77. Keil, T.A. Surface coats of pore tubules and olfactory sensory dendrites of a silkmoth revealed by cationic markers. *Tissue Cell* **16**, 705-17 (1984).
78. Du, G., Ng, C.-S. & Prestwich, G.D. Odorant Binding by a Pheromone Binding Protein: Active Site Mapping by Photoaffinity Labeling. *Biochemistry* **33**, 4812-4819 (1994).
79. Prestwich, G.D. Bacterial expression and photoaffinity labeling of a pheromone binding protein. *Protein science : a publication of the Protein Society* **2**, 420-428 (1993).
80. Mohanty, S., Zubkov, S. & Gronenborn, A.M. The Solution NMR Structure of Antheraea polyphemus PBP Provides New Insight into Pheromone Recognition by Pheromone-binding Proteins. *Journal of Molecular Biology* **337**, 443-451 (2004).
81. Damberger, F.F., Ishida, Y., Leal, W.S. & Wüthrich, K. Structural Basis of Ligand Binding and Release in Insect Pheromone-binding Proteins: NMR Structure of Antheraea polyphemus PBP1 at pH 4.5. *Journal of Molecular Biology* **373**, 811-819 (2007).
82. Zubkov, S., Gronenborn, A.M., Byeon, I.J. & Mohanty, S. Structural consequences of the pH-induced conformational switch in A.polyphemus pheromone-binding protein: mechanisms of ligand release. *J Mol Biol* **354**, 1081-90 (2005).
83. Xu, X. et al. NMR structure of navel orangeworm moth pheromone-binding protein (AtraPBP1): implications for pH-sensitive pheromone detection. *Biochemistry* **49**, 1469-1476 (2010).
84. Xu, W., Xu, X., Leal, W.S. & Ames, J.B. Extrusion of the C-terminal helix in navel orangeworm moth pheromone-binding protein (AtraPBP1) controls pheromone binding. *Biochemical and biophysical research communications* **404**, 335-338 (2011).
85. di Luccio, E., Ishida, Y., Leal, W.S. & Wilson, D.K. Crystallographic observation of pH-induced conformational changes in the Amyelois transitella pheromone-binding protein AtraPBP1. *PloS one* **8**, e53840-e53840 (2013).
86. Bierl, B.A., Beroza, M. & Collier, C.W. Potent sex attractant of the gypsy moth: its isolation, identification, and synthesis. *Science* **170**, 87-9 (1970).
87. Vogt, R.G., Köhne, A.C., Dubnau, J.T. & Prestwich, G.D. Expression of pheromone binding proteins during antennal development in the gypsy moth Lymantria dispar. *J Neurosci* **9**, 3332-46 (1989).
88. Plettner, E., Lazar, J., Prestwich, E.G. & Prestwich, G.D. Discrimination of Pheromone Enantiomers by Two Pheromone Binding Proteins from the Gypsy Moth Lymantria dispar. *Biochemistry* **39**, 8953-8962 (2000).
89. Terrado, M., Okon, M., McIntosh, L.P. & Plettner, E. Ligand- and pH-Induced Structural Transition of Gypsy Moth Lymantria dispar Pheromone-Binding Protein 1 (LdisPBP1). *Biochemistry* **59**, 3411-3426 (2020).
90. Lartigue, A. et al. The crystal structure of a cockroach pheromone-binding protein suggests a new ligand binding and release mechanism. *J Biol Chem* **278**, 30213-8 (2003).

91. Lescop, E., Briand, L., Pernollet, J.C. & Guittet, E. Structural basis of the broad specificity of a general odorant-binding protein from honeybee. *Biochemistry* **48**, 2431-41 (2009).
92. Pesenti, M.E. et al. Structural basis of the honey bee PBP pheromone and pH-induced conformational change. *J Mol Biol* **380**, 158-69 (2008).
93. Pesenti, M.E. et al. Queen bee pheromone binding protein pH-induced domain swapping favors pheromone release. *J Mol Biol* **390**, 981-90 (2009).
94. Lartigue, A. et al. Sulfur single-wavelength anomalous diffraction crystal structure of a pheromone-binding protein from the honeybee Apis mellifera L. *J Biol Chem* **279**, 4459-64 (2004).
95. Zhou, J.-J. et al. Characterisation of Bombyx mori Odorant-binding Proteins Reveals that a General Odorant-binding Protein Discriminates Between Sex Pheromone Components. *Journal of Molecular Biology* **389**, 529-545 (2009).
96. Wogulis, M., Morgan, T., Ishida, Y., Leal, W.S. & Wilson, D.K. The crystal structure of an odorant binding protein from Anopheles gambiae: evidence for a common ligand release mechanism. *Biochem Biophys Res Commun* **339**, 157-64 (2006).
97. Leite, N.R. et al. Structure of an Odorant-Binding Protein from the Mosquito Aedes aegypti Suggests a Binding Pocket Covered by a pH-Sensitive "Lid". *PLOS ONE* **4**, e8006 (2009).
98. Mao, Y. et al. Crystal and solution structures of an odorant-binding protein from the southern house mosquito complexed with an oviposition pheromone. *Proc Natl Acad Sci U S A* **107**, 19102-7 (2010).
99. Wang, J., Murphy, E.J., Nix, J.C. & Jones, D.N.M. Aedes aegypti Odorant Binding Protein 22 selectively binds fatty acids through a conformational change in its C-terminal tail. *Scientific Reports* **10**, 3300 (2020).
100. Spinelli, S. et al. Crystal structure of Apis mellifera OBP14, a C-minus odorant-binding protein, and its complexes with odorant molecules. *Insect Biochemistry and Molecular Biology* **42**, 41-50 (2012).
101. Bucci, B.K., Kruse, S.W., Thode, A.B., Alvarado, S.M. & Jones, D.N.M. Effect of n-Alcohols on the Structure and Stability of the Drosophila Odorant Binding Protein LUSH. *Biochemistry* **45**, 1693-1701 (2006).
102. Kruse, S.W., Zhao, R., Smith, D.P. & Jones, D.N. Structure of a specific alcohol-binding site defined by the odorant binding protein LUSH from Drosophila melanogaster. *Nat Struct Biol* **10**, 694-700 (2003).
103. Venthur, H. et al. Structural investigation of selective binding dynamics for the pheromone-binding protein 1 of the grapevine moth, Lobesia botrana. *Archives of Insect Biochemistry and Physiology* **101**, e21557 (2019).
104. Maida, R., Ziegelberger, G. & Kaissling, K.E. Ligand binding to six recombinant pheromone-binding proteins of Antheraea polyphemus and Antheraea pernyi. *J Comp Physiol B* **173**, 565-73 (2003).
105. Campanacci, V. et al. Moth chemosensory protein exhibits drastic conformational changes and cooperativity on ligand binding. *Proc Natl Acad Sci U S A* **100**, 5069-74 (2003).
106. Wogulis, M., Morgan, T., Ishida, Y., Leal, W.S. & Wilson, D.K. The crystal structure of an odorant binding protein from Anopheles gambiae: Evidence for a common ligand release mechanism. *Biochemical and Biophysical Research Communications* **339**, 157-164 (2006).

107. Allen, J.E. & Wanner, K.W. Asian corn borer pheromone binding protein 3, a candidate for evolving specificity to the 12-tetradecenyl acetate sex pheromone. *Insect Biochemistry and Molecular Biology* **41**, 141-149 (2011).
108. Roelofs, W.L. et al. Evolution of moth sex pheromones via ancestral genes. *Proc Natl Acad Sci U S A* **99**, 13621-6 (2002).
109. Klun, J.A. et al. Sex pheromone of the Asian corn borer moth. *Life Sci* **27**, 1603-6 (1980).
110. Sakurai, T. et al. A single sex pheromone receptor determines chemical response specificity of sexual behavior in the silkmoth Bombyx mori. *PLoS Genet* **7**, e1002115 (2011).
111. Leary, G.P. et al. Single mutation to a sex pheromone receptor provides adaptive specificity between closely related moth species. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 14081-14086 (2012).
112. Lassance, J.M. et al. Functional consequences of sequence variation in the pheromone biosynthetic gene pgFAR for Ostrinia moths. *Proc Natl Acad Sci U S A* **110**, 3967-72 (2013).
113. Lassance, J.M. Journey in the Ostrinia world: from pest to model in chemical ecology. *J Chem Ecol* **36**, 1155-69 (2010).
114. Klun, J.A. et al. Sex pheromone of the Asian corn borer moth. *Life Sciences* **27**, 1603-1606 (1980).
115. El-Sayed, A.M. The pherobase: database of insect pheromones and semiochemicals. *See http://www. pherobase. com* (2008).
116. Lassance, J.M. et al. Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. *Nature* **466**, 486-9 (2010).
117. Willett, C.S. & Harrison, R.G. Pheromone binding proteins in the European and Asian corn borers: no protein change associated with pheromone differences. *Insect Biochemistry and Molecular Biology* **29**, 277-284 (1999).
118. Yeh, S.F., Lee, K.C., Chang, K.-T., Yen, F.-C. & Hwang, J.-S. Sex pheromone components from asian corn borer,Ostrinia furnacalis (Hubner) (Lepidoptera: Pyralidae) in Taiwan. *Journal of Chemical Ecology* **15**, 497-505 (1989).
119. Cheng, Z.-Q. et al. *Sex pheromone components isolated from china corn borer,Ostrinia furnacalis guen*[illegible]*e (lepidoptera: Pyralidae), (E)- and (Z)-12-tetradecenyl acetates*, 841-851 (1981).
120. Linn, C.E., Musto, C.J. & Roelofs, W.L. More Rare Males in Ostrinia: Response of Asian Corn Borer Moths to the Sex Pheromone of the European Corn Borer. *Journal of Chemical Ecology* **33**, 199-212 (2007).
121. Ando, T., Saito, O., Arai, K. & Takahashi, N. (Z)- and (E)-12-tetradecenyl acetates: sex pheromone components of oriental corn borer (Lepidoptera: Pyralidae). *Agricultural and biological chemistry* **44**, 2643-2649 (1980).
122. Nafus, D.M. & Schreiner, I.H. Review of the biology and control of the Asian corn borer, Ostrinia furnacalis (Lep: Pyralidae). *Tropical Pest Management* **37**, 41-56 (1991).
123. Ando, T., Saito, O., Arai, K. & Takahashi, N. (Z)- and (E)-12-Tetradecenyl Acetates: Sex Pheromone Components of Oriental Corn Borer (Lepidoptera : Pyralidae). *Agricultural and Biological Chemistry* **44**, 2643-2649 (1980).
124. Kou, R., Ho, H.Y., Yang, H.T., Chow, Y.S. & Wu, H.J. Investigation of sex pheromone components of female asian corn borer,Ostrinia furnacalis (Hübner) (Lepidoptera: Pyralidae) in Taiwan. *Journal of Chemical Ecology* **18**, 833-840 (1992).

125. Willett, C.S. & Harrison, R.G. Pheromone binding proteins in the European and Asian corn borers: no protein change associated with pheromone differences. *Insect Biochem Mol Biol* **29**, 277-84 (1999).
126. Bohnenblust, E.W. et al. Current European corn borer, Ostrinia nubilalis, injury levels in the northeastern United States and the value of Bt field corn. *Pest Manag Sci* **70**, 1711-9 (2014).
127. Hutchison, W.D. et al. Areawide suppression of European corn borer with Bt maize reaps savings to non-Bt maize growers. *Science* **330**, 222-5 (2010).
128. Bell, J.R. et al. Putting the brakes on a cycle: bottom-up effects damp cycle amplitude. *Ecol Lett* **15**, 310-8 (2012).
129. Ostlie, K.R., Hutchison, W.D. & Hellmich, R.L. Bt corn and European corn borer. (1997).
130. Morrison, W.P., Mock, D.E., Stone, J.D. & Whitworth, J. A Bibliography of the Southwestern Corn Borer, Diatraea grandiosella Dyar (Lepidoptera: Pyralidae)1. *Bulletin of the Entomological Society of America* **23**, 185-190 (1977).
131. Calcagno, V. et al. Parallel evolution of behaviour during independent host-shifts following maize introduction into Asia and Europe. *Evolutionary applications* **10**, 881-889 (2017).
132. Patanakamjorn, S., Guthrie, W.D. & Young, W.R. Biology of the tropical corn borer, Ostrinia furnacalis, in relation to host plant resistance research. *Iowa State Journal of Research* (1978).
133. Saito, O. & Oku, T. Population trends of the oriental corn borer, Ostrinia furnacalis (Guenée), in a corn field. *Bulletin, Tohoku National Agricultural Experiment Station, Morioka*, 43-57 (1985).
134. He, K. et al. Evaluation of transgenic Bt corn for resistance to the Asian corn borer (Lepidoptera: Pyralidae). *J Econ Entomol* **96**, 935-40 (2003).
135. Gross, K. & Rosenheim, J.A. Quantifying secondary pest outbreaks in cotton and their monetary cost with causal-inference statistics. *Ecol Appl* **21**, 2770-80 (2011).
136. Allen, J.E. & Wanner, K.W. Asian corn borer pheromone binding protein 3, a candidate for evolving specificity to the 12-tetradecenyl acetate sex pheromone. *Insect Biochem Mol Biol* **41**, 141-9 (2011).
137. Gong, Y. et al. Ligand-Interaction Kinetics of the Pheromone- Binding Protein from the Gypsy Moth, L. dispar: Insights into the Mechanism of Binding and Release. *Chemistry & Biology* **16**, 162-172 (2009).
138. Mohanty, S., Zubkov, S. & Gronenborn, A.M. The solution NMR structure of Antheraea polyphemus PBP provides new insight into pheromone recognition by pheromone-binding proteins. *J Mol Biol* **337**, 443-51 (2004).
139. Zubkov, S., Gronenborn, A.M., Byeon, I.-J.L. & Mohanty, S. Structural Consequences of the pH-induced Conformational Switch in A.polyphemus Pheromone-binding Protein: Mechanisms of Ligand Release. *Journal of Molecular Biology* **354**, 1081-1090 (2005).
140. Pace, C.N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* **4**, 2411-23 (1995).
141. Stepanchenko, N.S., Novikova, G.V. & Moshkov, I.E. Protein quantification. *Russian Journal of Plant Physiology* **58**, 737 (2011).
142. Stoscheck, C.M. Quantitation of protein. *Methods Enzymol* **182**, 50-68 (1990).
143. Bette, S., Breer, H. & Krieger, J. Probing a pheromone binding protein of the silkmoth Antheraea polyphemus by endogenous tryptophan fluorescence. *Insect Biochem Mol Biol* **32**, 241-6 (2002).

144. Sreerama, N. & Woody, R.W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* **287**, 252-60 (2000).
145. Mazumder, S., Dahal, S.R., Chaudhary, B.P. & Mohanty, S. Structure and Function Studies of Asian Corn Borer Ostrinia furnacalis Pheromone Binding Protein2. *Scientific Reports* **8**, 17105 (2018).
146. Fischer, B., Sumner, I. & Goodenough, P. Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in Escherichia coli as inclusion bodies. *Biotechnol Bioeng* **41**, 3-13 (1993).
147. Tyedmers, J., Mogk, A. & Bukau, B. Cellular strategies for controlling protein aggregation. *Nat Rev Mol Cell Biol* **11**, 777-88 (2010).
148. Nardella, J., Terrado, M., Honson, N.S. & Plettner, E. Endogenous fatty acids in olfactory hairs influence pheromone binding protein structure and function in Lymantria dispar. *Archives of Biochemistry and Biophysics* **579**, 73-84 (2015).
149. Wallimann, P., Kennedy, R.J., Miller, J.S., Shalongo, W. & Kemp, D.S. Dual Wavelength Parametric Test of Two-State Models for Circular Dichroism Spectra of Helical Polypeptides: Anomalous Dichroic Properties of Alanine-Rich Peptides. *Journal of the American Chemical Society* **125**, 1203-1220 (2003).
150. Consalvi, V. et al. Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile Thermotoga maritima. *Protein Engineering, Design and Selection* **13**, 501-507 (2000).
151. Rodger, A., Marrington, R., Roper, D. & Windsor, S. Circular dichroism spectroscopy for the study of protein-ligand interactions. *Methods Mol Biol* **305**, 343-64 (2005).
152. Ranjbar, B. & Gill, P. Circular Dichroism Techniques: Biomolecular and Nanostructural Analyses- A Review. *Chemical Biology & Drug Design* **74**, 101-120 (2009).
153. Patel, M.M., Sgourakis, N.G., Garcia, A.E. & Makhatadze, G.I. Experimental Test of the Thermodynamic Model of Protein Cooperativity Using Temperature-Induced Unfolding of a Ubq−UIM Fusion Protein. *Biochemistry* **49**, 8455-8467 (2010).
154. Neira, J.L. et al. Structural characterisation of the natively unfolded enterocin EJ97. *Protein Engineering, Design and Selection* **23**, 507-518 (2010).
155. Zhang, T. et al. Binding affinity of five PBPs to Ostrinia sex pheromones. *BMC Mol Biol* **18**, 4 (2017).
156. Veltri, T. et al. Amide hydrogens reveal a temperature-dependent structural transition that enhances site-II Ca2+-binding affinity in a C-domain mutant of cardiac troponin C. *Scientific Reports* **7**, 691 (2017).
157. Baxter, N.J. & Williamson, M.P. Temperature dependence of 1H chemical shifts in proteins. *J Biomol NMR* **9**, 359-69 (1997).
158. Tycko, R. NMR at low and ultralow temperatures. *Accounts of chemical research* **46**, 1923-1932 (2013).
159. Chen, K., Freedberg, D.I. & Keire, D.A. NMR profiling of biomolecules at natural abundance using 2D 1H-15N and 1H-13C multiplicity-separated (MS) HSQC spectra. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **251**, 65-70 (2015).
160. Kjaergaard, M., Teilum, K. & Poulsen, F.M. Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. *Proceedings of the National Academy of Sciences* **107**, 12535 (2010).
161. Arai, M. & Kuwajima, K. Role of the molten globule state in protein folding. *Adv Protein Chem* **53**, 209-82 (2000).

162. Kim, S., Bracken, C. & Baum, J. Characterization of millisecond time-scale dynamics in the molten globule state of α-lactalbumin by NMR11Edited by P. E. Wright. *Journal of Molecular Biology* **294**, 551-560 (1999).
163. Bhattacharjya, S. et al. pH-induced conformational transitions of a molten-globule-like state of the inhibitory prodomain of furin: implications for zymogen activation. *Protein science : a publication of the Protein Society* **10**, 934-942 (2001).
164. Horst, R. et al. NMR structure reveals intramolecular regulation mechanism for pheromone binding and release. *Proceedings of the National Academy of Sciences* **98**, 14374 (2001).
165. Lautenschlager, C., Leal, W.S. & Clardy, J. Coil-to-helix transition and ligand release of Bombyx mori pheromone-binding protein. *Biochemical and Biophysical Research Communications* **335**, 1044-1050 (2005).
166. di Luccio, E., Ishida, Y., Leal, W.S. & Wilson, D.K. Crystallographic Observation of pH-Induced Conformational Changes in the Amyelois transitella Pheromone-Binding Protein AtraPBP1. *PLOS ONE* **8**, e53840 (2013).
167. Michel, E. et al. Dynamic Conformational Equilibria in the Physiological Function of the Bombyx mori Pheromone-Binding Protein. *Journal of Molecular Biology* **408**, 922-931 (2011).
168. Damberger, F.F., Michel, E., Ishida, Y., Leal, W.S. & Wüthrich, K. Pheromone discrimination by a pH-tuned polymorphism of the Bombyx mori pheromone-binding protein. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 18680-18685 (2013).
169. Leal, W.S., Nikonova, L. & Peng, G. Disulfide structure of the pheromone binding protein from the silkworm moth, Bombyx mori. *FEBS Letters* **464**, 85-90 (1999).
170. Gong, Y., Tang, H., Bohne, C. & Plettner, E. Binding conformation and kinetics of two pheromone-binding proteins from the Gypsy moth Lymantria dispar with biological and nonbiological ligands. *Biochemistry* **49**, 793-801 (2010).
171. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213-23 (2009).
172. Marsh, J.A., Singh, V.K., Jia, Z. & Forman-Kay, J.D. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein science : a publication of the Protein Society* **15**, 2795-2804 (2006).
173. Mazumder, S., Dahal, S.R., Chaudhary, B.P. & Mohanty, S. Structure and Function Studies of Asian Corn Borer Ostrinia furnacalis Pheromone Binding Protein2. *Sci Rep* **8**, 17105 (2018).
174. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93 (1995).
175. Frueh, D.P. Practical aspects of NMR signal assignment in larger and challenging proteins. *Progress in nuclear magnetic resonance spectroscopy* **78**, 47-75 (2014).
176. Redfield, C. Assignment of Protein NMR Spectra Using Heteronuclear NMR—A Tutorial. in *Protein NMR: Modern Techniques and Biomedical Applications* (ed. Berliner, L.) 1-42 (Springer US, Boston, MA, 2015).
177. Redfield, C. Assignment of Protein NMR Spectra Using Heteronuclear NMR—A Tutorial. 1-42 (2015).

178. Wang, J. et al. Multidimensional NMR studies of an exchangeable apolipoprotein and its interactions with lipids. in *Techniques in Protein Chemistry*, Vol. 8 (ed. Marshak, D.R.) 427-438 (Academic Press, 1997).
179. Schubert, M., Labudde, D., Oschkinat, H. & Schmieder, P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. *J Biomol NMR* **24**, 149-54 (2002).
180. Taler-Verčič, A. et al. Proline residues as switches in conformational changes leading to amyloid fibril formation. *International journal of molecular sciences* **18**, 549 (2017).
181. Ünal, C.M. & Steinert, M. Microbial Peptidyl-Prolyl cis/trans Isomerases (PPIases): Virulence Factors and Potential Alternative Drug Targets. *Microbiology and Molecular Biology Reviews* **78**, 544 (2014).
182. Paramasivam, S., Gronenborn, A.M. & Polenova, T. Backbone amide (15)N chemical shift tensors report on hydrogen bonding interactions in proteins: A magic angle spinning NMR study. *Solid state nuclear magnetic resonance* **92**, 1-6 (2018).
183. Sharma, D. & Rajarathnam, K. 13C NMR chemical shifts can predict disulfide bond formation. *J Biomol NMR* **18**, 165-71 (2000).
184. Bax, A., Clore, G.M. & Gronenborn, A.M. 1H・1H correlation via isotropic mixing of 13C magnetization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins. *Journal of Magnetic Resonance (1969)* **88**, 425-431 (1990).
185. Lin, Y. & Wagner, G. Efficient side-chain and backbone assignment in large proteins: Application to tGCN5. *Journal of Biomolecular NMR* **15**, 227-239 (1999).
186. Gardner, K.H. & Kay, L.E. THE USE OF 2H, 13C, 15N MULTIDIMENSIONAL NMR GTO STUDY THE STRUCTURE AND DYNAMICS OF PROTEINS. *Annual Review of Biophysics and Biomolecular Structure* **27**, 357-406 (1998).
187. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of biomolecular NMR* **44**, 213-223 (2009).
188. Lee, D. et al. NMR structure of the unliganded Bombyx mori pheromone-binding protein at physiological pH. *FEBS Letters* **531**, 314-318 (2002).
189. Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. & Sykes, B.D. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of Biomolecular NMR* **5**, 67-81 (1995).
190. Hafsa, N.E., Arndt, D. & Wishart, D.S. CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. *Nucleic acids research* **43**, W370-W377 (2015).
191. Mok, K.H., Nagashima, T., Day, I.J., Hore, P.J. & Dobson, C.M. Multiple subsets of side-chain packing in partially folded states of α-lactalbumins. *Proceedings of the National Academy of Sciences* **102**, 8899-8904 (2005).
192. Redfield, C. Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. *Methods* **34**, 121-132 (2004).
193. Karplus, M. Contact electron-spin coupling of nuclear magnetic moments. *The Journal of chemical physics* **30**, 11-15 (1959).
194. Zalucki, M.P. et al. Estimating the Economic Cost of One of the World's Major Insect Pests, Plutella xylostella (Lepidoptera: Plutellidae): Just How Long Is a Piece of String? *Journal of Economic Entomology* **105**, 1115-1129 (2012).

195. Zhang, T., He, K. & Wang, Z. Transcriptome Comparison Analysis of Ostrinia furnacalis in Four Developmental Stages. *Scientific Reports* **6**, 35008 (2016).
196. Damberger, F.F., Ishida, Y., Leal, W.S. & Wuthrich, K. Structural basis of ligand binding and release in insect pheromone-binding proteins: NMR structure of Antheraea polyphemus PBP1 at pH 4.5. *J Mol Biol* **373**, 811-9 (2007).
197. Gräter, F., de Groot, B.L., Jiang, H. & Grubmüller, H. Ligand-Release Pathways in the Pheromone-Binding Protein of Bombyx mori. *Structure* **14**, 1567-1576 (2006).
198. Xu, W. & Leal, W.S. Molecular switches for pheromone release from a moth pheromone-binding protein. *Biochem Biophys Res Commun* **372**, 559-64 (2008).
199. Xu, X. et al. 1H, 15N, and 13C Chemical shift assignments of the navel orange worm pheromone-binding protein-1 (Atra-PBP1). *Biomol NMR Assign* **2**, 105-6 (2008).
200. Ziegelberger, G. The multiple role of the pheromone-binding protein in olfactory transduction. *Ciba Found Symp* **200**, 267-75; discussion 275-80 (1996).
201. Honson, N., Johnson, M.A., Oliver, J.E., Prestwich, G.D. & Plettner, E. Structure–Activity Studies with Pheromone-binding Proteins of the Gypsy Moth, Lymantria dispar. *Chemical Senses* **28**, 479-489 (2003).
202. Yu, Y. et al. Structural and functional difference of pheromone binding proteins in discriminating chemicals in the gypsy moth, Lymantria dispar. *International journal of biological sciences* **8**, 979-991 (2012).
203. Lassance, J.-M. Journey in the Ostrinia World: From Pest to Model in Chemical Ecology. *Journal of Chemical Ecology* **36**, 1155-1169 (2010).
204. Roelofs, W. & Bjostad, L. Biosynthesis of lepidopteran pheromones. *Bioorganic Chemistry* **12**, 279-298 (1984).
205. Roelofs, W.L. & Rooney, A.P. Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera. *Proceedings of the National Academy of Sciences* **100**, 9179 (2003).
206. Roelofs, W.L. et al. Evolution of moth sex pheromones via ancestral genes. *Proceedings of the National Academy of Sciences* **99**, 13621 (2002).
207. Dahal, S.R., Lewellen, J.L., Chaudhary, B.P. & Mohanty, S. (1)H, (13)C, and (15)N resonance assignment and secondary structure of the pheromone-binding protein2 from the agricultural pest Ostrinia furnacalis (OfurPBP2). *Biomol NMR Assign* (2020).
208. Lee, W., Tonelli, M. & Markley, J.L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics (Oxford, England)* **31**, 1325-1327 (2015).
209. Güntert, P. Automated NMR protein structure calculation. *Progress in Nuclear Magnetic Resonance Spectroscopy* **43**, 105-125 (2003).
210. Scaloni, A., Monti, M., Angeli, S. & Pelosi, P. Structural Analysis and Disulfide-Bridge Pairing of Two Odorant-Binding Proteins from Bombyx mori. *Biochemical and Biophysical Research Communications* **266**, 386-391 (1999).
211. Krieger, E. et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **77 Suppl 9**, 114-22 (2009).
212. Essmann, U. et al. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577-8593 (1995).
213. Williams, C.J. et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein science : a publication of the Protein Society* **27**, 293-315 (2018).

214. Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R. & Thornton, J.M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**, 477-86 (1996).
215. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Research* **46**, W363-W367 (2018).
216. Morris, G.M. et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry* **30**, 2785-2791 (2009).
217. Trott, O. & Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**, 455-461 (2010).
218. Huey, R., Morris, G.M., Olson, A.J. & Goodsell, D.S. A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* **28**, 1145-52 (2007).
219. Rizvi, S.M.D., Shakil, S. & Haneef, M. A simple click by click protocol to perform docking: AutoDock 4.2 made easy for non-bioinformaticians. *EXCLI journal* **12**, 831-857 (2013).
220. Fuhrmann, J., Rurainski, A., Lenhof, H.P. & Neumann, D. A new Lamarckian genetic algorithm for flexible ligand-receptor docking. *J Comput Chem* **31**, 1911-8 (2010).
221. Cosconati, S. et al. Virtual screening with AutoDock: theory and practice. *Expert Opinion on Drug Discovery* **5**, 597-607 (2010).
222. Berendsen, H.J.C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**, 43-56 (1995).
223. Malde, A.K. et al. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *Journal of Chemical Theory and Computation* **7**, 4026-4037 (2011).
224. Schmid, N. et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal* **40**, 843 (2011).
225. Hess, B., Bekker, H., Berendsen, H.J.C. & Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472 (1997).
226. Rout, S. & Mahapatra, R.K. In silico screening of novel inhibitors of M17 Leucine Amino Peptidase (LAP) of Plasmodium vivax as therapeutic candidate. *Biomedicine & Pharmacotherapy* **82**, 192-201 (2016).
227. Lee, W., Tonelli, M. & Markley, J.L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325-1327 (2015).
228. Wüthrich, K. *NMR of proteins and nucleic acids*, (1986).
229. Sharma, D. & Rajarathnam, K. 13C NMR chemical shifts can predict disulfide bond formation. *Journal of Biomolecular NMR* **18**, 165-171 (2000).
230. Walewska, A. et al. NMR-based mapping of disulfide bridges in cysteine-rich peptides: application to the μ-conotoxin SxIIIA. *Journal of the American Chemical Society* **130**, 14280-14286 (2008).
231. Takeda, M., Terauchi, T. & Kainosho, M. Conformational analysis by quantitative NOE measurements of the β-proton pairs across individual disulfide bonds in proteins. *Journal of Biomolecular NMR* **52**, 127-139 (2012).
232. Klaus, W., Broger, C., Gerber, P. & Senn, H. Determination of the Disulphide Bonding Pattern in Proteins by Local and Global Analysis of Nuclear Magnetic Resonance Data: Application to Flavoridin. *Journal of Molecular Biology* **232**, 897-906 (1993).
233. Wang, J., Gagné, S.M., Sykes, B.D. & Ryan, R.O. Insight into Lipid Surface Recognition and Reversible Conformational Adaptations of an Exchangeable Apolipoprotein by

Multidimensional Heteronuclear NMR Techniques*. *Journal of Biological Chemistry* **272**, 17912-17920 (1997).
234. Nanga, R.P.R., Brender, J.R., Vivekanandan, S. & Ramamoorthy, A. Structure and membrane orientation of IAPP in its natively amidated form at physiological pH in a membrane environment. *Biochimica et biophysica acta* **1808**, 2337-2342 (2011).
235. Lücke, C. et al. Solution structure of a recombinant mouse major urinary protein. *European Journal of Biochemistry* **266**, 1210-1218 (1999).
236. Krieger, E. et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins: Structure, Function, and Bioinformatics* **77**, 114-122 (2009).
237. Canutescu, A.A., Shelenkov, A.A. & Dunbrack, R.L., Jr. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci* **12**, 2001-14 (2003).
238. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**, 283-291 (1993).
239. Ptitsyn, O.B. Molten globule and protein folding. *Adv Protein Chem* **47**, 83-229 (1995).
240. Birktoft, J. & Banaszak, L. The presence of a histidine-aspartic acid pair in the active site of 2-hydroxyacid dehydrogenases. X-ray refinement of cytoplasmic malate dehydrogenase. *The Journal of biological chemistry* **258**, 472-82 (1983).
241. Liao, S.-M., Du, Q.-S., Meng, J.-Z., Pang, Z.-W. & Huang, R.-B. The multiple roles of histidine in protein interactions. *Chemistry Central journal* **7**, 44-44 (2013).
242. Kumar, K. et al. Cation–π interactions in protein–ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chemical Science* **9**, 2655-2665 (2018).
243. Hou, Q., Bourgeas, R., Pucci, F. & Rooman, M. Computational analysis of the amino acid interactions that promote or decrease protein solubility. *Scientific Reports* **8**, 14661 (2018).
244. Ando, T., Saito, O., Arai, K. & Takahashi, N. (Z)- and (E)-12-Tetradecenyl Acetates: Sex Pheromone Components of Oriental Corn Borer (Lepidoptera: Pyralidae). *Agricultural and Biological Chemistry* **44**, 2643-2649 (1980).
245. Anderson, A.C. The process of structure-based drug design. *Chem Biol* **10**, 787-97 (2003).
246. Mohanty, S., Zubkov, S. & Campos-Olivas, R. 1H, 13C and 15N backbone assignments of the pheromone binding protein from the silk moth Antheraea polyphemus (ApolPBP). *J Biomol NMR* **27**, 393-4 (2003).
247. Zhou, J.-J. et al. Revisiting the odorant-binding protein LUSH of Drosophila melanogaster: evidence for odour recognition and discrimination. *FEBS Letters* **558**, 23-26 (2004).
248. Liu, Y., Gu, S., Zhang, Y., Guo, Y. & Wang, G. Candidate Olfaction Genes Identified within the Helicoverpa armigera Antennal Transcriptome. *PLOS ONE* **7**, e48260 (2012).
249. Ozaki, K., Utoguchi, A., Yamada, A. & Yoshikawa, H. Identification and genomic structure of chemosensory proteins (CSP) and odorant binding proteins (OBP) genes expressed in foreleg tarsi of the swallowtail butterfly Papilio xuthus. *Insect Biochem Mol Biol* **38**, 969-76 (2008).
250. Malini, P., Ramasamy, S., Schafleitner, R. & Muthukalingan, K. Pheromone-binding proteins based phylogenetics and phylogeography of Maruca spp. from Asia, Africa, Oceania, and South America. *Ecol Evol* **9**, 9239-9272 (2019).
251. Xiao-Jian Jia, H.-X.W., Zeng-Guang Yan, Min-Zhao Zhang, Chun-Hua Wei, Xiao-Chun Qin, Wei-Rong Ji, Patrizia Falabella & Yan-Li Du. cDNA cloning, expression profiling and binding affinity assay of the pheromone binding protein Cpun-PBP1 in the yellow peach

moth, Conogethes punctiferalis (Lepidoptera: Crambidae). Acta Entomol. Sin. 58, 1167–1176 (2015). (In Chinese).
252. Yasukochi, Y. et al. Conservation and lineage-specific rearrangements in the GOBP/PBP gene complex of distantly related ditrysian Lepidoptera. *PLoS One* **13**, e0192762 (2018).
253. Liao, S.M., Du, Q.S., Meng, J.Z., Pang, Z.W. & Huang, R.B. The multiple roles of histidine in protein interactions. *Chem Cent J* **7**, 44 (2013).
254. Blanchet, C.E. & Svergun, D.I. Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution. *Annual Review of Physical Chemistry* **64**, 37-54 (2013).
255. Klumpler, T. Biological Small-Angle X-Ray Scattering (SAXS). in *Plant Structural Biology: Hormonal Regulations* (eds. Hejátko, J. & Hakoshima, T.) 277-293 (Springer International Publishing, Cham, 2018).
256. Mathew, E., Mirza, A. & Menhart, N. Liquid-chromatography-coupled SAXS for accurate sizing of aggregating proteins. *J Synchrotron Radiat* **11**, 314-8 (2004).
257. Franke, D. et al. *ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions*, (2017).
258. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* **50**, 1212-1225 (2017).
259. Nielsen, S. et al. *BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis*, 959-964 (2009).
260. J. B. Hopkins, R.E.G.a.S.S. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Cryst. Cited by 31* **50**, 1545-1553 (2017).
261. Forster, S., Apostol, L. & Bras, W. Scatter: software for the analysis of nano- and mesoscale small-angle scattering. *Journal of Applied Crystallography* **43**, 639-646 (2010).
262. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. & Svergun, D.I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**, 1277-1282 (2003).
263. Butt, L.E., Holland, R.A., Khunti, N.S., Quinn, D.L. & Pickford, A.R. Using Small Angle X-Ray Scattering (SAXS) to Characterize the Solution Conformation and Flexibility of Matrix Metalloproteinases (MMPs). in *Matrix Metalloproteases: Methods and Protocols* (ed. Galea, C.A.) 87-108 (Springer New York, New York, NY, 2017).
264. Kaissling, K.E. Olfactory perireceptor and receptor events in moths: a kinetic model revised. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **195**, 895-922 (2009).
265. Mertens, H.D.T. & Svergun, D.I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of Structural Biology* **172**, 128-141 (2010).
266. Kikhney, A.G. & Svergun, D.I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters* **589**, 2570-2577 (2015).
267. Rambo, R.P. & Tainer, J.A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559-571 (2011).
268. Svergun, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of applied crystallography* **25**, 495-503 (1992).
269. Mertens, H.D. & Svergun, D.I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J Struct Biol* **172**, 128-41 (2010).

270. Hansen, S. Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering. *Journal of Applied Crystallography - J APPL CRYST* **33**, 1415-1421 (2000).
271. Jacques, D.A. & Trewhella, J. Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls. *Protein Science* **19**, 642-657 (2010).
272. Putnam, C.D., Hammel, M., Hura, G.L. & Tainer, J.A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **40**, 191-285 (2007).
273. Svergun, D.I. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophysical Journal* **76**, 2879-2886 (1999).
274. Franke, D. & Svergun, D.I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of applied crystallography* **42**, 342-346 (2009).
275. Volkov, V.V. & Svergun, D.I. Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **36**, 860-864 (2003).
276. Svergun, D., Barberato, C. & Koch, M.H.J. CRYSOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* **28**, 768-773 (1995).
277. Korasick, D.A. & Tanner, J.J. Determination of protein oligomeric structure from small-angle X-ray scattering. *Protein science : a publication of the Protein Society* **27**, 814-824 (2018).
278. Grant, T.D. et al. The accurate assessment of small-angle X-ray scattering data. *Acta crystallographica. Section D, Biological crystallography* **71**, 45-56 (2015).
279. Hutin, S., Brennich, M., Maillot, B. & Round, A. Online ion-exchange chromatography for small-angle X-ray scattering. *Acta crystallographica. Section D, Structural biology* **72**, 1090-1099 (2016).
280. Piiadov, V., Ares de Araújo, E., Oliveira Neto, M., Craievich, A.F. & Polikarpov, I. SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. *Protein Sci* **28**, 454-463 (2019).
281. Martino, L. et al. The Biophysical Characterisation and SAXS Analysis of Human NLRP1 Uncover a New Level of Complexity of NLR Proteins. *PLOS ONE* **11**, e0164662 (2016).
282. Franke, D. & Svergun, D.I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of applied crystallography* **42**, 342-346 (2009).

APPENDICES

Appendix Table A-1 Backbone chemical shift assignments of the OfurPBP2

1	GLN	HA	H	3.78	0	2	GLN	HA
2	GLN	HB2	H	2.193	0	2	GLN	HB2
3	GLN	HB3	H	2.193	0	2	GLN	HB3
4	GLN	C	C	178.213	0.013	2	GLN	C
5	GLN	CA	C	58.109	0.084	2	GLN	CA
6	GLN	CB	C	28.641	0.109	2	GLN	CB
7	GLN	CG	C	33.819	0.044	2	GLN	CG
8	ALA	H	H	7.921	0.194	3	ALA	H
9	ALA	HA	H	3.851	0.003	3	ALA	HA
10	ALA	HB1	H	1.262	0.007	3	ALA	HB1
11	ALA	HB2	H	1.262	0.007	3	ALA	HB2
12	ALA	HB3	H	1.262	0.007	3	ALA	HB3
13	ALA	C	C	179.845	0.089	3	ALA	C
14	ALA	CA	C	55.329	0.049	3	ALA	CA
15	ALA	CB	C	18.014	0.359	3	ALA	CB
16	ALA	N	N	122.319	0.045	3	ALA	N
17	VAL	H	H	7.16	0.01	4	VAL	H
18	VAL	HA	H	3.709	0.003	4	VAL	HA
19	VAL	HB	H	1.571	0.003	4	VAL	HB
20	VAL	HG11	H	0.755	0.006	4	VAL	HG11
21	VAL	HG12	H	0.755	0.006	4	VAL	HG12
22	VAL	HG13	H	0.755	0.006	4	VAL	HG13
23	VAL	HG21	H	0.679	0.007	4	VAL	HG21
24	VAL	HG22	H	0.679	0.007	4	VAL	HG22
25	VAL	HG23	H	0.679	0.007	4	VAL	HG23
26	VAL	C	C	177.401	0	4	VAL	C
27	VAL	CA	C	65.418	0.192	4	VAL	CA
28	VAL	CB	C	32.031	0.077	4	VAL	CB
29	VAL	CG1	C	21.195	0.23	4	VAL	CG1
30	VAL	CG2	C	20.818	0	4	VAL	CG2
31	VAL	N	N	118.922	0.055	4	VAL	N
32	MET	H	H	7.714	0.006	5	MET	H
33	MET	HA	H	4.374	0.011	5	MET	HA
34	MET	HB2	H	2.046	0.003	5	MET	HB2
35	MET	HB3	H	1.79	0.168	5	MET	HB3
36	MET	HG2	H	2.391	0	5	MET	HG2
37	MET	HG3	H	2.213	0.002	5	MET	HG3
38	MET	C	C	180.745	0.001	5	MET	C
39	MET	CA	C	56.371	0.244	5	MET	CA

40	**MET**	**CB**	**C**	**29.92**	**0.052**	**5**	**MET**	**CB**
41	**MET**	**CG**	**C**	**31.733**	**0**	**5**	**MET**	**CG**
42	**MET**	**N**	**N**	**117.037**	**0.067**	**5**	**MET**	**N**
43	**LYS**	**H**	**H**	**8.612**	**0.012**	**6**	**LYS**	**H**
44	**LYS**	**HA**	**H**	**4.622**	**0**	**6**	**LYS**	**HA**
45	**LYS**	**C**	**C**	**179.723**	**0**	**6**	**LYS**	**C**
46	**LYS**	**CA**	**C**	**60.73**	**0.069**	**6**	**LYS**	**CA**
47	**LYS**	**CB**	**C**	**30.961**	**0.026**	**6**	**LYS**	**CB**
48	**LYS**	**CG**	**C**	**24.094**	**0**	**6**	**LYS**	**CG**
49	**LYS**	**CD**	**C**	**28.428**	**0**	**6**	**LYS**	**CD**
50	**LYS**	**CE**	**C**	**40.95**	**0**	**6**	**LYS**	**CE**
51	**LYS**	**N**	**N**	**124.724**	**0.049**	**6**	**LYS**	**N**
52	**ASP**	**H**	**H**	**7.751**	**0.01**	**7**	**ASP**	**H**
53	**ASP**	**HA**	**H**	**4.39**	**0.006**	**7**	**ASP**	**HA**
54	**ASP**	**HB2**	**H**	**2.749**	**0.008**	**7**	**ASP**	**HB2**
55	**ASP**	**HB3**	**H**	**2.699**	**0.002**	**7**	**ASP**	**HB3**
56	**ASP**	**C**	**C**	**180.001**	**0.011**	**7**	**ASP**	**C**
57	**ASP**	**CA**	**C**	**58.085**	**0.122**	**7**	**ASP**	**CA**
58	**ASP**	**CB**	**C**	**40.008**	**0.113**	**7**	**ASP**	**CB**
59	**ASP**	**N**	**N**	**121.444**	**0.075**	**7**	**ASP**	**N**
60	**MET**	**H**	**H**	**8.875**	**0.134**	**8**	**MET**	**H**
61	**MET**	**HA**	**H**	**3.193**	**0.004**	**8**	**MET**	**HA**
62	**MET**	**HB2**	**H**	**2.024**	**0.084**	**8**	**MET**	**HB2**
63	**MET**	**HB3**	**H**	**1.726**	**0.192**	**8**	**MET**	**HB3**
64	**MET**	**HG2**	**H**	**2.565**	**0.001**	**8**	**MET**	**HG2**
65	**MET**	**HG3**	**H**	**2.075**	**0.001**	**8**	**MET**	**HG3**
66	**MET**	**C**	**C**	**178.145**	**0.039**	**8**	**MET**	**C**
67	**MET**	**CA**	**C**	**60.314**	**0.148**	**8**	**MET**	**CA**
68	**MET**	**CB**	**C**	**34.299**	**0.094**	**8**	**MET**	**CB**
69	**MET**	**CG**	**C**	**32.311**	**0**	**8**	**MET**	**CG**
70	**MET**	**N**	**N**	**119.403**	**0.066**	**8**	**MET**	**N**
71	**THR**	**H**	**H**	**8.001**	**0.008**	**9**	**THR**	**H**
72	**THR**	**HA**	**H**	**4.565**	**0.002**	**9**	**THR**	**HA**
73	**THR**	**HB**	**H**	**3.845**	**0.007**	**9**	**THR**	**HB**
74	**THR**	**HG1**	**H**	**5.265**	**0**	**9**	**THR**	**HG1**
75	**THR**	**HG21**	**H**	**1.436**	**0.004**	**9**	**THR**	**HG21**
76	**THR**	**HG22**	**H**	**1.436**	**0.004**	**9**	**THR**	**HG22**
77	**THR**	**HG23**	**H**	**1.436**	**0.004**	**9**	**THR**	**HG23**
78	**THR**	**C**	**C**	**174.953**	**0**	**9**	**THR**	**C**
79	**THR**	**CA**	**C**	**68.471**	**0.088**	**9**	**THR**	**CA**
80	**THR**	**CB**	**C**	**68.35**	**0.083**	**9**	**THR**	**CB**
81	**THR**	**CG2**	**C**	**21.133**	**0**	**9**	**THR**	**CG2**
82	**THR**	**N**	**N**	**115.82**	**0.051**	**9**	**THR**	**N**
83	**LYS**	**H**	**H**	**8.12**	**0.01**	**10**	**LYS**	**H**
84	**LYS**	**HA**	**H**	**3.84**	**0.016**	**10**	**LYS**	**HA**
85	**LYS**	**HB2**	**H**	**2.136**	**0.007**	**10**	**LYS**	**HB2**
86	**LYS**	**HB3**	**H**	**2.136**	**0.007**	**10**	**LYS**	**HB3**
87	**LYS**	**HG2**	**H**	**1.921**	**0.004**	**10**	**LYS**	**HG2**
88	**LYS**	**HG3**	**H**	**1.921**	**0.004**	**10**	**LYS**	**HG3**
89	**LYS**	**HE2**	**H**	**2.518**	**0.001**	**10**	**LYS**	**HE2**
90	**LYS**	**HE3**	**H**	**2.518**	**0.001**	**10**	**LYS**	**HE3**
91	**LYS**	**C**	**C**	**175.879**	**0.308**	**10**	**LYS**	**C**
92	**LYS**	**CA**	**C**	**60.471**	**0.06**	**10**	**LYS**	**CA**
93	**LYS**	**CB**	**C**	**31.642**	**0.959**	**10**	**LYS**	**CB**
94	**LYS**	**CG**	**C**	**25.835**	**0**	**10**	**LYS**	**CG**
95	**LYS**	**CD**	**C**	**29.649**	**0**	**10**	**LYS**	**CD**
96	**LYS**	**CE**	**C**	**41.621**	**0**	**10**	**LYS**	**CE**
97	**LYS**	**N**	**N**	**119.127**	**0.125**	**10**	**LYS**	**N**
98	**ASN**	**H**	**H**	**7.244**	**0.016**	**11**	**ASN**	**H**
99	**ASN**	**HA**	**H**	**4.571**	**0.004**	**11**	**ASN**	**HA**

100	**ASN**	**HB2**	**H**	**2.533**	**0.418**	**11**	**ASN**	**HB2**
101	**ASN**	**HB3**	**H**	**2.395**	**0.523**	**11**	**ASN**	**HB3**
102	**ASN**	**C**	**C**	**176.394**	**0.344**	**11**	**ASN**	**C**
103	**ASN**	**CA**	**C**	**56.14**	**0.038**	**11**	**ASN**	**CA**
104	**ASN**	**CB**	**C**	**39.803**	**0.035**	**11**	**ASN**	**CB**
105	**ASN**	**N**	**N**	**114.423**	**0.131**	**11**	**ASN**	**N**
106	**PHE**	**H**	**H**	**8.752**	**0.008**	**12**	**PHE**	**H**
107	**PHE**	**HA**	**H**	**4.16**	**0.007**	**12**	**PHE**	**HA**
108	**PHE**	**HB2**	**H**	**3.41**	**0**	**12**	**PHE**	**HB2**
109	**PHE**	**HB3**	**H**	**3.197**	**0**	**12**	**PHE**	**HB3**
110	**PHE**	**CA**	**C**	**59.986**	**1.453**	**12**	**PHE**	**CA**
111	**PHE**	**CB**	**C**	**40.921**	**0.043**	**12**	**PHE**	**CB**
112	**PHE**	**N**	**N**	**123.307**	**0.104**	**12**	**PHE**	**N**
113	**ILE**	**H**	**H**	**7.662**	**0.013**	**13**	**ILE**	**H**
114	**ILE**	**HA**	**H**	**4.136**	**0.005**	**13**	**ILE**	**HA**
115	**ILE**	**HB**	**H**	**2.027**	**0.006**	**13**	**ILE**	**HB**
116	**ILE**	**HG12**	**H**	**1.559**	**0.009**	**13**	**ILE**	**HG12**
117	**ILE**	**HG13**	**H**	**1.559**	**0.009**	**13**	**ILE**	**HG13**
118	**ILE**	**HG21**	**H**	**0.996**	**0.008**	**13**	**ILE**	**HG21**
119	**ILE**	**HG22**	**H**	**0.996**	**0.008**	**13**	**ILE**	**HG22**
120	**ILE**	**HG23**	**H**	**0.996**	**0.008**	**13**	**ILE**	**HG23**
121	**ILE**	**HD11**	**H**	**0.812**	**0.004**	**13**	**ILE**	**HD11**
122	**ILE**	**HD12**	**H**	**0.812**	**0.004**	**13**	**ILE**	**HD12**
123	**ILE**	**HD13**	**H**	**0.812**	**0.004**	**13**	**ILE**	**HD13**
124	**ILE**	**CA**	**C**	**61.988**	**0.022**	**13**	**ILE**	**CA**
125	**ILE**	**CB**	**C**	**39.067**	**0.058**	**13**	**ILE**	**CB**
126	**ILE**	**CG1**	**C**	**28.008**	**0**	**13**	**ILE**	**CG1**
127	**ILE**	**CG2**	**C**	**18.115**	**0**	**13**	**ILE**	**CG2**
128	**ILE**	**CD1**	**C**	**14.144**	**0.055**	**13**	**ILE**	**CD1**
129	**ILE**	**N**	**N**	**109.389**	**1.7**	**13**	**ILE**	**N**
130	**LYS**	**H**	**H**	**7.249**	**0.01**	**14**	**LYS**	**H**
131	**LYS**	**HA**	**H**	**4.865**	**0**	**14**	**LYS**	**HA**
132	**LYS**	**HB2**	**H**	**1.601**	**0**	**14**	**LYS**	**HB2**
133	**LYS**	**HB3**	**H**	**1.601**	**0**	**14**	**LYS**	**HB3**
134	**LYS**	**HG2**	**H**	**1.248**	**0**	**14**	**LYS**	**HG2**
135	**LYS**	**HG3**	**H**	**1.248**	**0**	**14**	**LYS**	**HG3**
136	**LYS**	**HE2**	**H**	**2.695**	**0**	**14**	**LYS**	**HE2**
137	**LYS**	**HE3**	**H**	**2.695**	**0**	**14**	**LYS**	**HE3**
138	**LYS**	**CA**	**C**	**59.861**	**0.009**	**14**	**LYS**	**CA**
139	**LYS**	**CB**	**C**	**32.312**	**0.017**	**14**	**LYS**	**CB**
140	**LYS**	**CG**	**C**	**25.172**	**0**	**14**	**LYS**	**CG**
141	**LYS**	**CD**	**C**	**29.314**	**0**	**14**	**LYS**	**CD**
142	**LYS**	**CE**	**C**	**42.125**	**0**	**14**	**LYS**	**CE**
143	**LYS**	**N**	**N**	**121.863**	**0.048**	**14**	**LYS**	**N**
144	**ALA**	**H**	**H**	**7.105**	**0.008**	**15**	**ALA**	**H**
145	**ALA**	**HA**	**H**	**4.578**	**0.011**	**15**	**ALA**	**HA**
146	**ALA**	**HB1**	**H**	**1.273**	**0.005**	**15**	**ALA**	**HB1**
147	**ALA**	**HB2**	**H**	**1.273**	**0.005**	**15**	**ALA**	**HB2**
148	**ALA**	**HB3**	**H**	**1.273**	**0.005**	**15**	**ALA**	**HB3**
149	**ALA**	**C**	**C**	**176.736**	**0.019**	**15**	**ALA**	**C**
150	**ALA**	**CA**	**C**	**51.889**	**0.046**	**15**	**ALA**	**CA**
151	**ALA**	**CB**	**C**	**17.517**	**0.116**	**15**	**ALA**	**CB**
152	**ALA**	**N**	**N**	**118.137**	**0.074**	**15**	**ALA**	**N**
153	**TYR**	**H**	**H**	**8.153**	**0.115**	**16**	**TYR**	**H**
154	**TYR**	**HA**	**H**	**3.392**	**0.003**	**16**	**TYR**	**HA**
155	**TYR**	**HB2**	**H**	**3.133**	**0.004**	**16**	**TYR**	**HB2**
156	**TYR**	**HB3**	**H**	**2.942**	**0.004**	**16**	**TYR**	**HB3**
157	**TYR**	**C**	**C**	**175.873**	**0.036**	**16**	**TYR**	**C**
158	**TYR**	**CA**	**C**	**65.177**	**0.078**	**16**	**TYR**	**CA**
159	**TYR**	**CB**	**C**	**38.137**	**0.123**	**16**	**TYR**	**CB**
160	**TYR**	**N**	**N**	**121.237**	**0.083**	**16**	**TYR**	**N**
161	**GLU**	**H**	**H**	**8.882**	**0.01**	**17**	**GLU**	**H**

162	**GLU**	**HA**	**H**	**3.753**	**0.004**	**17**	**GLU**	**HA**
163	**GLU**	**HB2**	**H**	**2.07**	**0.018**	**17**	**GLU**	**HB2**
164	**GLU**	**HB3**	**H**	**1.993**	**0.001**	**17**	**GLU**	**HB3**
165	**GLU**	**HG2**	**H**	**2.322**	**0.008**	**17**	**GLU**	**HG2**
166	**GLU**	**HG3**	**H**	**2.251**	**0.01**	**17**	**GLU**	**HG3**
167	**GLU**	**C**	**C**	**179.079**	**0.009**	**17**	**GLU**	**C**
168	**GLU**	**CA**	**C**	**59.893**	**0.095**	**17**	**GLU**	**CA**
169	**GLU**	**CB**	**C**	**28.888**	**0.043**	**17**	**GLU**	**CB**
170	**GLU**	**CG**	**C**	**36.359**	**0.034**	**17**	**GLU**	**CG**
171	**GLU**	**N**	**N**	**117.085**	**0.071**	**17**	**GLU**	**N**
172	**VAL**	**H**	**H**	**7.28**	**0.01**	**18**	**VAL**	**H**
173	**VAL**	**HA**	**H**	**3.677**	**0.006**	**18**	**VAL**	**HA**
174	**VAL**	**HB**	**H**	**2.039**	**0.003**	**18**	**VAL**	**HB**
175	**VAL**	**HG11**	**H**	**0.953**	**0.006**	**18**	**VAL**	**HG11**
176	**VAL**	**HG12**	**H**	**0.953**	**0.006**	**18**	**VAL**	**HG12**
177	**VAL**	**HG13**	**H**	**0.953**	**0.006**	**18**	**VAL**	**HG13**
178	**VAL**	**HG21**	**H**	**0.873**	**0.002**	**18**	**VAL**	**HG21**
179	**VAL**	**HG22**	**H**	**0.873**	**0.002**	**18**	**VAL**	**HG22**
180	**VAL**	**HG23**	**H**	**0.873**	**0.002**	**18**	**VAL**	**HG23**
181	**VAL**	**C**	**C**	**179.622**	**0**	**18**	**VAL**	**C**
182	**VAL**	**CA**	**C**	**65.994**	**0.105**	**18**	**VAL**	**CA**
183	**VAL**	**CB**	**C**	**31.677**	**0.128**	**18**	**VAL**	**CB**
184	**VAL**	**CG1**	**C**	**21.415**	**0.189**	**18**	**VAL**	**CG1**
185	**VAL**	**CG2**	**C**	**21.055**	**0**	**18**	**VAL**	**CG2**
186	**VAL**	**N**	**N**	**118.55**	**0.053**	**18**	**VAL**	**N**
187	**CYS**	**H**	**H**	**7.513**	**0.012**	**19**	**CYS**	**H**
188	**CYS**	**HA**	**H**	**4.222**	**0.003**	**19**	**CYS**	**HA**
189	**CYS**	**HB2**	**H**	**2.859**	**0.006**	**19**	**CYS**	**HB2**
190	**CYS**	**HB3**	**H**	**2.315**	**0.003**	**19**	**CYS**	**HB3**
191	**CYS**	**C**	**C**	**175.973**	**0.008**	**19**	**CYS**	**C**
192	**CYS**	**CA**	**C**	**59.835**	**0.071**	**19**	**CYS**	**CA**
193	**CYS**	**CB**	**C**	**41.433**	**0.078**	**19**	**CYS**	**CB**
194	**CYS**	**N**	**N**	**119.218**	**0.068**	**19**	**CYS**	**N**
195	**ALA**	**H**	**H**	**8.984**	**0.017**	**20**	**ALA**	**H**
196	**ALA**	**HA**	**H**	**3.571**	**0.002**	**20**	**ALA**	**HA**
197	**ALA**	**HB1**	**H**	**1.126**	**0.001**	**20**	**ALA**	**HB1**
198	**ALA**	**HB2**	**H**	**1.126**	**0.001**	**20**	**ALA**	**HB2**
199	**ALA**	**HB3**	**H**	**1.126**	**0.001**	**20**	**ALA**	**HB3**
200	**ALA**	**C**	**C**	**180.092**	**0.01**	**20**	**ALA**	**C**
201	**ALA**	**CA**	**C**	**55.074**	**0.066**	**20**	**ALA**	**CA**
202	**ALA**	**CB**	**C**	**17.45**	**0.02**	**20**	**ALA**	**CB**
203	**ALA**	**N**	**N**	**122.541**	**0.066**	**20**	**ALA**	**N**
204	**LYS**	**H**	**H**	**7.323**	**0.008**	**21**	**LYS**	**H**
205	**LYS**	**HA**	**H**	**4.148**	**0.04**	**21**	**LYS**	**HA**
206	**LYS**	**HB2**	**H**	**1.843**	**0.002**	**21**	**LYS**	**HB2**
207	**LYS**	**HG2**	**H**	**1.495**	**0**	**21**	**LYS**	**HG2**
208	**LYS**	**HG3**	**H**	**1.44**	**0**	**21**	**LYS**	**HG3**
209	**LYS**	**HD2**	**H**	**1.626**	**0.004**	**21**	**LYS**	**HD2**
210	**LYS**	**HD3**	**H**	**1.52**	**0.007**	**21**	**LYS**	**HD3**
211	**LYS**	**C**	**C**	**179.13**	**0.013**	**21**	**LYS**	**C**
212	**LYS**	**CA**	**C**	**58.37**	**0.073**	**21**	**LYS**	**CA**
213	**LYS**	**CB**	**C**	**32.266**	**0.166**	**21**	**LYS**	**CB**
214	**LYS**	**CG**	**C**	**24.618**	**0**	**21**	**LYS**	**CG**
215	**LYS**	**CD**	**C**	**28.92**	**0**	**21**	**LYS**	**CD**
216	**LYS**	**CE**	**C**	**42.152**	**0**	**21**	**LYS**	**CE**
217	**LYS**	**N**	**N**	**116.951**	**0**	**21**	**LYS**	**N**
218	**GLU**	**H**	**H**	**8.05**	**0.138**	**22**	**GLU**	**H**
219	**GLU**	**HA**	**H**	**3.785**	**0.002**	**22**	**GLU**	**HA**
220	**GLU**	**HB2**	**H**	**1.814**	**0.004**	**22**	**GLU**	**HB2**
221	**GLU**	**HB3**	**H**	**1.706**	**0.001**	**22**	**GLU**	**HB3**
222	**GLU**	**HG2**	**H**	**2.235**	**0.004**	**22**	**GLU**	**HG2**
223	**GLU**	**HG3**	**H**	**2.034**	**0.002**	**22**	**GLU**	**HG3**

224	GLU	C	C	178.101	0.03	22	GLU	C
225	GLU	CA	C	59.514	0.076	22	GLU	CA
226	GLU	CB	C	30.31	0.029	22	GLU	CB
227	GLU	CG	C	36.257	0.061	22	GLU	CG
228	GLU	N	N	119.633	0.086	22	GLU	N
229	TYR	H	H	8.047	0.01	23	TYR	H
230	TYR	HA	H	4.343	0.003	23	TYR	HA
231	TYR	HB2	H	3.124	0.003	23	TYR	HB2
232	TYR	HB3	H	2.495	0.006	23	TYR	HB3
233	TYR	C	C	174.433	0.007	23	TYR	C
234	TYR	CA	C	58.241	0.066	23	TYR	CA
235	TYR	CB	C	37.793	0.111	23	TYR	CB
236	TYR	N	N	113.47	0.075	23	TYR	N
237	ASN	H	H	7.445	0.008	24	ASN	H
238	ASN	HA	H	4.368	0.006	24	ASN	HA
239	ASN	HB2	H	3.009	0.005	24	ASN	HB2
240	ASN	HB3	H	2.596	0.006	24	ASN	HB3
241	ASN	C	C	175.191	0.006	24	ASN	C
242	ASN	CA	C	54.021	0.268	24	ASN	CA
243	ASN	CB	C	36.726	0.174	24	ASN	CB
244	ASN	N	N	119.729	0.07	24	ASN	N
245	LEU	H	H	8.251	0.009	25	LEU	H
246	LEU	HA	H	4.44	0.004	25	LEU	HA
247	LEU	HB2	H	1.258	0	25	LEU	HB2
248	LEU	HB3	H	1.258	0	25	LEU	HB3
249	LEU	HG	H	1.225	0.002	25	LEU	HG
250	LEU	HD11	H	0.752	0.004	25	LEU	HD11
251	LEU	HD12	H	0.752	0.004	25	LEU	HD12
252	LEU	HD13	H	0.752	0.004	25	LEU	HD13
253	LEU	HD21	H	0.492	0.001	25	LEU	HD21
254	LEU	HD22	H	0.492	0.001	25	LEU	HD22
255	LEU	HD23	H	0.492	0.001	25	LEU	HD23
256	LEU	CA	C	52.84	0.056	25	LEU	CA
257	LEU	CB	C	40.784	0	25	LEU	CB
258	LEU	N	N	117.191	0.07	25	LEU	N
259	PRO	HA	H	4.654	0.004	26	PRO	HA
260	PRO	HB2	H	2.381	0.002	26	PRO	HB2
261	PRO	HB3	H	2.174	0.017	26	PRO	HB3
262	PRO	HG2	H	1.945	0	26	PRO	HG2
263	PRO	HD2	H	3.82	0.316	26	PRO	HD2
264	PRO	HD3	H	3.515	0.295	26	PRO	HD3
265	PRO	C	C	177.511	0.009	26	PRO	C
266	PRO	CA	C	62.01	0.268	26	PRO	CA
267	PRO	CB	C	31.971	0.048	26	PRO	CB
268	PRO	CG	C	26.799	0.074	26	PRO	CG
269	PRO	CD	C	50.136	0.082	26	PRO	CD
270	PRO	N	N	122.748	0	26	PRO	N
271	GLU	H	H	8.871	0.009	27	GLU	H
272	GLU	HA	H	4.114	0.005	27	GLU	HA
273	GLU	HB2	H	2	0.004	27	GLU	HB2
274	GLU	HB3	H	2	0.004	27	GLU	HB3
275	GLU	HG2	H	2.325	0	27	GLU	HG2
276	GLU	HG3	H	2.288	0	27	GLU	HG3
277	GLU	C	C	179.036	0.016	27	GLU	C
278	GLU	CA	C	59.591	0.116	27	GLU	CA
279	GLU	CB	C	29.265	0.062	27	GLU	CB
280	GLU	CG	C	36.397	0.04	27	GLU	CG
281	GLU	N	N	122.884	0.069	27	GLU	N
282	ALA	H	H	8.422	0.016	28	ALA	H
283	ALA	HA	H	4.051	0.005	28	ALA	HA
284	ALA	HB1	H	1.372	0	28	ALA	HB1
285	ALA	HB2	H	1.372	0	28	ALA	HB2

286	ALA	HB3	H	1.372	0	28	ALA	HB3
287	ALA	C	C	179.603	0.022	28	ALA	C
288	ALA	CA	C	54.639	0.154	28	ALA	CA
289	ALA	CB	C	18.676	0.125	28	ALA	CB
290	ALA	N	N	119.282	0.059	28	ALA	N
291	ALA	H	H	7.621	0.007	29	ALA	H
292	ALA	HA	H	4.06	0.002	29	ALA	HA
293	ALA	HB1	H	1.372	0	29	ALA	HB1
294	ALA	HB2	H	1.372	0	29	ALA	HB2
295	ALA	HB3	H	1.372	0	29	ALA	HB3
296	ALA	C	C	178.691	0.011	29	ALA	C
297	ALA	CA	C	55.072	0.078	29	ALA	CA
298	ALA	CB	C	18.14	0.046	29	ALA	CB
299	ALA	N	N	119.371	0.064	29	ALA	N
300	GLY	H	H	7.217	0.009	30	GLY	H
301	GLY	HA2	H	3.341	0.004	30	GLY	HA2
302	GLY	HA3	H	3.053	0.007	30	GLY	HA3
303	GLY	C	C	175.438	0.031	30	GLY	C
304	GLY	CA	C	47.456	0.074	30	GLY	CA
305	GLY	N	N	106.145	0.069	30	GLY	N
306	ALA	H	H	7.364	0.014	31	ALA	H
307	ALA	HA	H	3.903	0.004	31	ALA	HA
308	ALA	HB1	H	1.322	0.005	31	ALA	HB1
309	ALA	HB2	H	1.322	0.005	31	ALA	HB2
310	ALA	HB3	H	1.322	0.005	31	ALA	HB3
311	ALA	C	C	179.864	0.028	31	ALA	C
312	ALA	CA	C	54.586	0.093	31	ALA	CA
313	ALA	CB	C	17.743	0.136	31	ALA	CB
314	ALA	N	N	122.609	0.085	31	ALA	N
315	GLU	H	H	7.439	0.012	32	GLU	H
316	GLU	HA	H	4.145	0.007	32	GLU	HA
317	GLU	HB2	H	2.285	0.003	32	GLU	HB2
318	GLU	HB3	H	2.114	0.005	32	GLU	HB3
319	GLU	HG2	H	2.621	0	32	GLU	HG2
320	GLU	HG3	H	2.285	0.005	32	GLU	HG3
321	GLU	C	C	178.421	0.019	32	GLU	C
322	GLU	CA	C	60.371	0.116	32	GLU	CA
323	GLU	CB	C	30.008	0.132	32	GLU	CB
324	GLU	CG	C	37.207	0.056	32	GLU	CG
325	GLU	N	N	117.326	0.115	32	GLU	N
326	VAL	H	H	7.569	0.009	33	VAL	H
327	VAL	HA	H	3.485	0.006	33	VAL	HA
328	VAL	HB	H	1.952	0.015	33	VAL	HB
329	VAL	HG11	H	0.992	0.008	33	VAL	HG11
330	VAL	HG12	H	0.992	0.008	33	VAL	HG12
331	VAL	HG13	H	0.992	0.008	33	VAL	HG13
332	VAL	HG21	H	0.763	0.009	33	VAL	HG21
333	VAL	HG22	H	0.763	0.009	33	VAL	HG22
334	VAL	HG23	H	0.763	0.009	33	VAL	HG23
335	VAL	C	C	177.206	1.005	33	VAL	C
336	VAL	CA	C	67.112	0.078	33	VAL	CA
337	VAL	CB	C	31.42	0.153	33	VAL	CB
338	VAL	CG1	C	22.946	0	33	VAL	CG1
339	VAL	CG2	C	20.961	0	33	VAL	CG2
340	VAL	N	N	115.91	0.123	33	VAL	N
341	MET	H	H	7.941	0.009	34	MET	H
342	MET	HA	H	4.05	0.009	34	MET	HA
343	MET	HB2	H	2.023	0	34	MET	HB2
344	MET	HB3	H	1.945	0.009	34	MET	HB3
345	MET	HG2	H	2.526	0.003	34	MET	HG2
346	MET	HG3	H	2.396	0.004	34	MET	HG3
347	MET	C	C	176.732	0.009	34	MET	C

348	**MET**	**CA**	**C**	**57.797**	**0.064**	**34**	**MET**	**CA**
349	**MET**	**CB**	**C**	**31.757**	**0.228**	**34**	**MET**	**CB**
350	**MET**	**CG**	**C**	**33.053**	**0**	**34**	**MET**	**CG**
351	**MET**	**N**	**N**	**116.932**	**0.081**	**34**	**MET**	**N**
352	**ASN**	**H**	**H**	**7.24**	**0.01**	**35**	**ASN**	**H**
353	**ASN**	**HA**	**H**	**4.672**	**0.002**	**35**	**ASN**	**HA**
354	**ASN**	**HB2**	**H**	**2.004**	**0.009**	**35**	**ASN**	**HB2**
355	**ASN**	**HB3**	**H**	**2.199**	**0.004**	**35**	**ASN**	**HB3**
356	**ASN**	**C**	**C**	**174.707**	**0.051**	**35**	**ASN**	**C**
357	**ASN**	**CA**	**C**	**52.927**	**0.042**	**35**	**ASN**	**CA**
358	**ASN**	**CB**	**C**	**38.8**	**0.095**	**35**	**ASN**	**CB**
359	**ASN**	**N**	**N**	**116.049**	**0.049**	**35**	**ASN**	**N**
360	**PHE**	**H**	**H**	**7.388**	**0.012**	**36**	**PHE**	**H**
361	**PHE**	**HA**	**H**	**4.684**	**0.023**	**36**	**PHE**	**HA**
362	**PHE**	**HB2**	**H**	**2.214**	**0**	**36**	**PHE**	**HB2**
363	**PHE**	**HB3**	**H**	**2.214**	**0**	**36**	**PHE**	**HB3**
364	**PHE**	**CA**	**C**	**63.504**	**0.091**	**36**	**PHE**	**CA**
365	**PHE**	**CB**	**C**	**40.536**	**0.052**	**36**	**PHE**	**CB**
366	**PHE**	**N**	**N**	**119.15**	**0.075**	**36**	**PHE**	**N**
367	**TRP**	**H**	**H**	**9.321**	**0.01**	**37**	**TRP**	**H**
368	**TRP**	**HA**	**H**	**4.788**	**0.013**	**37**	**TRP**	**HA**
369	**TRP**	**HB2**	**H**	**3.777**	**0.003**	**37**	**TRP**	**HB2**
370	**TRP**	**HB3**	**H**	**3.122**	**0.002**	**37**	**TRP**	**HB3**
371	**TRP**	**C**	**C**	**175.274**	**0.018**	**37**	**TRP**	**C**
372	**TRP**	**CA**	**C**	**56.18**	**0.116**	**37**	**TRP**	**CA**
373	**TRP**	**CB**	**C**	**29.148**	**0.105**	**37**	**TRP**	**CB**
374	**TRP**	**N**	**N**	**112.901**	**0.085**	**37**	**TRP**	**N**
375	**LYS**	**H**	**H**	**7.672**	**0.008**	**38**	**LYS**	**H**
376	**LYS**	**HA**	**H**	**3.644**	**0.003**	**38**	**LYS**	**HA**
377	**LYS**	**HB2**	**H**	**1.417**	**0**	**38**	**LYS**	**HB2**
378	**LYS**	**HB3**	**H**	**1.417**	**0**	**38**	**LYS**	**HB3**
379	**LYS**	**C**	**C**	**178.137**	**0.03**	**38**	**LYS**	**C**
380	**LYS**	**CA**	**C**	**57.817**	**0.062**	**38**	**LYS**	**CA**
381	**LYS**	**CB**	**C**	**31.86**	**0.038**	**38**	**LYS**	**CB**
382	**LYS**	**CG**	**C**	**24.63**	**0**	**38**	**LYS**	**CG**
383	**LYS**	**CD**	**C**	**29.293**	**0**	**38**	**LYS**	**CD**
384	**LYS**	**CE**	**C**	**42.066**	**0**	**38**	**LYS**	**CE**
385	**LYS**	**N**	**N**	**124.028**	**0.059**	**38**	**LYS**	**N**
386	**GLU**	**H**	**H**	**9.334**	**0.056**	**39**	**GLU**	**H**
387	**GLU**	**HA**	**H**	**3.78**	**0.011**	**39**	**GLU**	**HA**
388	**GLU**	**HB2**	**H**	**1.941**	**0.044**	**39**	**GLU**	**HB2**
389	**GLU**	**HB3**	**H**	**1.894**	**0**	**39**	**GLU**	**HB3**
390	**GLU**	**HG2**	**H**	**2.207**	**0.006**	**39**	**GLU**	**HG2**
391	**GLU**	**HG3**	**H**	**2.1**	**0.007**	**39**	**GLU**	**HG3**
392	**GLU**	**C**	**C**	**177.748**	**0.043**	**39**	**GLU**	**C**
393	**GLU**	**CA**	**C**	**58.498**	**0.148**	**39**	**GLU**	**CA**
394	**GLU**	**CB**	**C**	**29.203**	**0.065**	**39**	**GLU**	**CB**
395	**GLU**	**CG**	**C**	**36.274**	**0.013**	**39**	**GLU**	**CG**
396	**GLU**	**N**	**N**	**135.575**	**0.467**	**39**	**GLU**	**N**
397	**GLY**	**H**	**H**	**9.008**	**0.012**	**40**	**GLY**	**H**
398	**GLY**	**HA2**	**H**	**3.951**	**0.002**	**40**	**GLY**	**HA2**
399	**GLY**	**HA3**	**H**	**3.513**	**0.032**	**40**	**GLY**	**HA3**
400	**GLY**	**C**	**C**	**172.96**	**0.009**	**40**	**GLY**	**C**
401	**GLY**	**CA**	**C**	**45.4**	**0.134**	**40**	**GLY**	**CA**
402	**GLY**	**N**	**N**	**114.383**	**0.062**	**40**	**GLY**	**N**
403	**TYR**	**H**	**H**	**7.261**	**0.009**	**41**	**TYR**	**H**
404	**TYR**	**HA**	**H**	**4.202**	**0.003**	**41**	**TYR**	**HA**
405	**TYR**	**HB2**	**H**	**2.511**	**0.003**	**41**	**TYR**	**HB2**
406	**TYR**	**HB3**	**H**	**2.43**	**0.005**	**41**	**TYR**	**HB3**
407	**TYR**	**C**	**C**	**174.235**	**0.053**	**41**	**TYR**	**C**
408	**TYR**	**CA**	**C**	**58.201**	**0.097**	**41**	**TYR**	**CA**
409	**TYR**	**CB**	**C**	**40.81**	**0.155**	**41**	**TYR**	**CB**

410	**TYR**	**N**	**N**	**121.488**	**0.07**	**41**	**TYR**	**N**
411	**VAL**	**H**	**H**	**7.159**	**0.013**	**42**	**VAL**	**H**
412	**VAL**	**HA**	**H**	**3.836**	**0.004**	**42**	**VAL**	**HA**
413	**VAL**	**HB**	**H**	**1.549**	**0.005**	**42**	**VAL**	**HB**
414	**VAL**	**HG11**	**H**	**0.609**	**0.007**	**42**	**VAL**	**HG11**
415	**VAL**	**HG12**	**H**	**0.609**	**0.007**	**42**	**VAL**	**HG12**
416	**VAL**	**HG13**	**H**	**0.609**	**0.007**	**42**	**VAL**	**HG13**
417	**VAL**	**HG21**	**H**	**0.609**	**0.007**	**42**	**VAL**	**HG21**
418	**VAL**	**HG22**	**H**	**0.609**	**0.007**	**42**	**VAL**	**HG22**
419	**VAL**	**HG23**	**H**	**0.609**	**0.007**	**42**	**VAL**	**HG23**
420	**VAL**	**CA**	**C**	**60.607**	**0.046**	**42**	**VAL**	**CA**
421	**VAL**	**CB**	**C**	**33.867**	**0.025**	**42**	**VAL**	**CB**
422	**VAL**	**CG1**	**C**	**21.894**	**1.697**	**42**	**VAL**	**CG1**
423	**VAL**	**CG2**	**C**	**20.66**	**0**	**42**	**VAL**	**CG2**
424	**VAL**	**N**	**N**	**127.2**	**0.054**	**42**	**VAL**	**N**
425	**LEU**	**H**	**H**	**7.267**	**0.007**	**43**	**LEU**	**H**
426	**LEU**	**HA**	**H**	**4.126**	**0.003**	**43**	**LEU**	**HA**
427	**LEU**	**HB2**	**H**	**1.441**	**0.001**	**43**	**LEU**	**HB2**
428	**LEU**	**HB3**	**H**	**1.441**	**0.001**	**43**	**LEU**	**HB3**
429	**LEU**	**HG**	**H**	**1.335**	**0.002**	**43**	**LEU**	**HG**
430	**LEU**	**HD11**	**H**	**0.841**	**0.001**	**43**	**LEU**	**HD11**
431	**LEU**	**HD12**	**H**	**0.841**	**0.001**	**43**	**LEU**	**HD12**
432	**LEU**	**HD13**	**H**	**0.841**	**0.001**	**43**	**LEU**	**HD13**
433	**LEU**	**HD21**	**H**	**0.789**	**0.005**	**43**	**LEU**	**HD21**
434	**LEU**	**HD22**	**H**	**0.789**	**0.005**	**43**	**LEU**	**HD22**
435	**LEU**	**HD23**	**H**	**0.789**	**0.005**	**43**	**LEU**	**HD23**
436	**LEU**	**C**	**C**	**175.17**	**0.012**	**43**	**LEU**	**C**
437	**LEU**	**CA**	**C**	**56.457**	**0.06**	**43**	**LEU**	**CA**
438	**LEU**	**CB**	**C**	**42.768**	**0.055**	**43**	**LEU**	**CB**
439	**LEU**	**CG**	**C**	**29.577**	**0**	**43**	**LEU**	**CG**
440	**LEU**	**CD1**	**C**	**25.579**	**0**	**43**	**LEU**	**CD1**
441	**LEU**	**N**	**N**	**128.255**	**0.071**	**43**	**LEU**	**N**
442	**THR**	**H**	**H**	**9.03**	**0.031**	**44**	**THR**	**H**
443	**THR**	**HA**	**H**	**4.33**	**0.002**	**44**	**THR**	**HA**
444	**THR**	**HB**	**H**	**4.073**	**0.003**	**44**	**THR**	**HB**
445	**THR**	**HG21**	**H**	**0.929**	**0.004**	**44**	**THR**	**HG21**
446	**THR**	**HG22**	**H**	**0.929**	**0.004**	**44**	**THR**	**HG22**
447	**THR**	**HG23**	**H**	**0.929**	**0.004**	**44**	**THR**	**HG23**
448	**THR**	**C**	**C**	**174.527**	**0.889**	**44**	**THR**	**C**
449	**THR**	**CA**	**C**	**61.962**	**0.195**	**44**	**THR**	**CA**
450	**THR**	**CB**	**C**	**71.586**	**0.151**	**44**	**THR**	**CB**
451	**THR**	**CG2**	**C**	**21.155**	**0.443**	**44**	**THR**	**CG2**
452	**THR**	**N**	**N**	**110.994**	**0.085**	**44**	**THR**	**N**
453	**SER**	**H**	**H**	**9.094**	**0.013**	**45**	**SER**	**H**
454	**SER**	**HA**	**H**	**4.435**	**0.009**	**45**	**SER**	**HA**
455	**SER**	**HB2**	**H**	**4.145**	**0.012**	**45**	**SER**	**HB2**
456	**SER**	**HB3**	**H**	**3.946**	**0.007**	**45**	**SER**	**HB3**
457	**SER**	**C**	**C**	**174.779**	**0.003**	**45**	**SER**	**C**
458	**SER**	**CA**	**C**	**58.172**	**0.153**	**45**	**SER**	**CA**
459	**SER**	**CB**	**C**	**64.135**	**0.278**	**45**	**SER**	**CB**
460	**SER**	**N**	**N**	**121.721**	**0.06**	**45**	**SER**	**N**
461	**ARG**	**H**	**H**	**8.945**	**0.011**	**46**	**ARG**	**H**
462	**ARG**	**HA**	**H**	**3.609**	**0.002**	**46**	**ARG**	**HA**
463	**ARG**	**HB2**	**H**	**1.859**	**0.002**	**46**	**ARG**	**HB2**
464	**ARG**	**HB3**	**H**	**1.632**	**0.038**	**46**	**ARG**	**HB3**
465	**ARG**	**HD2**	**H**	**3.187**	**0.001**	**46**	**ARG**	**HD2**
466	**ARG**	**HD3**	**H**	**3.034**	**0.002**	**46**	**ARG**	**HD3**
467	**ARG**	**C**	**C**	**178.265**	**0.005**	**46**	**ARG**	**C**
468	**ARG**	**CA**	**C**	**60.004**	**0.074**	**46**	**ARG**	**CA**
469	**ARG**	**CB**	**C**	**29.452**	**0.134**	**46**	**ARG**	**CB**
470	**ARG**	**CG**	**C**	**28.543**	**0**	**46**	**ARG**	**CG**
471	**ARG**	**CD**	**C**	**42.839**	**0.032**	**46**	**ARG**	**CD**

472	ARG	N	N	125.834	0.082	46	ARG	N
473	GLU	H	H	9.046	0.009	47	GLU	H
474	GLU	HA	H	3.803	0.004	47	GLU	HA
475	GLU	HB2	H	2.075	0.004	47	GLU	HB2
476	GLU	HB3	H	1.782	0.004	47	GLU	HB3
477	GLU	HG2	H	2.456	0.006	47	GLU	HG2
478	GLU	HG3	H	2.185	0.004	47	GLU	HG3
479	GLU	C	C	178.479	0.008	47	GLU	C
480	GLU	CA	C	61.374	0.095	47	GLU	CA
481	GLU	CB	C	28.601	0.074	47	GLU	CB
482	GLU	CG	C	37.847	0.013	47	GLU	CG
483	GLU	N	N	116.251	0.062	47	GLU	N
484	ALA	H	H	8.065	0.009	48	ALA	H
485	ALA	HA	H	3.97	0.004	48	ALA	HA
486	ALA	HB1	H	1.538	0.003	48	ALA	HB1
487	ALA	HB2	H	1.538	0.003	48	ALA	HB2
488	ALA	HB3	H	1.538	0.003	48	ALA	HB3
489	ALA	C	C	179.763	0.009	48	ALA	C
490	ALA	CA	C	55.462	0.122	48	ALA	CA
491	ALA	CB	C	18.414	0.035	48	ALA	CB
492	ALA	N	N	122.898	0.071	48	ALA	N
493	GLY	H	H	7.543	0.009	49	GLY	H
494	GLY	HA2	H	4.122	0.039	49	GLY	HA2
495	GLY	HA3	H	3.45	0.036	49	GLY	HA3
496	GLY	C	C	175.664	0.005	49	GLY	C
497	GLY	CA	C	48.383	0.685	49	GLY	CA
498	GLY	N	N	105.63	0.058	49	GLY	N
499	CYS	H	H	8.551	0.016	50	CYS	H
500	CYS	HA	H	4.442	0.003	50	CYS	HA
501	CYS	HB2	H	2.837	0.003	50	CYS	HB2
502	CYS	HB3	H	2.71	0.003	50	CYS	HB3
503	CYS	C	C	177.735	0.244	50	CYS	C
504	CYS	CA	C	55.229	0.033	50	CYS	CA
505	CYS	CB	C	34.937	0.139	50	CYS	CB
506	CYS	N	N	119.096	0.049	50	CYS	N
507	ALA	H	H	8.466	0.018	51	ALA	H
508	ALA	HA	H	3.582	0.004	51	ALA	HA
509	ALA	HB1	H	1.543	0.01	51	ALA	HB1
510	ALA	HB2	H	1.543	0.01	51	ALA	HB2
511	ALA	HB3	H	1.543	0.01	51	ALA	HB3
512	ALA	C	C	178.123	0.01	51	ALA	C
513	ALA	CA	C	55.978	0.114	51	ALA	CA
514	ALA	CB	C	18.085	0.084	51	ALA	CB
515	ALA	N	N	123.392	0.074	51	ALA	N
516	ILE	H	H	8.052	0.01	52	ILE	H
517	ILE	HA	H	3.545	0.007	52	ILE	HA
518	ILE	HB	H	1.92	0.002	52	ILE	HB
519	ILE	HG21	H	0.864	0.002	52	ILE	HG21
520	ILE	HG22	H	0.864	0.002	52	ILE	HG22
521	ILE	HG23	H	0.864	0.002	52	ILE	HG23
522	ILE	HD11	H	0.765	0	52	ILE	HD11
523	ILE	HD12	H	0.765	0	52	ILE	HD12
524	ILE	HD13	H	0.765	0	52	ILE	HD13
525	ILE	C	C	174.456	0	52	ILE	C
526	ILE	CA	C	65.808	0.049	52	ILE	CA
527	ILE	CB	C	37.741	0.052	52	ILE	CB
528	ILE	CG1	C	29.898	0	52	ILE	CG1
529	ILE	CG2	C	17.364	0	52	ILE	CG2
530	ILE	CD1	C	12.445	0	52	ILE	CD1
531	ILE	N	N	118.417	0.094	52	ILE	N
532	LEU	H	H	8.038	0.007	53	LEU	H
533	LEU	HA	H	3.845	0	53	LEU	HA

534	LEU	C	C	174.678	0.035	53	LEU	C
535	LEU	CA	C	57.652	1.656	53	LEU	CA
536	LEU	CB	C	41.791	0.466	53	LEU	CB
537	LEU	CG	C	26.51	0	53	LEU	CG
538	LEU	CD1	C	24.885	0	53	LEU	CD1
539	LEU	N	N	123.06	0.1	53	LEU	N
540	CYS	H	H	8.067	0.01	54	CYS	H
541	CYS	HA	H	4.015	0.005	54	CYS	HA
542	CYS	HB2	H	2.83	0.016	54	CYS	HB2
543	CYS	HB3	H	2.83	0.016	54	CYS	HB3
544	CYS	C	C	176.429	0.237	54	CYS	C
545	CYS	CA	C	60.003	0.036	54	CYS	CA
546	CYS	CB	C	42.17	0.012	54	CYS	CB
547	CYS	N	N	119.61	0.057	54	CYS	N
548	LEU	H	H	8.88	0.009	55	LEU	H
549	LEU	HA	H	3.53	0.006	55	LEU	HA
550	LEU	HB2	H	1.842	0.001	55	LEU	HB2
551	LEU	HB3	H	1.271	0.006	55	LEU	HB3
552	LEU	HG	H	0.694	0.001	55	LEU	HG
553	LEU	HD11	H	0.613	0	55	LEU	HD11
554	LEU	HD12	H	0.613	0	55	LEU	HD12
555	LEU	HD13	H	0.613	0	55	LEU	HD13
556	LEU	HD21	H	0.244	0.004	55	LEU	HD21
557	LEU	HD22	H	0.244	0.004	55	LEU	HD22
558	LEU	HD23	H	0.244	0.004	55	LEU	HD23
559	LEU	C	C	179.013	0	55	LEU	C
560	LEU	CA	C	57.628	0.14	55	LEU	CA
561	LEU	CB	C	40.417	0.066	55	LEU	CB
562	LEU	CG	C	26.048	0	55	LEU	CG
563	LEU	CD1	C	26.048	0	55	LEU	CD1
564	LEU	N	N	122.92	0.054	55	LEU	N
565	SER	H	H	7.851	0.01	56	SER	H
566	SER	HA	H	3.883	0.01	56	SER	HA
567	SER	HB2	H	3.685	0.097	56	SER	HB2
568	SER	HB3	H	3.547	0.01	56	SER	HB3
569	SER	C	C	176.08	0	56	SER	C
570	SER	CA	C	62.966	0.045	56	SER	CA
571	SER	CB	C	63.429	0.05	56	SER	CB
572	SER	N	N	114.271	0.05	56	SER	N
573	SER	H	H	8.037	0.008	57	SER	H
574	SER	HA	H	4.277	0.003	57	SER	HA
575	SER	HB2	H	3.975	0.015	57	SER	HB2
576	SER	HB3	H	3.975	0.015	57	SER	HB3
577	SER	CA	C	61.337	0.069	57	SER	CA
578	SER	CB	C	62.742	0.134	57	SER	CB
579	SER	N	N	118.906	0.044	57	SER	N
580	LYS	H	H	8.359	0.009	58	LYS	H
581	LYS	CA	C	57.399	0.053	58	LYS	CA
582	LYS	CB	C	31.008	0.014	58	LYS	CB
583	LYS	CG	C	24.75	0	58	LYS	CG
584	LYS	CD	C	27.801	0	58	LYS	CD
585	LYS	CE	C	42.31	0	58	LYS	CE
586	LYS	N	N	122.202	0.041	58	LYS	N
587	LEU	H	H	6.65	0.009	59	LEU	H
588	LEU	HA	H	4.216	0.008	59	LEU	HA
589	LEU	HB2	H	1.402	0.005	59	LEU	HB2
590	LEU	HB3	H	1.249	0.005	59	LEU	HB3
591	LEU	HG	H	1.197	0	59	LEU	HG
592	LEU	HD11	H	0.7	0.007	59	LEU	HD11
593	LEU	HD12	H	0.7	0.007	59	LEU	HD12
594	LEU	HD13	H	0.7	0.007	59	LEU	HD13
595	LEU	HD21	H	0.376	0.004	59	LEU	HD21

596	LEU	HD22	H	0.376	0.004	59	LEU	HD22
597	LEU	HD23	H	0.376	0.004	59	LEU	HD23
598	LEU	CA	C	54.019	0.032	59	LEU	CA
599	LEU	CB	C	42.926	0.063	59	LEU	CB
600	LEU	CG	C	26.977	0	59	LEU	CG
601	LEU	CD1	C	24.734	0.019	59	LEU	CD1
602	LEU	CD2	C	22.645	0	59	LEU	CD2
603	LEU	N	N	118.06	0.051	59	LEU	N
604	ASN	H	H	7.879	0.007	60	ASN	H
605	ASN	HA	H	4.543	0.001	60	ASN	HA
606	ASN	HB2	H	2.887	0.005	60	ASN	HB2
607	ASN	HB3	H	2.677	0.005	60	ASN	HB3
608	ASN	C	C	175.155	0	60	ASN	C
609	ASN	CA	C	54.115	0.052	60	ASN	CA
610	ASN	CB	C	36.824	0.061	60	ASN	CB
611	ASN	N	N	114.352	0.058	60	ASN	N
612	LEU	H	H	7.756	0.011	61	LEU	H
613	LEU	CA	C	55.506	0	61	LEU	CA
614	LEU	CB	C	40.941	0	61	LEU	CB
615	LEU	N	N	111.024	0.048	61	LEU	N
616	ASP	H	H	9.267	0.009	63	ASP	H
617	ASP	CA	C	52.687	0	63	ASP	CA
618	ASP	CB	C	41.357	0	63	ASP	CB
619	ASP	N	N	126.175	0.047	63	ASP	N
620	PRO	HA	H	4.272	0.008	64	PRO	HA
621	PRO	HB2	H	2.357	0.005	64	PRO	HB2
622	PRO	HB3	H	2.065	0.004	64	PRO	HB3
623	PRO	HG2	H	1.939	0.003	64	PRO	HG2
624	PRO	HG3	H	1.851	0.007	64	PRO	HG3
625	PRO	HD2	H	3.841	0.005	64	PRO	HD2
626	PRO	HD3	H	3.707	0.006	64	PRO	HD3
627	PRO	C	C	177.507	0.007	64	PRO	C
628	PRO	CA	C	65.306	0.092	64	PRO	CA
629	PRO	CB	C	31.86	0.07	64	PRO	CB
630	PRO	CG	C	27.663	0.021	64	PRO	CG
631	PRO	CD	C	51.328	0.07	64	PRO	CD
632	GLU	H	H	8.489	0.012	65	GLU	H
633	GLU	HA	H	4.25	0.004	65	GLU	HA
634	GLU	HB2	H	1.948	0.005	65	GLU	HB2
635	GLU	HB3	H	1.948	0.005	65	GLU	HB3
636	GLU	HG2	H	2.188	0.003	65	GLU	HG2
637	GLU	HG3	H	2.114	0.008	65	GLU	HG3
638	GLU	C	C	177.086	0.011	65	GLU	C
639	GLU	CA	C	56.528	0.181	65	GLU	CA
640	GLU	CB	C	29.353	0.145	65	GLU	CB
641	GLU	CG	C	36.853	0.003	65	GLU	CG
642	GLU	N	N	115.671	0.092	65	GLU	N
643	GLY	H	H	8.327	0.009	66	GLY	H
644	GLY	HA2	H	4.175	0.02	66	GLY	HA2
645	GLY	HA3	H	3.475	0.015	66	GLY	HA3
646	GLY	C	C	173.196	0.016	66	GLY	C
647	GLY	CA	C	45.661	0	66	GLY	CA
648	GLY	N	N	108.009	0.077	66	GLY	N
649	THR	H	H	7.509	0.011	67	THR	H
650	THR	HA	H	4.505	0.008	67	THR	HA
651	THR	HB	H	4.147	0.004	67	THR	HB
652	THR	HG1	H	5.528	0.461	67	THR	HG1
653	THR	HG21	H	0.951	0.006	67	THR	HG21
654	THR	HG22	H	0.951	0.006	67	THR	HG22
655	THR	HG23	H	0.951	0.006	67	THR	HG23
656	THR	C	C	173.64	0.005	67	THR	C
657	THR	CA	C	59.641	0.045	67	THR	CA

658	THR	CB	C	69.814	0.245	67	THR	CB
659	THR	CG2	C	21.115	0	67	THR	CG2
660	THR	N	N	113.73	0.077	67	THR	N
661	LEU	H	H	9.258	0.016	68	LEU	H
662	LEU	HA	H	4.771	0	68	LEU	HA
663	LEU	HB2	H	1.548	0	68	LEU	HB2
664	LEU	HB3	H	1.548	0	68	LEU	HB3
665	LEU	HG	H	1.454	0	68	LEU	HG
666	LEU	HD11	H	1.482	0	68	LEU	HD11
667	LEU	HD12	H	1.482	0	68	LEU	HD12
668	LEU	HD13	H	1.482	0	68	LEU	HD13
669	LEU	HD21	H	1.482	0	68	LEU	HD21
670	LEU	HD22	H	1.482	0	68	LEU	HD22
671	LEU	HD23	H	1.482	0	68	LEU	HD23
672	LEU	C	C	175.972	0	68	LEU	C
673	LEU	CA	C	56.656	0.023	68	LEU	CA
674	LEU	CB	C	42.611	0.094	68	LEU	CB
675	LEU	CG	C	27.401	0	68	LEU	CG
676	LEU	CD1	C	24.344	0	68	LEU	CD1
677	LEU	N	N	126.829	0.093	68	LEU	N
678	HIS	H	H	9.687	0.009	69	HIS	H
679	HIS	HA	H	4.345	0.021	69	HIS	HA
680	HIS	HB2	H	3.059	0	69	HIS	HB2
681	HIS	HB3	H	2.561	0	69	HIS	HB3
682	HIS	C	C	176.294	0.011	69	HIS	C
683	HIS	CA	C	57.73	0.242	69	HIS	CA
684	HIS	CB	C	31.85	0.082	69	HIS	CB
685	HIS	N	N	131.618	0.072	69	HIS	N
686	ARG	H	H	9.021	0.02	70	ARG	H
687	ARG	HA	H	4.611	0.004	70	ARG	HA
688	ARG	C	C	176.633	0.008	70	ARG	C
689	ARG	CA	C	60.337	0.123	70	ARG	CA
690	ARG	CB	C	30.449	0.047	70	ARG	CB
691	ARG	CG	C	26.688	0	70	ARG	CG
692	ARG	CD	C	42.9	0	70	ARG	CD
693	ARG	N	N	132.301	0.111	70	ARG	N
694	GLY	HA2	H	4.019	0.005	71	GLY	HA2
695	GLY	HA3	H	3.672	0.01	71	GLY	HA3
696	GLY	C	C	177.305	0	71	GLY	C
697	GLY	CA	C	47.758	0.171	71	GLY	CA
698	GLY	N	N	107.102	0.062	71	GLY	N
699	ASN	H	H	11.487	0.01	72	ASN	H
700	ASN	HA	H	4.481	0.009	72	ASN	HA
701	ASN	HB2	H	2.686	0.002	72	ASN	HB2
702	ASN	HB3	H	2.686	0.002	72	ASN	HB3
703	ASN	C	C	179.501	0.015	72	ASN	C
704	ASN	CA	C	55.348	0.035	72	ASN	CA
705	ASN	CB	C	36.805	0.047	72	ASN	CB
706	ASN	N	N	123.943	0.02	72	ASN	N
707	THR	H	H	8.15	0.007	73	THR	H
708	THR	HA	H	4.046	0.004	73	THR	HA
709	THR	HB	H	3.892	0	73	THR	HB
710	THR	HG1	H	5.212	0.007	73	THR	HG1
711	THR	HG21	H	1.083	0.002	73	THR	HG21
712	THR	HG22	H	1.083	0.002	73	THR	HG22
713	THR	HG23	H	1.083	0.002	73	THR	HG23
714	THR	CA	C	67	0.036	73	THR	CA
715	THR	CB	C	68.079	0.054	73	THR	CB
716	THR	CG2	C	21.171	0	73	THR	CG2
717	THR	N	N	120.667	0.056	73	THR	N
718	VAL	H	H	8.727	0.013	74	VAL	H
719	VAL	HA	H	3.448	0.005	74	VAL	HA

720	VAL	HB	H	2.048	0.004	74	VAL	HB
721	VAL	HG11	H	0.651	0.003	74	VAL	HG11
722	VAL	HG12	H	0.651	0.003	74	VAL	HG12
723	VAL	HG13	H	0.651	0.003	74	VAL	HG13
724	VAL	HG21	H	0.309	0.006	74	VAL	HG21
725	VAL	HG22	H	0.309	0.006	74	VAL	HG22
726	VAL	HG23	H	0.309	0.006	74	VAL	HG23
727	VAL	CA	C	66.919	0.056	74	VAL	CA
728	VAL	CB	C	30.738	0.085	74	VAL	CB
729	VAL	CG1	C	20.739	0.449	74	VAL	CG1
730	VAL	CG2	C	19.304	0	74	VAL	CG2
731	VAL	N	N	123.304	0.047	74	VAL	N
732	GLU	H	H	7.697	0.006	75	GLU	H
733	GLU	HA	H	3.939	0.002	75	GLU	HA
734	GLU	HB2	H	2.057	0	75	GLU	HB2
735	GLU	HB3	H	1.937	0	75	GLU	HB3
736	GLU	HG2	H	2.323	0.004	75	GLU	HG2
737	GLU	HG3	H	2.323	0.004	75	GLU	HG3
738	GLU	C	C	178.608	0	75	GLU	C
739	GLU	CA	C	59.669	0.048	75	GLU	CA
740	GLU	CB	C	29.297	0.019	75	GLU	CB
741	GLU	CG	C	35.988	0.001	75	GLU	CG
742	GLU	N	N	117.756	0.045	75	GLU	N
743	PHE	H	H	7.709	0.009	76	PHE	H
744	PHE	HA	H	4.442	0.003	76	PHE	HA
745	PHE	HB2	H	3.46	0.001	76	PHE	HB2
746	PHE	HB3	H	3.276	0.001	76	PHE	HB3
747	PHE	C	C	177.154	0	76	PHE	C
748	PHE	CA	C	60.663	0.066	76	PHE	CA
749	PHE	CB	C	39.491	0.098	76	PHE	CB
750	PHE	N	N	120.766	0.114	76	PHE	N
751	ALA	H	H	8.657	0.172	77	ALA	H
752	ALA	HA	H	3.924	0.004	77	ALA	HA
753	ALA	HB1	H	1.358	0.039	77	ALA	HB1
754	ALA	HB2	H	1.358	0.039	77	ALA	HB2
755	ALA	HB3	H	1.358	0.039	77	ALA	HB3
756	ALA	C	C	181.433	0	77	ALA	C
757	ALA	CA	C	55.297	0.042	77	ALA	CA
758	ALA	CB	C	16.945	0.022	77	ALA	CB
759	ALA	N	N	122.425	0.039	77	ALA	N
760	LYS	H	H	8.611	0.006	78	LYS	H
761	LYS	HA	H	4.05	0.002	78	LYS	HA
762	LYS	HG3	H	1.441	0	78	LYS	HG3
763	LYS	HD2	H	1.847	0	78	LYS	HD2
764	LYS	C	C	178.971	1.716	78	LYS	C
765	LYS	CA	C	59.619	0.103	78	LYS	CA
766	LYS	CB	C	31.486	0.237	78	LYS	CB
767	LYS	CG	C	25.225	0	78	LYS	CG
768	LYS	CD	C	28.976	0	78	LYS	CD
769	LYS	CE	C	41.627	0	78	LYS	CE
770	LYS	N	N	119.167	0.056	78	LYS	N
771	GLN	H	H	7.98	0.013	79	GLN	H
772	GLN	HA	H	3.865	0.003	79	GLN	HA
773	GLN	HB2	H	2.025	0.003	79	GLN	HB2
774	GLN	HB3	H	1.733	0.002	79	GLN	HB3
775	GLN	HG2	H	2.266	0.002	79	GLN	HG2
776	GLN	HG3	H	2.159	0.001	79	GLN	HG3
777	GLN	C	C	176.489	0.006	79	GLN	C
778	GLN	CA	C	57.95	0.169	79	GLN	CA
779	GLN	CB	C	27.682	0.095	79	GLN	CB
780	GLN	CG	C	33.677	0.049	79	GLN	CG
781	GLN	N	N	120.527	1.129	79	GLN	N

782	HIS	H	H	7.059	0.01	80	HIS	H
783	HIS	HA	H	4.542	0.005	80	HIS	HA
784	HIS	HB2	H	3.374	0.001	80	HIS	HB2
785	HIS	HB3	H	2.349	0.004	80	HIS	HB3
786	HIS	C	C	173.751	0.005	80	HIS	C
787	HIS	CA	C	56.287	0.058	80	HIS	CA
788	HIS	CB	C	29.801	0.051	80	HIS	CB
789	HIS	N	N	115.3	0.072	80	HIS	N
790	GLY	H	H	7.29	0.01	81	GLY	H
791	GLY	HA2	H	4.154	0.004	81	GLY	HA2
792	GLY	HA3	H	3.581	0.002	81	GLY	HA3
793	GLY	C	C	174.987	0.011	81	GLY	C
794	GLY	CA	C	45.221	0.186	81	GLY	CA
795	GLY	N	N	104.005	0.085	81	GLY	N
796	SER	H	H	7.727	0.007	82	SER	H
797	SER	HA	H	4.094	0.005	82	SER	HA
798	SER	HB2	H	3.448	0.005	82	SER	HB2
799	SER	HB3	H	3.157	0.002	82	SER	HB3
800	SER	HG	H	5.487	0.044	82	SER	HG
801	SER	C	C	174.64	0.007	82	SER	C
802	SER	CA	C	59.422	0.141	82	SER	CA
803	SER	CB	C	64.818	0.151	82	SER	CB
804	SER	N	N	115.906	0.062	82	SER	N
805	ASP	H	H	7.986	0.011	83	ASP	H
806	ASP	HA	H	4.815	0.021	83	ASP	HA
807	ASP	HB2	H	3.095	0.006	83	ASP	HB2
808	ASP	HB3	H	2.747	0.007	83	ASP	HB3
809	ASP	C	C	177.246	0.026	83	ASP	C
810	ASP	CA	C	52.545	0.131	83	ASP	CA
811	ASP	CB	C	41.244	0.163	83	ASP	CB
812	ASP	N	N	122.766	0.046	83	ASP	N
813	ASP	H	H	8.516	0.016	84	ASP	H
814	ASP	HA	H	4.174	0.01	84	ASP	HA
815	ASP	HB2	H	2.618	0.007	84	ASP	HB2
816	ASP	HB3	H	2.582	0.015	84	ASP	HB3
817	ASP	C	C	178.134	1.393	84	ASP	C
818	ASP	CA	C	58.717	0.442	84	ASP	CA
819	ASP	CB	C	41.58	0.111	84	ASP	CB
820	ASP	N	N	120.558	0.082	84	ASP	N
821	ALA	H	H	8.021	0.008	85	ALA	H
822	ALA	HA	H	4.114	0.001	85	ALA	HA
823	ALA	HB1	H	1.44	0.003	85	ALA	HB1
824	ALA	HB2	H	1.44	0.003	85	ALA	HB2
825	ALA	HB3	H	1.44	0.003	85	ALA	HB3
826	ALA	C	C	181.078	0.004	85	ALA	C
827	ALA	CA	C	55.345	0.026	85	ALA	CA
828	ALA	CB	C	17.849	0.016	85	ALA	CB
829	ALA	N	N	120.68	0.046	85	ALA	N
830	MET	H	H	8.581	0.009	86	MET	H
831	MET	HA	H	4.052	0.003	86	MET	HA
832	MET	HB2	H	2.04	0	86	MET	HB2
833	MET	HB3	H	1.941	0	86	MET	HB3
834	MET	HG2	H	2.76	0	86	MET	HG2
835	MET	HG3	H	2.76	0	86	MET	HG3
836	MET	HE1	H	1.043	0	86	MET	HE1
837	MET	HE2	H	1.043	0	86	MET	HE2
838	MET	HE3	H	1.043	0	86	MET	HE3
839	MET	C	C	177.136	0.028	86	MET	C
840	MET	CA	C	59.328	0.045	86	MET	CA
841	MET	CB	C	31.271	0.044	86	MET	CB
842	MET	CG	C	33.014	0	86	MET	CG
843	MET	N	N	119.929	0.038	86	MET	N

844	ALA	H	H	7.468	0.008	87	ALA	H
845	ALA	HA	H	3.678	0.005	87	ALA	HA
846	ALA	HB1	H	1.426	0.005	87	ALA	HB1
847	ALA	HB2	H	1.426	0.005	87	ALA	HB2
848	ALA	HB3	H	1.426	0.005	87	ALA	HB3
849	ALA	C	C	178.747	0.009	87	ALA	C
850	ALA	CA	C	55.681	0.084	87	ALA	CA
851	ALA	CB	C	20.052	0.145	87	ALA	CB
852	ALA	N	N	120.068	0.03	87	ALA	N
853	HIS	H	H	8.044	0.007	88	HIS	H
854	HIS	HA	H	3.877	0.007	88	HIS	HA
855	HIS	HB2	H	3.135	0.009	88	HIS	HB2
856	HIS	HB3	H	2.979	0.001	88	HIS	HB3
857	HIS	C	C	178.174	0.018	88	HIS	C
858	HIS	CA	C	60.677	0.051	88	HIS	CA
859	HIS	CB	C	30.979	0.19	88	HIS	CB
860	HIS	N	N	113.454	0.089	88	HIS	N
861	GLN	H	H	8.037	0.007	89	GLN	H
862	GLN	HA	H	4.212	0.003	89	GLN	HA
863	GLN	HB2	H	2.551	0.004	89	GLN	HB2
864	GLN	HB3	H	2.378	0.005	89	GLN	HB3
865	GLN	C	C	178.513	0.004	89	GLN	C
866	GLN	CA	C	59.302	0.034	89	GLN	CA
867	GLN	CB	C	28.111	0.105	89	GLN	CB
868	GLN	CG	C	33.569	0.016	89	GLN	CG
869	GLN	N	N	121.14	0.038	89	GLN	N
870	LEU	H	H	8.38	0.008	90	LEU	H
871	LEU	HA	H	4.111	0.003	90	LEU	HA
872	LEU	HB2	H	2.22	0.008	90	LEU	HB2
873	LEU	HB3	H	2.22	0.008	90	LEU	HB3
874	LEU	HG	H	1.411	0.006	90	LEU	HG
875	LEU	HD11	H	0.891	0.004	90	LEU	HD11
876	LEU	HD12	H	0.891	0.004	90	LEU	HD12
877	LEU	HD13	H	0.891	0.004	90	LEU	HD13
878	LEU	HD21	H	1.09	0.001	90	LEU	HD21
879	LEU	HD22	H	1.09	0.001	90	LEU	HD22
880	LEU	HD23	H	1.09	0.001	90	LEU	HD23
881	LEU	C	C	178.822	0	90	LEU	C
882	LEU	CA	C	59.179	0.171	90	LEU	CA
883	LEU	CB	C	42.224	0.051	90	LEU	CB
884	LEU	CG	C	26.652	0	90	LEU	CG
885	LEU	CD1	C	24.366	0	90	LEU	CD1
886	LEU	N	N	118.187	0.071	90	LEU	N
887	VAL	H	H	7.155	0.015	91	VAL	H
888	VAL	HA	H	3.197	0.004	91	VAL	HA
889	VAL	HB	H	2.051	0.008	91	VAL	HB
890	VAL	HG11	H	0.985	0.007	91	VAL	HG11
891	VAL	HG12	H	0.985	0.007	91	VAL	HG12
892	VAL	HG13	H	0.985	0.007	91	VAL	HG13
893	VAL	HG21	H	0.33	0.006	91	VAL	HG21
894	VAL	HG22	H	0.33	0.006	91	VAL	HG22
895	VAL	HG23	H	0.33	0.006	91	VAL	HG23
896	VAL	CA	C	66.319	0.071	91	VAL	CA
897	VAL	CB	C	30.962	0.173	91	VAL	CB
898	VAL	CG1	C	21.712	0.92	91	VAL	CG1
899	VAL	CG2	C	20.678	0	91	VAL	CG2
900	VAL	N	N	120.766	0	91	VAL	N
901	ASP	H	H	8.509	0.015	92	ASP	H
902	ASP	HA	H	4.205	0.006	92	ASP	HA
903	ASP	HB2	H	3.066	0.045	92	ASP	HB2
904	ASP	HB3	H	2.659	0.006	92	ASP	HB3
905	ASP	C	C	180.097	0.008	92	ASP	C

906	**ASP**	**CA**	**C**	**57.959**	**0.076**	**92**	**ASP**	**CA**
907	**ASP**	**CB**	**C**	**39.775**	**0.064**	**92**	**ASP**	**CB**
908	**ASP**	**N**	**N**	**120.76**	**0.037**	**92**	**ASP**	**N**
909	**ILE**	**H**	**H**	**8.384**	**0.011**	**93**	**ILE**	**H**
910	**ILE**	**HA**	**H**	**3.772**	**0.006**	**93**	**ILE**	**HA**
911	**ILE**	**HB**	**H**	**2.214**	**0.002**	**93**	**ILE**	**HB**
912	**ILE**	**HG21**	**H**	**1.019**	**0.011**	**93**	**ILE**	**HG21**
913	**ILE**	**HG22**	**H**	**1.019**	**0.011**	**93**	**ILE**	**HG22**
914	**ILE**	**HG23**	**H**	**1.019**	**0.011**	**93**	**ILE**	**HG23**
915	**ILE**	**HD11**	**H**	**0.701**	**0.007**	**93**	**ILE**	**HD11**
916	**ILE**	**HD12**	**H**	**0.701**	**0.007**	**93**	**ILE**	**HD12**
917	**ILE**	**HD13**	**H**	**0.701**	**0.007**	**93**	**ILE**	**HD13**
918	**ILE**	**C**	**C**	**175.486**	**0**	**93**	**ILE**	**C**
919	**ILE**	**CA**	**C**	**65.514**	**0.087**	**93**	**ILE**	**CA**
920	**ILE**	**CB**	**C**	**38.804**	**0.106**	**93**	**ILE**	**CB**
921	**ILE**	**CG1**	**C**	**28.971**	**0**	**93**	**ILE**	**CG1**
922	**ILE**	**CG2**	**C**	**18.72**	**0**	**93**	**ILE**	**CG2**
923	**ILE**	**CD1**	**C**	**14.641**	**0.09**	**93**	**ILE**	**CD1**
924	**ILE**	**N**	**N**	**121.202**	**0.047**	**93**	**ILE**	**N**
925	**VAL**	**H**	**H**	**7.923**	**0.009**	**94**	**VAL**	**H**
926	**VAL**	**HA**	**H**	**3.292**	**0.002**	**94**	**VAL**	**HA**
927	**VAL**	**HB**	**H**	**2.007**	**0.007**	**94**	**VAL**	**HB**
928	**VAL**	**HG11**	**H**	**0.902**	**0.003**	**94**	**VAL**	**HG11**
929	**VAL**	**HG12**	**H**	**0.902**	**0.003**	**94**	**VAL**	**HG12**
930	**VAL**	**HG13**	**H**	**0.902**	**0.003**	**94**	**VAL**	**HG13**
931	**VAL**	**HG21**	**H**	**0.798**	**0.002**	**94**	**VAL**	**HG21**
932	**VAL**	**HG22**	**H**	**0.798**	**0.002**	**94**	**VAL**	**HG22**
933	**VAL**	**HG23**	**H**	**0.798**	**0.002**	**94**	**VAL**	**HG23**
934	**VAL**	**CA**	**C**	**68.56**	**0.057**	**94**	**VAL**	**CA**
935	**VAL**	**CB**	**C**	**30.907**	**0.051**	**94**	**VAL**	**CB**
936	**VAL**	**CG1**	**C**	**23.221**	**0.638**	**94**	**VAL**	**CG1**
937	**VAL**	**N**	**N**	**120.717**	**0.04**	**94**	**VAL**	**N**
938	**HIS**	**H**	**H**	**9.233**	**0.014**	**95**	**HIS**	**H**
939	**HIS**	**HA**	**H**	**4.564**	**0.007**	**95**	**HIS**	**HA**
940	**HIS**	**HB2**	**H**	**3.215**	**0.006**	**95**	**HIS**	**HB2**
941	**HIS**	**HB3**	**H**	**3.136**	**0.008**	**95**	**HIS**	**HB3**
942	**HIS**	**C**	**C**	**178.106**	**0.011**	**95**	**HIS**	**C**
943	**HIS**	**CA**	**C**	**57.676**	**0.054**	**95**	**HIS**	**CA**
944	**HIS**	**CB**	**C**	**29.791**	**0.039**	**95**	**HIS**	**CB**
945	**HIS**	**N**	**N**	**118.137**	**0.077**	**95**	**HIS**	**N**
946	**ALA**	**H**	**H**	**8.24**	**0.008**	**96**	**ALA**	**H**
947	**ALA**	**HA**	**H**	**4.193**	**0.013**	**96**	**ALA**	**HA**
948	**ALA**	**HB1**	**H**	**1.536**	**0.008**	**96**	**ALA**	**HB1**
949	**ALA**	**HB2**	**H**	**1.536**	**0.008**	**96**	**ALA**	**HB2**
950	**ALA**	**HB3**	**H**	**1.536**	**0.008**	**96**	**ALA**	**HB3**
951	**ALA**	**C**	**C**	**181.877**	**0.003**	**96**	**ALA**	**C**
952	**ALA**	**CA**	**C**	**55.376**	**0.136**	**96**	**ALA**	**CA**
953	**ALA**	**CB**	**C**	**17.55**	**0.08**	**96**	**ALA**	**CB**
954	**ALA**	**N**	**N**	**122.289**	**0.075**	**96**	**ALA**	**N**
955	**CYS**	**H**	**H**	**8.49**	**0.009**	**97**	**CYS**	**H**
956	**CYS**	**HA**	**H**	**4.647**	**0.008**	**97**	**CYS**	**HA**
957	**CYS**	**HB2**	**H**	**3.584**	**0.006**	**97**	**CYS**	**HB2**
958	**CYS**	**HB3**	**H**	**2.913**	**0.005**	**97**	**CYS**	**HB3**
959	**CYS**	**C**	**C**	**176.485**	**0.005**	**97**	**CYS**	**C**
960	**CYS**	**CA**	**C**	**57.873**	**0.087**	**97**	**CYS**	**CA**
961	**CYS**	**CB**	**C**	**39.818**	**0.027**	**97**	**CYS**	**CB**
962	**CYS**	**N**	**N**	**120.441**	**0.084**	**97**	**CYS**	**N**
963	**GLU**	**H**	**H**	**8.932**	**0.009**	**98**	**GLU**	**H**
964	**GLU**	**HA**	**H**	**3.733**	**0.008**	**98**	**GLU**	**HA**
965	**GLU**	**HB2**	**H**	**2.428**	**0**	**98**	**GLU**	**HB2**
966	**GLU**	**HB3**	**H**	**1.967**	**0**	**98**	**GLU**	**HB3**
967	**GLU**	**HG2**	**H**	**2.603**	**0.002**	**98**	**GLU**	**HG2**

968	**GLU**	**HG3**	**H**	**2.436**	**0.004**	**98**	**GLU**	**HG3**
969	**GLU**	**C**	**C**	**177.99**	**0.011**	**98**	**GLU**	**C**
970	**GLU**	**CA**	**C**	**59.477**	**0.071**	**98**	**GLU**	**CA**
971	**GLU**	**CB**	**C**	**29.577**	**0.065**	**98**	**GLU**	**CB**
972	**GLU**	**CG**	**C**	**36.499**	**0**	**98**	**GLU**	**CG**
973	**GLU**	**N**	**N**	**120.288**	**0.08**	**98**	**GLU**	**N**
974	**LYS**	**H**	**H**	**7.132**	**0.009**	**99**	**LYS**	**H**
975	**LYS**	**HA**	**H**	**4.276**	**0.005**	**99**	**LYS**	**HA**
976	**LYS**	**HB2**	**H**	**1.874**	**0.04**	**99**	**LYS**	**HB2**
977	**LYS**	**HB3**	**H**	**1.777**	**0.027**	**99**	**LYS**	**HB3**
978	**LYS**	**HG2**	**H**	**1.418**	**0.004**	**99**	**LYS**	**HG2**
979	**LYS**	**HG3**	**H**	**1.377**	**0.008**	**99**	**LYS**	**HG3**
980	**LYS**	**HD2**	**H**	**1.709**	**0.03**	**99**	**LYS**	**HD2**
981	**LYS**	**HD3**	**H**	**1.616**	**0.003**	**99**	**LYS**	**HD3**
982	**LYS**	**HE2**	**H**	**2.81**	**0**	**99**	**LYS**	**HE2**
983	**LYS**	**HE3**	**H**	**2.712**	**0**	**99**	**LYS**	**HE3**
984	**LYS**	**C**	**C**	**177.273**	**0.008**	**99**	**LYS**	**C**
985	**LYS**	**CA**	**C**	**56.369**	**0.258**	**99**	**LYS**	**CA**
986	**LYS**	**CB**	**C**	**33.073**	**0.041**	**99**	**LYS**	**CB**
987	**LYS**	**CG**	**C**	**24.811**	**0**	**99**	**LYS**	**CG**
988	**LYS**	**CD**	**C**	**28.602**	**0**	**99**	**LYS**	**CD**
989	**LYS**	**CE**	**C**	**42.197**	**0**	**99**	**LYS**	**CE**
990	**LYS**	**N**	**N**	**113.815**	**0.08**	**99**	**LYS**	**N**
991	**SER**	**H**	**H**	**7.916**	**0.01**	**100**	**SER**	**H**
992	**SER**	**HA**	**H**	**4.423**	**0.007**	**100**	**SER**	**HA**
993	**SER**	**HB2**	**H**	**3.941**	**0.005**	**100**	**SER**	**HB2**
994	**SER**	**HB3**	**H**	**3.792**	**0.007**	**100**	**SER**	**HB3**
995	**SER**	**C**	**C**	**174.886**	**0.022**	**100**	**SER**	**C**
996	**SER**	**CA**	**C**	**59.308**	**0.067**	**100**	**SER**	**CA**
997	**SER**	**CB**	**C**	**64.216**	**0.123**	**100**	**SER**	**CB**
998	**SER**	**N**	**N**	**114.552**	**0.072**	**100**	**SER**	**N**
999	**VAL**	**H**	**H**	**7.559**	**0.014**	**101**	**VAL**	**H**
1000	**VAL**	**HA**	**H**	**4.292**	**0.004**	**101**	**VAL**	**HA**
1001	**VAL**	**HB**	**H**	**2.014**	**0.01**	**101**	**VAL**	**HB**
1002	**VAL**	**HG11**	**H**	**0.963**	**0.004**	**101**	**VAL**	**HG11**
1003	**VAL**	**HG12**	**H**	**0.963**	**0.004**	**101**	**VAL**	**HG12**
1004	**VAL**	**HG13**	**H**	**0.963**	**0.004**	**101**	**VAL**	**HG13**
1005	**VAL**	**HG21**	**H**	**0.897**	**0.009**	**101**	**VAL**	**HG21**
1006	**VAL**	**HG22**	**H**	**0.897**	**0.009**	**101**	**VAL**	**HG22**
1007	**VAL**	**HG23**	**H**	**0.897**	**0.009**	**101**	**VAL**	**HG23**
1008	**VAL**	**CA**	**C**	**60.282**	**0.06**	**101**	**VAL**	**CA**
1009	**VAL**	**CB**	**C**	**31.716**	**0.236**	**101**	**VAL**	**CB**
1010	**VAL**	**CG1**	**C**	**21.226**	**0**	**101**	**VAL**	**CG1**
1011	**VAL**	**N**	**N**	**124.015**	**0.121**	**101**	**VAL**	**N**
1012	**PRO**	**HA**	**H**	**4.543**	**0.002**	**102**	**PRO**	**HA**
1013	**PRO**	**HB2**	**H**	**2.27**	**0.002**	**102**	**PRO**	**HB2**
1014	**PRO**	**HB3**	**H**	**2.042**	**0**	**102**	**PRO**	**HB3**
1015	**PRO**	**HG2**	**H**	**1.935**	**0**	**102**	**PRO**	**HG2**
1016	**PRO**	**HG3**	**H**	**1.766**	**0.002**	**102**	**PRO**	**HG3**
1017	**PRO**	**HD2**	**H**	**3.897**	**0.009**	**102**	**PRO**	**HD2**
1018	**PRO**	**HD3**	**H**	**3.555**	**0.003**	**102**	**PRO**	**HD3**
1019	**PRO**	**HA**	**H**	**4.236**	**0.001**	**103**	**PRO**	**HA**
1020	**PRO**	**HB2**	**H**	**2.19**	**0.004**	**103**	**PRO**	**HB2**
1021	**PRO**	**HB3**	**H**	**1.958**	**0.007**	**103**	**PRO**	**HB3**
1022	**PRO**	**HD2**	**H**	**3.765**	**0.003**	**103**	**PRO**	**HD2**
1023	**PRO**	**HD3**	**H**	**3.553**	**0.007**	**103**	**PRO**	**HD3**
1024	**PRO**	**CA**	**C**	**63.289**	**0.026**	**103**	**PRO**	**CA**
1025	**PRO**	**CB**	**C**	**32.052**	**0.023**	**103**	**PRO**	**CB**
1026	**PRO**	**CG**	**C**	**27.429**	**0.143**	**103**	**PRO**	**CG**
1027	**PRO**	**CD**	**C**	**50.477**	**0.09**	**103**	**PRO**	**CD**
1028	**ASN**	**H**	**H**	**7.87**	**0.009**	**104**	**ASN**	**H**
1029	**ASN**	**HA**	**H**	**4.602**	**0.007**	**104**	**ASN**	**HA**

1030	**ASN**	**HB2**	**H**	**2.538**	**0.006**	**104**	**ASN**	**HB2**
1031	**ASN**	**HB3**	**H**	**2.438**	**0.006**	**104**	**ASN**	**HB3**
1032	**ASN**	**C**	**C**	**174.268**	**0**	**104**	**ASN**	**C**
1033	**ASN**	**CA**	**C**	**52.87**	**0.113**	**104**	**ASN**	**CA**
1034	**ASN**	**CB**	**C**	**42.6**	**0.086**	**104**	**ASN**	**CB**
1035	**ASN**	**N**	**N**	**122.117**	**0.072**	**104**	**ASN**	**N**
1036	**GLU**	**H**	**H**	**8.731**	**0.008**	**105**	**GLU**	**H**
1037	**GLU**	**HA**	**H**	**4.015**	**0.005**	**105**	**GLU**	**HA**
1038	**GLU**	**HB2**	**H**	**2.018**	**0.006**	**105**	**GLU**	**HB2**
1039	**GLU**	**HB3**	**H**	**1.908**	**0.005**	**105**	**GLU**	**HB3**
1040	**GLU**	**HG2**	**H**	**2.275**	**0**	**105**	**GLU**	**HG2**
1041	**GLU**	**HG3**	**H**	**2.203**	**0**	**105**	**GLU**	**HG3**
1042	**GLU**	**C**	**C**	**175.749**	**0.004**	**105**	**GLU**	**C**
1043	**GLU**	**CA**	**C**	**58.027**	**0.092**	**105**	**GLU**	**CA**
1044	**GLU**	**CB**	**C**	**29.701**	**0.074**	**105**	**GLU**	**CB**
1045	**GLU**	**CG**	**C**	**36.3**	**0.003**	**105**	**GLU**	**CG**
1046	**GLU**	**N**	**N**	**124.863**	**0.084**	**105**	**GLU**	**N**
1047	**ASP**	**H**	**H**	**8.719**	**0.013**	**106**	**ASP**	**H**
1048	**ASP**	**HA**	**H**	**4.653**	**0.007**	**106**	**ASP**	**HA**
1049	**ASP**	**HB2**	**H**	**2.931**	**0.006**	**106**	**ASP**	**HB2**
1050	**ASP**	**HB3**	**H**	**2.454**	**0.005**	**106**	**ASP**	**HB3**
1051	**ASP**	**C**	**C**	**176.332**	**0.008**	**106**	**ASP**	**C**
1052	**ASP**	**CA**	**C**	**52.744**	**0.1**	**106**	**ASP**	**CA**
1053	**ASP**	**CB**	**C**	**41.743**	**0.101**	**106**	**ASP**	**CB**
1054	**ASP**	**N**	**N**	**119.868**	**0.071**	**106**	**ASP**	**N**
1055	**ASN**	**H**	**H**	**8.79**	**0.008**	**107**	**ASN**	**H**
1056	**ASN**	**HA**	**H**	**4.271**	**0.003**	**107**	**ASN**	**HA**
1057	**ASN**	**HB2**	**H**	**2.798**	**0.002**	**107**	**ASN**	**HB2**
1058	**ASN**	**HB3**	**H**	**2.721**	**0.002**	**107**	**ASN**	**HB3**
1059	**ASN**	**C**	**C**	**178.115**	**0.016**	**107**	**ASN**	**C**
1060	**ASN**	**CA**	**C**	**56.051**	**0.047**	**107**	**ASN**	**CA**
1061	**ASN**	**CB**	**C**	**37.241**	**0.13**	**107**	**ASN**	**CB**
1062	**ASN**	**N**	**N**	**125.771**	**0.068**	**107**	**ASN**	**N**
1063	**CYS**	**H**	**H**	**8.293**	**0.008**	**108**	**CYS**	**H**
1064	**CYS**	**HA**	**H**	**4.028**	**0.002**	**108**	**CYS**	**HA**
1065	**CYS**	**HB2**	**H**	**3.726**	**0.003**	**108**	**CYS**	**HB2**
1066	**CYS**	**HB3**	**H**	**3.072**	**0.002**	**108**	**CYS**	**HB3**
1067	**CYS**	**C**	**C**	**176.518**	**0.021**	**108**	**CYS**	**C**
1068	**CYS**	**CA**	**C**	**61.008**	**0.077**	**108**	**CYS**	**CA**
1069	**CYS**	**CB**	**C**	**43.816**	**0.277**	**108**	**CYS**	**CB**
1070	**CYS**	**N**	**N**	**119.079**	**0.056**	**108**	**CYS**	**N**
1071	**LEU**	**H**	**H**	**7.526**	**0.007**	**109**	**LEU**	**H**
1072	**LEU**	**HA**	**H**	**4.065**	**0.006**	**109**	**LEU**	**HA**
1073	**LEU**	**HB2**	**H**	**1.782**	**0.011**	**109**	**LEU**	**HB2**
1074	**LEU**	**HB3**	**H**	**1.402**	**0.052**	**109**	**LEU**	**HB3**
1075	**LEU**	**HG**	**H**	**1.386**	**0**	**109**	**LEU**	**HG**
1076	**LEU**	**HD11**	**H**	**0.945**	**0.002**	**109**	**LEU**	**HD11**
1077	**LEU**	**HD12**	**H**	**0.945**	**0.002**	**109**	**LEU**	**HD12**
1078	**LEU**	**HD13**	**H**	**0.945**	**0.002**	**109**	**LEU**	**HD13**
1079	**LEU**	**HD21**	**H**	**0.755**	**0.004**	**109**	**LEU**	**HD21**
1080	**LEU**	**HD22**	**H**	**0.755**	**0.004**	**109**	**LEU**	**HD22**
1081	**LEU**	**HD23**	**H**	**0.755**	**0.004**	**109**	**LEU**	**HD23**
1082	**LEU**	**C**	**C**	**178.762**	**0**	**109**	**LEU**	**C**
1083	**LEU**	**CA**	**C**	**57.244**	**0.053**	**109**	**LEU**	**CA**
1084	**LEU**	**CB**	**C**	**40.887**	**0.196**	**109**	**LEU**	**CB**
1085	**LEU**	**CG**	**C**	**27.246**	**0**	**109**	**LEU**	**CG**
1086	**LEU**	**CD1**	**C**	**25.139**	**0**	**109**	**LEU**	**CD1**
1087	**LEU**	**CD2**	**C**	**22.295**	**0**	**109**	**LEU**	**CD2**
1088	**LEU**	**N**	**N**	**120.057**	**0.097**	**109**	**LEU**	**N**
1089	**MET**	**H**	**H**	**8.232**	**0.01**	**110**	**MET**	**H**
1090	**MET**	**HA**	**H**	**4.295**	**0.006**	**110**	**MET**	**HA**
1091	**MET**	**HB2**	**H**	**2.19**	**0.002**	**110**	**MET**	**HB2**

1092	**MET**	**HB3**	**H**	**2.096**	**0.005**	**110**	**MET**	**HB3**
1093	**MET**	**HG2**	**H**	**2.662**	**0.004**	**110**	**MET**	**HG2**
1094	**MET**	**HG3**	**H**	**2.562**	**0.006**	**110**	**MET**	**HG3**
1095	**MET**	**C**	**C**	**177.577**	**0.012**	**110**	**MET**	**C**
1096	**MET**	**CA**	**C**	**58.331**	**0.107**	**110**	**MET**	**CA**
1097	**MET**	**CB**	**C**	**31.594**	**0.138**	**110**	**MET**	**CB**
1098	**MET**	**N**	**N**	**121.688**	**0.064**	**110**	**MET**	**N**
1099	**ALA**	**H**	**H**	**7.822**	**0.018**	**111**	**ALA**	**H**
1100	**ALA**	**HA**	**H**	**3.994**	**0.046**	**111**	**ALA**	**HA**
1101	**ALA**	**HB1**	**H**	**1.398**	**0.002**	**111**	**ALA**	**HB1**
1102	**ALA**	**HB2**	**H**	**1.398**	**0.002**	**111**	**ALA**	**HB2**
1103	**ALA**	**HB3**	**H**	**1.398**	**0.002**	**111**	**ALA**	**HB3**
1104	**ALA**	**C**	**C**	**180.701**	**0.014**	**111**	**ALA**	**C**
1105	**ALA**	**CA**	**C**	**55.462**	**0.084**	**111**	**ALA**	**CA**
1106	**ALA**	**CB**	**C**	**18.354**	**0.066**	**111**	**ALA**	**CB**
1107	**ALA**	**N**	**N**	**120.45**	**0.337**	**111**	**ALA**	**N**
1108	**LEU**	**H**	**H**	**7.808**	**0.011**	**112**	**LEU**	**H**
1109	**LEU**	**HA**	**H**	**3.853**	**0.007**	**112**	**LEU**	**HA**
1110	**LEU**	**HB2**	**H**	**1.861**	**0.003**	**112**	**LEU**	**HB2**
1111	**LEU**	**HB3**	**H**	**1.516**	**0.003**	**112**	**LEU**	**HB3**
1112	**LEU**	**HG**	**H**	**1.541**	**0**	**112**	**LEU**	**HG**
1113	**LEU**	**HD11**	**H**	**0.633**	**0**	**112**	**LEU**	**HD11**
1114	**LEU**	**HD12**	**H**	**0.633**	**0**	**112**	**LEU**	**HD12**
1115	**LEU**	**HD13**	**H**	**0.633**	**0**	**112**	**LEU**	**HD13**
1116	**LEU**	**HD21**	**H**	**0.578**	**0**	**112**	**LEU**	**HD21**
1117	**LEU**	**HD22**	**H**	**0.578**	**0**	**112**	**LEU**	**HD22**
1118	**LEU**	**HD23**	**H**	**0.578**	**0**	**112**	**LEU**	**HD23**
1119	**LEU**	**CA**	**C**	**58.543**	**0.029**	**112**	**LEU**	**CA**
1120	**LEU**	**CB**	**C**	**42.758**	**0.117**	**112**	**LEU**	**CB**
1121	**LEU**	**CD1**	**C**	**24.412**	**0**	**112**	**LEU**	**CD1**
1122	**LEU**	**N**	**N**	**119.645**	**0.106**	**112**	**LEU**	**N**
1123	**GLY**	**H**	**H**	**7.963**	**0.009**	**113**	**GLY**	**H**
1124	**GLY**	**HA2**	**H**	**3.881**	**0.012**	**113**	**GLY**	**HA2**
1125	**GLY**	**HA3**	**H**	**3.811**	**0.009**	**113**	**GLY**	**HA3**
1126	**GLY**	**C**	**C**	**177.811**	**0.009**	**113**	**GLY**	**C**
1127	**GLY**	**CA**	**C**	**47.852**	**1.007**	**113**	**GLY**	**CA**
1128	**GLY**	**N**	**N**	**106.691**	**0.075**	**113**	**GLY**	**N**
1129	**ILE**	**H**	**H**	**8.922**	**0.008**	**114**	**ILE**	**H**
1130	**ILE**	**HA**	**H**	**3.539**	**0.006**	**114**	**ILE**	**HA**
1131	**ILE**	**HB**	**H**	**1.763**	**0.002**	**114**	**ILE**	**HB**
1132	**ILE**	**HG21**	**H**	**0.879**	**0.001**	**114**	**ILE**	**HG21**
1133	**ILE**	**HG22**	**H**	**0.879**	**0.001**	**114**	**ILE**	**HG22**
1134	**ILE**	**HG23**	**H**	**0.879**	**0.001**	**114**	**ILE**	**HG23**
1135	**ILE**	**CA**	**C**	**66.179**	**0.076**	**114**	**ILE**	**CA**
1136	**ILE**	**CB**	**C**	**38.192**	**0.029**	**114**	**ILE**	**CB**
1137	**ILE**	**CG1**	**C**	**30.113**	**0**	**114**	**ILE**	**CG1**
1138	**ILE**	**CG2**	**C**	**17.963**	**0**	**114**	**ILE**	**CG2**
1139	**ILE**	**CD1**	**C**	**14.539**	**0**	**114**	**ILE**	**CD1**
1140	**ILE**	**N**	**N**	**124.466**	**0.077**	**114**	**ILE**	**N**
1141	**SER**	**H**	**H**	**8.621**	**0.008**	**115**	**SER**	**H**
1142	**SER**	**HA**	**H**	**4.149**	**0.059**	**115**	**SER**	**HA**
1143	**SER**	**C**	**C**	**176.226**	**0.011**	**115**	**SER**	**C**
1144	**SER**	**CA**	**C**	**62.881**	**0.638**	**115**	**SER**	**CA**
1145	**SER**	**CB**	**C**	**63.767**	**0.091**	**115**	**SER**	**CB**
1146	**SER**	**N**	**N**	**117.184**	**0.051**	**115**	**SER**	**N**
1147	**MET**	**H**	**H**	**8.086**	**0.007**	**116**	**MET**	**H**
1148	**MET**	**HA**	**H**	**4.628**	**0.005**	**116**	**MET**	**HA**
1149	**MET**	**HB2**	**H**	**2.226**	**0.006**	**116**	**MET**	**HB2**
1150	**MET**	**HB3**	**H**	**2.03**	**0.006**	**116**	**MET**	**HB3**
1151	**MET**	**HG2**	**H**	**2.873**	**0.01**	**116**	**MET**	**HG2**
1152	**MET**	**HG3**	**H**	**2.536**	**0.007**	**116**	**MET**	**HG3**
1153	**MET**	**HE1**	**H**	**0.768**	**0**	**116**	**MET**	**HE1**

1154	**MET**	**HE2**	**H**	**0.768**	**0**	**116**	**MET**	**HE2**
1155	**MET**	**HE3**	**H**	**0.768**	**0**	**116**	**MET**	**HE3**
1156	**MET**	**C**	**C**	**178.502**	**0.004**	**116**	**MET**	**C**
1157	**MET**	**CA**	**C**	**56.375**	**0.049**	**116**	**MET**	**CA**
1158	**MET**	**CB**	**C**	**31.861**	**0.087**	**116**	**MET**	**CB**
1159	**MET**	**CG**	**C**	**31.298**	**0**	**116**	**MET**	**CG**
1160	**MET**	**N**	**N**	**121.551**	**0.063**	**116**	**MET**	**N**
1161	**CYS**	**H**	**H**	**7.787**	**0.009**	**117**	**CYS**	**H**
1162	**CYS**	**HA**	**H**	**4.345**	**0.001**	**117**	**CYS**	**HA**
1163	**CYS**	**HB2**	**H**	**3.409**	**0.002**	**117**	**CYS**	**HB2**
1164	**CYS**	**HB3**	**H**	**3.201**	**0.002**	**117**	**CYS**	**HB3**
1165	**CYS**	**C**	**C**	**175.458**	**0.006**	**117**	**CYS**	**C**
1166	**CYS**	**CA**	**C**	**59.93**	**0.143**	**117**	**CYS**	**CA**
1167	**CYS**	**CB**	**C**	**37.821**	**0.16**	**117**	**CYS**	**CB**
1168	**CYS**	**N**	**N**	**123.898**	**0.07**	**117**	**CYS**	**N**
1169	**PHE**	**H**	**H**	**8.771**	**0.014**	**118**	**PHE**	**H**
1170	**PHE**	**HA**	**H**	**4.062**	**0**	**118**	**PHE**	**HA**
1171	**PHE**	**HB2**	**H**	**3.487**	**0.002**	**118**	**PHE**	**HB2**
1172	**PHE**	**HB3**	**H**	**2.975**	**0.003**	**118**	**PHE**	**HB3**
1173	**PHE**	**C**	**C**	**175.495**	**0**	**118**	**PHE**	**C**
1174	**PHE**	**CA**	**C**	**56.093**	**0.019**	**118**	**PHE**	**CA**
1175	**PHE**	**CB**	**C**	**39.697**	**0.062**	**118**	**PHE**	**CB**
1176	**PHE**	**N**	**N**	**123.601**	**0.048**	**118**	**PHE**	**N**
1177	**LYS**	**H**	**H**	**8.704**	**0.397**	**119**	**LYS**	**H**
1178	**LYS**	**HA**	**H**	**3.188**	**0**	**119**	**LYS**	**HA**
1179	**LYS**	**C**	**C**	**177.343**	**0.651**	**119**	**LYS**	**C**
1180	**LYS**	**CA**	**C**	**60.019**	**0.009**	**119**	**LYS**	**CA**
1181	**LYS**	**CB**	**C**	**33.734**	**0.002**	**119**	**LYS**	**CB**
1182	**LYS**	**CG**	**C**	**25.86**	**0**	**119**	**LYS**	**CG**
1183	**LYS**	**CD**	**C**	**25.861**	**0**	**119**	**LYS**	**CD**
1184	**LYS**	**CE**	**C**	**43.226**	**0**	**119**	**LYS**	**CE**
1185	**LYS**	**N**	**N**	**123.399**	**0.063**	**119**	**LYS**	**N**
1186	**THR**	**H**	**H**	**7.926**	**0.043**	**120**	**THR**	**H**
1187	**THR**	**HA**	**H**	**3.902**	**0.005**	**120**	**THR**	**HA**
1188	**THR**	**HB**	**H**	**4.411**	**0.007**	**120**	**THR**	**HB**
1189	**THR**	**HG1**	**H**	**5.396**	**0.015**	**120**	**THR**	**HG1**
1190	**THR**	**HG21**	**H**	**1.2**	**0.007**	**120**	**THR**	**HG21**
1191	**THR**	**HG22**	**H**	**1.2**	**0.007**	**120**	**THR**	**HG22**
1192	**THR**	**HG23**	**H**	**1.2**	**0.007**	**120**	**THR**	**HG23**
1193	**THR**	**C**	**C**	**176.959**	**0.005**	**120**	**THR**	**C**
1194	**THR**	**CA**	**C**	**67.036**	**0.08**	**120**	**THR**	**CA**
1195	**THR**	**CB**	**C**	**69.101**	**0.124**	**120**	**THR**	**CB**
1196	**THR**	**CG2**	**C**	**21.512**	**0.1**	**120**	**THR**	**CG2**
1197	**THR**	**N**	**N**	**114**	**0.092**	**120**	**THR**	**N**
1198	**GLU**	**H**	**H**	**7.723**	**0.009**	**121**	**GLU**	**H**
1199	**GLU**	**HA**	**H**	**3.984**	**0.004**	**121**	**GLU**	**HA**
1200	**GLU**	**HB2**	**H**	**2.019**	**0**	**121**	**GLU**	**HB2**
1201	**GLU**	**HB3**	**H**	**1.922**	**0**	**121**	**GLU**	**HB3**
1202	**GLU**	**HG2**	**H**	**2.435**	**0**	**121**	**GLU**	**HG2**
1203	**GLU**	**HG3**	**H**	**2.363**	**0**	**121**	**GLU**	**HG3**
1204	**GLU**	**C**	**C**	**178.624**	**0.013**	**121**	**GLU**	**C**
1205	**GLU**	**CA**	**C**	**58.583**	**0.121**	**121**	**GLU**	**CA**
1206	**GLU**	**CB**	**C**	**29.338**	**0.078**	**121**	**GLU**	**CB**
1207	**GLU**	**CG**	**C**	**35.104**	**0.017**	**121**	**GLU**	**CG**
1208	**GLU**	**N**	**N**	**120.758**	**0.089**	**121**	**GLU**	**N**
1209	**ILE**	**H**	**H**	**8.261**	**0.011**	**122**	**ILE**	**H**
1210	**ILE**	**HA**	**H**	**3.575**	**0.007**	**122**	**ILE**	**HA**
1211	**ILE**	**HB**	**H**	**2.229**	**0.006**	**122**	**ILE**	**HB**
1212	**ILE**	**HG21**	**H**	**0.494**	**0.01**	**122**	**ILE**	**HG21**
1213	**ILE**	**HG22**	**H**	**0.494**	**0.01**	**122**	**ILE**	**HG22**
1214	**ILE**	**HG23**	**H**	**0.494**	**0.01**	**122**	**ILE**	**HG23**
1215	**ILE**	**C**	**C**	**176.091**	**0**	**122**	**ILE**	**C**

1216	ILE	CA	C	61.617	0.084	122	ILE	CA
1217	ILE	CB	C	34.197	0.079	122	ILE	CB
1218	ILE	CG1	C	26.139	0	122	ILE	CG1
1219	ILE	CG2	C	18.754	0	122	ILE	CG2
1220	ILE	CD1	C	9.303	0	122	ILE	CD1
1221	ILE	N	N	120.66	0.073	122	ILE	N
1222	HIS	H	H	8.536	0.007	123	HIS	H
1223	HIS	HA	H	4.26	0.004	123	HIS	HA
1224	HIS	HB2	H	3.239	0.007	123	HIS	HB2
1225	HIS	HB3	H	3.239	0.007	123	HIS	HB3
1226	HIS	C	C	179.753	0.01	123	HIS	C
1227	HIS	CA	C	60.269	1.57	123	HIS	CA
1228	HIS	CB	C	28.615	0.075	123	HIS	CB
1229	HIS	N	N	119.537	0.05	123	HIS	N
1230	LYS	H	H	7.786	0.205	124	LYS	H
1231	LYS	HA	H	3.893	0.013	124	LYS	HA
1232	LYS	HB2	H	2.109	0	124	LYS	HB2
1233	LYS	HB3	H	2.109	0	124	LYS	HB3
1234	LYS	HG2	H	1.591	0	124	LYS	HG2
1235	LYS	HG3	H	1.591	0	124	LYS	HG3
1236	LYS	HE2	H	2.73	0.001	124	LYS	HE2
1237	LYS	HE3	H	2.73	0.001	124	LYS	HE3
1238	LYS	C	C	178.18	0.023	124	LYS	C
1239	LYS	CA	C	58.23	1.983	124	LYS	CA
1240	LYS	CB	C	32.442	0.105	124	LYS	CB
1241	LYS	CG	C	25.31	0	124	LYS	CG
1242	LYS	CD	C	29.613	0	124	LYS	CD
1243	LYS	CE	C	41.847	0	124	LYS	CE
1244	LYS	N	N	121.222	0.968	124	LYS	N
1245	LEU	H	H	6.998	0.011	125	LEU	H
1246	LEU	HA	H	3.694	0.009	125	LEU	HA
1247	LEU	HB2	H	1.133	0.006	125	LEU	HB2
1248	LEU	HB3	H	0.851	0.144	125	LEU	HB3
1249	LEU	HG	H	0.248	0.21	125	LEU	HG
1250	LEU	HD11	H	-0.157	0.001	125	LEU	HD11
1251	LEU	HD12	H	-0.157	0.001	125	LEU	HD12
1252	LEU	HD13	H	-0.157	0.001	125	LEU	HD13
1253	LEU	HD21	H	-0.252	0.004	125	LEU	HD21
1254	LEU	HD22	H	-0.252	0.004	125	LEU	HD22
1255	LEU	HD23	H	-0.252	0.004	125	LEU	HD23
1256	LEU	C	C	175.175	0	125	LEU	C
1257	LEU	CA	C	54.946	0.436	125	LEU	CA
1258	LEU	CB	C	41.566	0.035	125	LEU	CB
1259	LEU	CG	C	25.897	0	125	LEU	CG
1260	LEU	CD1	C	24.752	0.077	125	LEU	CD1
1261	LEU	CD2	C	21.827	0	125	LEU	CD2
1262	LEU	N	N	118.057	0.057	125	LEU	N
1263	ASN	H	H	7.865	0.01	126	ASN	H
1264	ASN	HA	H	4.556	0.007	126	ASN	HA
1265	ASN	HB2	H	3.252	0.005	126	ASN	HB2
1266	ASN	HB3	H	2.783	0.005	126	ASN	HB3
1267	ASN	C	C	176.201	0.002	126	ASN	C
1268	ASN	CA	C	54.233	0.082	126	ASN	CA
1269	ASN	CB	C	37.154	0.055	126	ASN	CB
1270	ASN	N	N	112.609	0.061	126	ASN	N
1271	TRP	H	H	7.48	0.009	127	TRP	H
1272	TRP	HA	H	4.7	0.007	127	TRP	HA
1273	TRP	HB2	H	3.176	0.007	127	TRP	HB2
1274	TRP	HB3	H	2.763	0.008	127	TRP	HB3
1275	TRP	C	C	174.066	0.024	127	TRP	C
1276	TRP	CA	C	53.787	0.086	127	TRP	CA
1277	TRP	CB	C	31.048	0.197	127	TRP	CB

1278	TRP	N	N	115.626	0.066	127	TRP	N
1279	ALA	H	H	7.183	0.008	128	ALA	H
1280	ALA	HA	H	3.113	0.01	128	ALA	HA
1281	ALA	HB1	H	0.808	0.011	128	ALA	HB1
1282	ALA	HB2	H	0.808	0.011	128	ALA	HB2
1283	ALA	HB3	H	0.808	0.011	128	ALA	HB3
1284	ALA	C	C	174.904	0	128	ALA	C
1285	ALA	CA	C	49.328	0.1	128	ALA	CA
1286	ALA	CB	C	17.999	0.149	128	ALA	CB
1287	ALA	N	N	122.226	0.064	128	ALA	N
1288	PRO	HA	H	4.056	0.004	129	PRO	HA
1289	PRO	HB2	H	1.119	0	129	PRO	HB2
1290	PRO	HB3	H	1.031	0	129	PRO	HB3
1291	PRO	HD2	H	3.76	0	129	PRO	HD2
1292	PRO	HD3	H	3.547	0	129	PRO	HD3
1293	PRO	CA	C	61.369	0.035	129	PRO	CA
1294	PRO	CB	C	32.214	0.052	129	PRO	CB
1295	PRO	CG	C	26.101	0.11	129	PRO	CG
1296	PRO	CD	C	48.298	0.059	129	PRO	CD
1297	ASP	H	H	8.077	0.011	130	ASP	H
1298	ASP	HA	H	4.367	0.002	130	ASP	HA
1299	ASP	HB2	H	2.654	0.002	130	ASP	HB2
1300	ASP	HB3	H	2.536	0.002	130	ASP	HB3
1301	ASP	C	C	176.695	0.001	130	ASP	C
1302	ASP	CA	C	54.646	0.08	130	ASP	CA
1303	ASP	CB	C	41.561	0.169	130	ASP	CB
1304	ASP	N	N	120.602	0.054	130	ASP	N
1305	HIS	H	H	8.555	0.01	131	HIS	H
1306	HIS	HA	H	4.111	0.004	131	HIS	HA
1307	HIS	HB2	H	3.055	0.009	131	HIS	HB2
1308	HIS	HB3	H	3.001	0.003	131	HIS	HB3
1309	HIS	C	C	175.702	0.021	131	HIS	C
1310	HIS	CA	C	60.301	0.096	131	HIS	CA
1311	HIS	CB	C	32.222	0.068	131	HIS	CB
1312	HIS	N	N	129.419	0.063	131	HIS	N
1313	GLU	H	H	8.084	0.007	132	GLU	H
1314	GLU	HA	H	3.603	0.005	132	GLU	HA
1315	GLU	HB2	H	2.003	0.002	132	GLU	HB2
1316	GLU	HB3	H	1.941	0.002	132	GLU	HB3
1317	GLU	HG2	H	2.293	0.002	132	GLU	HG2
1318	GLU	HG3	H	2.217	0.003	132	GLU	HG3
1319	GLU	C	C	179.574	0.009	132	GLU	C
1320	GLU	CA	C	59.684	0.105	132	GLU	CA
1321	GLU	CB	C	28.82	0.097	132	GLU	CB
1322	GLU	CG	C	36.622	0.172	132	GLU	CG
1323	GLU	N	N	117.538	0.056	132	GLU	N
1324	LEU	H	H	7.707	0.011	133	LEU	H
1325	LEU	HA	H	4.06	0.003	133	LEU	HA
1326	LEU	HB2	H	1.532	0.002	133	LEU	HB2
1327	LEU	HB3	H	1.532	0.002	133	LEU	HB3
1328	LEU	HG	H	0.855	0	133	LEU	HG
1329	LEU	HD11	H	0.835	0	133	LEU	HD11
1330	LEU	HD12	H	0.835	0	133	LEU	HD12
1331	LEU	HD13	H	0.835	0	133	LEU	HD13
1332	LEU	HD21	H	0.75	0	133	LEU	HD21
1333	LEU	HD22	H	0.75	0	133	LEU	HD22
1334	LEU	HD23	H	0.75	0	133	LEU	HD23
1335	LEU	CA	C	56.821	0.682	133	LEU	CA
1336	LEU	CB	C	41.741	0.049	133	LEU	CB
1337	LEU	CG	C	26.908	0	133	LEU	CG
1338	LEU	CD1	C	24.105	0	133	LEU	CD1
1339	LEU	N	N	120.078	0.08	133	LEU	N

1340	LEU	H	H	7.856	0.01	134	LEU	H
1341	LEU	HA	H	3.846	0.003	134	LEU	HA
1342	LEU	HB2	H	1.985	0.002	134	LEU	HB2
1343	LEU	HB3	H	1.985	0.002	134	LEU	HB3
1344	LEU	HG	H	1.49	0.001	134	LEU	HG
1345	LEU	HD11	H	0.72	0.002	134	LEU	HD11
1346	LEU	HD12	H	0.72	0.002	134	LEU	HD12
1347	LEU	HD13	H	0.72	0.002	134	LEU	HD13
1348	LEU	HD21	H	0.72	0.002	134	LEU	HD21
1349	LEU	HD22	H	0.72	0.002	134	LEU	HD22
1350	LEU	HD23	H	0.72	0.002	134	LEU	HD23
1351	LEU	CA	C	58.323	0.087	134	LEU	CA
1352	LEU	CB	C	41.991	0.051	134	LEU	CB
1353	LEU	CG	C	27.166	0	134	LEU	CG
1354	LEU	CD1	C	24.984	0	134	LEU	CD1
1355	LEU	N	N	120.543	0.074	134	LEU	N
1356	LEU	H	H	8.209	0.007	135	LEU	H
1357	LEU	HA	H	3.842	0.002	135	LEU	HA
1358	LEU	HB2	H	1.427	0.003	135	LEU	HB2
1359	LEU	HB3	H	1.427	0.003	135	LEU	HB3
1360	LEU	HD11	H	0.746	0.002	135	LEU	HD11
1361	LEU	HD12	H	0.746	0.002	135	LEU	HD12
1362	LEU	HD13	H	0.746	0.002	135	LEU	HD13
1363	LEU	HD21	H	0.746	0.002	135	LEU	HD21
1364	LEU	HD22	H	0.746	0.002	135	LEU	HD22
1365	LEU	HD23	H	0.746	0.002	135	LEU	HD23
1366	LEU	C	C	173.588	0	135	LEU	C
1367	LEU	CA	C	57.831	1.246	135	LEU	CA
1368	LEU	CB	C	41.384	0.081	135	LEU	CB
1369	LEU	CG	C	26.786	0	135	LEU	CG
1370	LEU	CD1	C	24.263	0	135	LEU	CD1
1371	LEU	CD2	C	24.263	0	135	LEU	CD2
1372	LEU	N	N	119.361	0.038	135	LEU	N
1373	GLU	H	H	7.604	0.008	136	GLU	H
1374	GLU	HA	H	3.871	0.003	136	GLU	HA
1375	GLU	HB2	H	2.154	0.004	136	GLU	HB2
1376	GLU	HB3	H	2.067	0	136	GLU	HB3
1377	GLU	HG2	H	2.162	0.001	136	GLU	HG2
1378	GLU	HG3	H	2.3	0.003	136	GLU	HG3
1379	GLU	C	C	179.605	0.015	136	GLU	C
1380	GLU	CA	C	59.444	0.296	136	GLU	CA
1381	GLU	CB	C	29.424	0.022	136	GLU	CB
1382	GLU	CG	C	36.008	0.006	136	GLU	CG
1383	GLU	N	N	118.208	0.064	136	GLU	N
1384	GLU	H	H	7.995	0.009	137	GLU	H
1385	GLU	HA	H	4.011	0.003	137	GLU	HA
1386	GLU	HB2	H	1.97	0.002	137	GLU	HB2
1387	GLU	HB3	H	1.879	0.002	137	GLU	HB3
1388	GLU	HG2	H	2.503	0.002	137	GLU	HG2
1389	GLU	HG3	H	2.062	0.002	137	GLU	HG3
1390	GLU	C	C	179.255	0.044	137	GLU	C
1391	GLU	CA	C	59.381	0.16	137	GLU	CA
1392	GLU	CB	C	29.907	0.057	137	GLU	CB
1393	GLU	CG	C	36.454	0.056	137	GLU	CG
1394	GLU	N	N	119.139	0.086	137	GLU	N
1395	MET	H	H	8.521	0.008	138	MET	H
1396	MET	HA	H	4.061	0.002	138	MET	HA
1397	MET	HB2	H	2.15	0.002	138	MET	HB2
1398	MET	HB3	H	2.012	0.002	138	MET	HB3
1399	MET	HG2	H	2.746	0.003	138	MET	HG2
1400	MET	HG3	H	2.412	0.002	138	MET	HG3
1401	MET	HE1	H	1.621	0	138	MET	HE1

1402	MET	HE2	H	1.621	0	138	MET	HE2
1403	MET	HE3	H	1.621	0	138	MET	HE3
1404	MET	C	C	178.832	0.02	138	MET	C
1405	MET	CA	C	59.156	0.3	138	MET	CA
1406	MET	CB	C	33.093	0.046	138	MET	CB
1407	MET	CG	C	32.173	0	138	MET	CG
1408	MET	N	N	117.935	0.054	138	MET	N
1409	MET	H	H	8.279	0.007	139	MET	H
1410	MET	HA	H	4.182	0.005	139	MET	HA
1411	MET	HB2	H	2.098	0.002	139	MET	HB2
1412	MET	HB3	H	1.991	0.002	139	MET	HB3
1413	MET	HG2	H	2.674	0.008	139	MET	HG2
1414	MET	HG3	H	2.553	0.003	139	MET	HG3
1415	MET	HE1	H	1.987	0	139	MET	HE1
1416	MET	HE2	H	1.987	0	139	MET	HE2
1417	MET	HE3	H	1.987	0	139	MET	HE3
1418	MET	C	C	178.359	0.006	139	MET	C
1419	MET	CA	C	57.702	0.155	139	MET	CA
1420	MET	CB	C	32.016	0.138	139	MET	CB
1421	MET	CG	C	30.831	0	139	MET	CG
1422	MET	CE	C	18.749	0	139	MET	CE
1423	MET	N	N	117.386	0.095	139	MET	N
1424	ALA	H	H	7.597	0.009	140	ALA	H
1425	ALA	HA	H	4.118	0.002	140	ALA	HA
1426	ALA	HB1	H	1.447	0.002	140	ALA	HB1
1427	ALA	HB2	H	1.447	0.002	140	ALA	HB2
1428	ALA	HB3	H	1.447	0.002	140	ALA	HB3
1429	ALA	C	C	179.554	0.014	140	ALA	C
1430	ALA	CA	C	54.233	0.055	140	ALA	CA
1431	ALA	CB	C	18.394	0.082	140	ALA	CB
1432	ALA	N	N	121.202	0.072	140	ALA	N
1433	GLU	H	H	7.698	0.01	141	GLU	H
1434	GLU	HA	H	4.134	0.002	141	GLU	HA
1435	GLU	HB2	H	2.054	0.003	141	GLU	HB2
1436	GLU	HB3	H	1.967	0.002	141	GLU	HB3
1437	GLU	HG2	H	2.387	0.002	141	GLU	HG2
1438	GLU	HG3	H	2.222	0.001	141	GLU	HG3
1439	GLU	C	C	177.423	0.012	141	GLU	C
1440	GLU	CA	C	57.303	0.076	141	GLU	CA
1441	GLU	CB	C	30	0.04	141	GLU	CB
1442	GLU	CG	C	36.317	0.005	141	GLU	CG
1443	GLU	N	N	116.59	0.094	141	GLU	N
1444	MET	H	H	7.68	0.02	142	MET	H
1445	MET	HA	H	4.363	0.004	142	MET	HA
1446	MET	HB2	H	2.105	0.003	142	MET	HB2
1447	MET	HB3	H	2.031	0.002	142	MET	HB3
1448	MET	HG2	H	2.651	0.005	142	MET	HG2
1449	MET	HG3	H	2.565	0.002	142	MET	HG3
1450	MET	C	C	176.327	0.36	142	MET	C
1451	MET	CA	C	56.241	0.134	142	MET	CA
1452	MET	CB	C	33.169	0.022	142	MET	CB
1453	MET	CG	C	32.245	0	142	MET	CG
1454	MET	N	N	118.316	0.064	142	MET	N
1455	LYS	H	H	8.001	0.353	143	LYS	H
1456	LYS	HA	H	4.261	0.027	143	LYS	HA
1457	LYS	HB2	H	1.838	0.003	143	LYS	HB2
1458	LYS	HB3	H	1.743	0.004	143	LYS	HB3
1459	LYS	HG2	H	1.387	0.011	143	LYS	HG2
1460	LYS	HG3	H	1.387	0.011	143	LYS	HG3
1461	LYS	HD2	H	1.617	0.005	143	LYS	HD2
1462	LYS	HD3	H	1.617	0.005	143	LYS	HD3
1463	LYS	HE2	H	2.931	0.003	143	LYS	HE2

1464	**LYS**	**HE3**	**H**	**2.931**	**0.003**	**143**	**LYS**	**HE3**
1465	**LYS**	**C**	**C**	**174.433**	**1.303**	**143**	**LYS**	**C**
1466	**LYS**	**CA**	**C**	**56.216**	**0.318**	**143**	**LYS**	**CA**
1467	**LYS**	**CB**	**C**	**32.72**	**0.056**	**143**	**LYS**	**CB**
1468	**LYS**	**CG**	**C**	**24.382**	**0.088**	**143**	**LYS**	**CG**
1469	**LYS**	**CD**	**C**	**28.912**	**0.028**	**143**	**LYS**	**CD**
1470	**LYS**	**CE**	**C**	**42.175**	**0.024**	**143**	**LYS**	**CE**
1471	**LYS**	**N**	**N**	**121.648**	**0.722**	**143**	**LYS**	**N**
1472	**GLN**	**H**	**H**	**7.775**	**0.014**	**144**	**GLN**	**H**
1473	**GLN**	**HA**	**H**	**4.097**	**0.002**	**144**	**GLN**	**HA**
1474	**GLN**	**HB2**	**H**	**2.055**	**0.002**	**144**	**GLN**	**HB2**
1475	**GLN**	**HB3**	**H**	**1.858**	**0.005**	**144**	**GLN**	**HB3**
1476	**GLN**	**C**	**C**	**180.768**	**0.182**	**144**	**GLN**	**C**
1477	**GLN**	**CA**	**C**	**57.336**	**0.158**	**144**	**GLN**	**CA**
1478	**GLN**	**CB**	**C**	**30.454**	**0.07**	**144**	**GLN**	**CB**
1479	**GLN**	**N**	**N**	**126.471**	**0.149**	**144**	**GLN**	**N**

Appendix Table A-2 **Backbone chemical shift assignments of the delipidated OfurPBP2**

1	**2**	**GLN**	**C**	**175.719**	**0.011**
2	**2**	**GLN**	**CA**	**51.794**	**0.036**
3	**2**	**GLN**	**CB**	**36.604**	**0.057**
4	**2**	**GLN**	**HA**	**4.573**	**0**
5	**2**	**GLN**	**QB**	**2.632**	**0**
6	**2**	**GLN**	**QG**	**2.692**	**0**
7	**3**	**ALA**	**C**	**176.646**	**0.02**
8	**3**	**ALA**	**CA**	**50.022**	**0.06**
9	**3**	**ALA**	**CB**	**16.6**	**0.03**

10	**3**	**ALA**	**H**	**8.061**	**0.003**
11	**3**	**ALA**	**HA**	**4.361**	**0.024**
12	**3**	**ALA**	**N**	**123.989**	**0.025**
13	**3**	**ALA**	**QB**	**1.397**	**0.003**
14	**4**	**VAl**	**C**	**176.162**	**0**
15	**4**	**VAl**	**CA**	**60.943**	**0**
16	**4**	**VAl**	**CB**	**30.656**	**0**
17	**4**	**VAl**	**H**	**7.65**	**0.002**
18	**4**	**VAl**	**HA**	**4.042**	**0**
19	**4**	**VAl**	**HB**	**2.085**	**0**
20	**4**	**VAl**	**N**	**123.477**	**0.05**
21	**4**	**VAl**	**QQG**	**0.913**	**0**
22	**5**	**MET**	**C**	**176.149**	**0.011**
23	**5**	**MET**	**CA**	**53.668**	**0.018**
24	**5**	**MET**	**CB**	**30.535**	**0.011**
25	**5**	**MET**	**CG**	**29.645**	**0**
26	**5**	**MET**	**H**	**7.718**	**0**
27	**5**	**MET**	**HA**	**4.461**	**0**
28	**5**	**MET**	**HB2**	**2.202**	**0**
29	**5**	**MET**	**HB3**	**2.128**	**0**
30	**5**	**MET**	**N**	**117.096**	**0**
31	**6**	**LYS**	**C**	**175.618**	**0.068**
32	**6**	**LYS**	**CA**	**53.315**	**0.1**
33	**6**	**LYS**	**CB**	**30.447**	**0.114**
34	**6**	**LYS**	**H**	**7.817**	**0.003**
35	**6**	**LYS**	**HA**	**4.451**	**0.062**
36	**6**	**LYS**	**N**	**120.994**	**0.025**
37	**6**	**LYS**	**QB**	**2.035**	**0.001**
38	**6**	**LYS**	**QE**	**2.934**	**0**
39	**7**	**ASP**	**C**	**176.021**	**0**
40	**7**	**ASP**	**CB**	**38.464**	**0**
41	**7**	**ASP**	**H**	**8.433**	**0.007**
42	**7**	**ASP**	**HA**	**4.323**	**0.018**
43	**7**	**ASP**	**N**	**123.132**	**0.015**
44	**7**	**ASP**	**QB**	**1.79**	**0.003**
45	**8**	**MET**	**C**	**176.133**	**0.006**
46	**8**	**MET**	**CA**	**53.217**	**0.063**
47	**8**	**MET**	**CB**	**30.159**	**0.022**
48	**8**	**MET**	**CG**	**29.438**	**0**
49	**8**	**MET**	**H**	**8.015**	**0**
50	**8**	**MET**	**HA**	**4.569**	**0**
51	**8**	**MET**	**N**	**127.226**	**0**
52	**8**	**MET**	**QB**	**2.051**	**0**
53	**8**	**MET**	**QG**	**2.145**	**0**

54	**9**	**THR**	**C**	**174.165**	**0.029**
55	**9**	**THR**	**CA**	**59.587**	**0.138**
56	**9**	**THR**	**CB**	**67.177**	**0.045**
57	**9**	**THR**	**H**	**8.278**	**0.004**
58	**9**	**THR**	**HA**	**4.349**	**0.054**
59	**9**	**THR**	**HG2**	**1.213**	**0**
60	**9**	**THR**	**N**	**115.979**	**0.019**
61	**9**	**THR**	**QB**	**1.226**	**0**
62	**10**	**LYS**	**C**	**175.082**	**0.009**
63	**10**	**LYS**	**CA**	**59.766**	**0.03**
64	**10**	**LYS**	**CB**	**30.22**	**0.042**
65	**10**	**LYS**	**H**	**8.229**	**0.004**
66	**10**	**LYS**	**HA**	**4.162**	**0**
67	**10**	**LYS**	**N**	**122.973**	**0.03**
68	**10**	**LYS**	**QB**	**2.137**	**0**
69	**11**	**ASN**	**C**	**176.18**	**0**
70	**11**	**ASN**	**CA**	**54.492**	**0**
71	**11**	**ASN**	**CB**	**41.251**	**0**
72	**11**	**ASN**	**H**	**7.972**	**0.011**
73	**11**	**ASN**	**N**	**129.542**	**0.036**
74	**12**	**PHE**	**C**	**177.058**	**0.002**
75	**12**	**PHE**	**CA**	**59.208**	**0.06**
76	**12**	**PHE**	**CB**	**36.723**	**0.045**
77	**12**	**PHE**	**H**	**8.757**	**0**
78	**12**	**PHE**	**HA**	**3.926**	**0.006**
79	**12**	**PHE**	**HB2**	**3.362**	**0.004**
80	**12**	**PHE**	**N**	**123.225**	**0**
81	**12**	**PHE**	**HB3**	**3.232**	**0.001**
82	**13**	**ILE**	**C**	**177.977**	**0.006**
83	**13**	**ILE**	**CA**	**62.662**	**0.034**
84	**13**	**ILE**	**CB**	**35.392**	**0.049**
85	**13**	**ILE**	**CD1**	**14.805**	**0**
86	**13**	**ILE**	**CG1**	**27.283**	**0**
87	**13**	**ILE**	**CG2**	**16.678**	**0**
88	**13**	**ILE**	**H**	**8.18**	**0.006**
89	**13**	**ILE**	**HA**	**3.69**	**0.011**
90	**13**	**ILE**	**N**	**118.603**	**0.02**
91	**13**	**ILE**	**QB**	**1.984**	**0.009**
92	**13**	**ILE**	**QD1**	**0.901**	**0**
93	**13**	**ILE**	**QG1**	**1.385**	**0**
94	**13**	**ILE**	**QG2**	**1.032**	**0.009**
95	**14**	**LYS**	**C**	**179.069**	**0.006**
96	**14**	**LYS**	**CA**	**56.555**	**0.049**
97	**14**	**LYS**	**CB**	**29.412**	**0.05**

98	**14**	**LYS**	**CD**	**26.349**	**0**
99	**14**	**LYS**	**CE**	**38.722**	**0**
100	**14**	**LYS**	**CG**	**22.462**	**0**
101	**14**	**LYS**	**H**	**7.963**	**0.007**
102	**14**	**LYS**	**HA**	**4.114**	**0.002**
103	**14**	**LYS**	**HG**	**1.495**	**0**
104	**14**	**LYS**	**N**	**119.336**	**0.027**
105	**14**	**LYS**	**QB**	**1.955**	**0.004**
106	**15**	**ALA**	**C**	**178.809**	**0.014**
107	**15**	**ALA**	**CA**	**52.195**	**0.037**
108	**15**	**ALA**	**CB**	**16.125**	**0.014**
109	**15**	**ALA**	**HA**	**3.897**	**0.007**
110	**15**	**ALA**	**N**	**121.235**	**0.019**
111	**15**	**ALA**	**QB**	**1.189**	**0.012**
112	**15**	**ALA**	**H**	**7.941**	**0**
113	**16**	**TYR**	**C**	**176.473**	**0.009**
114	**16**	**TYR**	**CA**	**60.878**	**0.059**
115	**16**	**TYR**	**CB**	**35.123**	**0.054**
116	**16**	**TYR**	**H**	**8.783**	**0.005**
117	**16**	**TYR**	**HA**	**4.733**	**0.01**
118	**16**	**TYR**	**HB2**	**2.962**	**0.009**
119	**16**	**TYR**	**HB3**	**2.564**	**0.028**
120	**16**	**TYR**	**N**	**119.134**	**0.027**
121	**17**	**GLU**	**C**	**179.944**	**0.002**
122	**17**	**GLU**	**CA**	**57.192**	**0.033**
123	**17**	**GLU**	**CB**	**26.752**	**0.027**
124	**17**	**GLU**	**CG**	**34.226**	**0**
125	**17**	**GLU**	**H**	**8.448**	**0.007**
126	**17**	**GLU**	**HA**	**3.78**	**0.004**
127	**17**	**GLU**	**HB2**	**2.074**	**0.001**
128	**17**	**GLU**	**HB3**	**1.911**	**0.003**
129	**17**	**GLU**	**N**	**118.927**	**0.02**
130	**17**	**GLU**	**QG**	**2.284**	**0.009**
131	**18**	**VAL**	**C**	**178.973**	**0.002**
132	**18**	**VAL**	**CA**	**64.09**	**0.014**
133	**18**	**VAL**	**CB**	**28.739**	**0.041**
134	**18**	**VAL**	**CG1**	**19.891**	**0**
135	**18**	**VAL**	**CG2**	**18.944**	**0**
136	**18**	**VAL**	**H**	**7.481**	**0.013**
137	**18**	**VAL**	**HA**	**3.673**	**0.008**
138	**18**	**VAL**	**HB**	**2.155**	**0.004**
139	**18**	**VAL**	**N**	**119.434**	**0.038**
140	**18**	**VAL**	**QQG**	**0.914**	**0**
141	**19**	**CYS**	**C**	**175.884**	**0.011**

142	**19**	**CYS**	**CA**	**57.641**	**0.049**
143	**19**	**CYS**	**CB**	**38.783**	**0.015**
144	**19**	**CYS**	**H**	**7.811**	**0.005**
145	**19**	**CYS**	**HA**	**4.263**	**0.006**
146	**19**	**CYS**	**HB2**	**2.983**	**0.002**
147	**19**	**CYS**	**HB3**	**2.354**	**0.004**
148	**19**	**CYS**	**N**	**119.307**	**0.027**
149	**20**	**ALA**	**C**	**180.165**	**0.009**
150	**20**	**ALA**	**CA**	**52.512**	**0.05**
151	**20**	**ALA**	**CB**	**14.682**	**0.025**
152	**20**	**ALA**	**H**	**9.241**	**0.007**
153	**20**	**ALA**	**HA**	**4.658**	**0.012**
154	**20**	**ALA**	**N**	**123.005**	**0.034**
155	**20**	**ALA**	**QB**	**1.222**	**0.006**
156	**21**	**LYS**	**C**	**179.009**	**0.001**
157	**21**	**LYS**	**CA**	**55.684**	**0.073**
158	**21**	**LYS**	**CB**	**29.808**	**0.024**
159	**21**	**LYS**	**CD**	**26.277**	**0**
160	**21**	**LYS**	**CE**	**39.463**	**0**
161	**21**	**LYS**	**CG**	**22.152**	**0**
162	**21**	**LYS**	**H**	**7.327**	**0.007**
163	**21**	**LYS**	**HA**	**4.245**	**0.004**
164	**21**	**LYS**	**HD2**	**1.584**	**0**
165	**21**	**LYS**	**HD3**	**1.464**	**0**
166	**21**	**LYS**	**HG2**	**1.584**	**0**
167	**21**	**LYS**	**HG3**	**1.464**	**0**
168	**21**	**LYS**	**N**	**116.811**	**0.022**
169	**21**	**LYS**	**QB**	**1.954**	**0.003**
170	**21**	**LYS**	**QE**	**3.056**	**0.045**
171	**22**	**GLU**	**C**	**178.009**	**0**
172	**22**	**GLU**	**CA**	**56.964**	**0.045**
173	**22**	**GLU**	**CB**	**27.906**	**0.05**
174	**22**	**GLU**	**CG**	**33.652**	**0**
175	**22**	**GLU**	**H**	**8.23**	**0.007**
176	**22**	**GLU**	**HA**	**3.907**	**0.004**
177	**22**	**GLU**	**HB2**	**1.901**	**0.007**
178	**22**	**GLU**	**HB3**	**2.146**	**0.003**
179	**22**	**GLU**	**N**	**119.896**	**0.015**
180	**22**	**GLU**	**QG**	**2.326**	**0.012**
181	**23**	**TYR**	**C**	**174.579**	**0.009**
182	**23**	**TYR**	**CA**	**55.649**	**0.082**
183	**23**	**TYR**	**CB**	**35.367**	**0.009**
184	**23**	**TYR**	**CG**	**33.652**	**0**
185	**23**	**TYR**	**H**	**8.329**	**0.006**

186	**23**	**TYR**	**HA**	**4.406**	**0.013**
187	**23**	**TYR**	**HB2**	**3.167**	**0.006**
188	**23**	**TYR**	**HB3**	**2.615**	**0.005**
189	**23**	**TYR**	**N**	**114.005**	**0.029**
190	**24**	**ASN**	**C**	**175.167**	**0.012**
191	**24**	**ASN**	**CA**	**51.49**	**0.034**
192	**24**	**ASN**	**CB**	**34.311**	**0.014**
193	**24**	**ASN**	**H**	**7.509**	**0.006**
194	**24**	**ASN**	**HA**	**4.462**	**0.01**
195	**24**	**ASN**	**HB2**	**3.103**	**0.008**
196	**24**	**ASN**	**HB3**	**2.673**	**0.003**
197	**24**	**ASN**	**HD21**	**6.659**	**0**
198	**24**	**ASN**	**HD22**	**7.429**	**0**
199	**24**	**ASN**	**N**	**119.599**	**0.024**
200	**24**	**ASN**	**ND2**	**112.017**	**0**
201	**25**	**LEU**	**C**	**176.195**	**0**
202	**25**	**LEU**	**CA**	**50.198**	**0**
203	**25**	**LEU**	**CB**	**38.498**	**0**
204	**25**	**LEU**	**H**	**8.242**	**0.005**
205	**25**	**LEU**	**HA**	**4.514**	**0.006**
206	**25**	**LEU**	**HB2**	**2.829**	**0**
207	**25**	**LEU**	**HB3**	**2.62**	**0.01**
208	**25**	**LEU**	**HG**	**1.595**	**0.018**
209	**25**	**LEU**	**N**	**117.381**	**0.031**
210	**26**	**PRO**	**C**	**177.789**	**0.007**
211	**26**	**PRO**	**CA**	**59.366**	**0.078**
212	**26**	**PRO**	**CB**	**29.394**	**0.107**
213	**26**	**PRO**	**CG**	**24.261**	**0**
214	**26**	**PRO**	**HA**	**4.707**	**0.012**
215	**26**	**PRO**	**QB**	**2.478**	**0.003**
216	**26**	**PRO**	**QD**	**2.981**	**0**
217	**26**	**PRO**	**QG**	**2.184**	**0.003**
218	**27**	**GLU**	**C**	**179.048**	**0.008**
219	**27**	**GLU**	**CA**	**57.052**	**0.041**
220	**27**	**GLU**	**CB**	**26.772**	**0.075**
221	**27**	**GLU**	**CG**	**33.707**	**0**
222	**27**	**GLU**	**H**	**8.985**	**0.006**
223	**27**	**GLU**	**HA**	**4.152**	**0.007**
224	**27**	**GLU**	**N**	**123.664**	**0.032**
225	**27**	**GLU**	**QB**	**2.095**	**0.002**
226	**27**	**GLU**	**QG**	**2.415**	**0.021**
227	**28**	**ALA**	**C**	**179.535**	**0.002**
228	**28**	**ALA**	**CA**	**52.15**	**0.024**
229	**28**	**ALA**	**CB**	**16.104**	**0.004**

230	**28**	**ALA**	**H**	**8.628**	**0.005**
231	**28**	**ALA**	**HA**	**4.126**	**0.004**
232	**28**	**ALA**	**N**	**119.505**	**0.022**
233	**28**	**ALA**	**QB**	**1.463**	**0.003**
234	**29**	**ALA**	**C**	**178.769**	**0.014**
235	**29**	**ALA**	**CA**	**52.319**	**0.056**
236	**29**	**ALA**	**CB**	**15.556**	**0.05**
237	**29**	**ALA**	**H**	**7.56**	**0.011**
238	**29**	**ALA**	**HA**	**4.182**	**0.007**
239	**29**	**ALA**	**N**	**119.049**	**0.028**
240	**29**	**ALA**	**QB**	**1.416**	**0.004**
241	**30**	**GLY**	**C**	**175.378**	**0.003**
242	**30**	**GLY**	**CA**	**44.996**	**0.046**
243	**30**	**GLY**	**H**	**7.405**	**0.007**
244	**30**	**GLY**	**HA2**	**3.586**	**0.006**
245	**30**	**GLY**	**HA3**	**3.342**	**0.009**
246	**30**	**GLY**	**N**	**106.223**	**0.037**
247	**31**	**ALA**	**C**	**179.883**	**0.015**
248	**31**	**ALA**	**CA**	**51.947**	**0.024**
249	**31**	**ALA**	**CB**	**15.396**	**0.081**
250	**31**	**ALA**	**H**	**7.495**	**0.007**
251	**31**	**ALA**	**HA**	**4.003**	**0.01**
252	**31**	**ALA**	**N**	**122.649**	**0.028**
253	**31**	**ALA**	**QB**	**1.404**	**0.008**
254	**32**	**GLU**	**C**	**178.698**	**0.018**
255	**32**	**GLU**	**CA**	**57.289**	**0.013**
256	**32**	**GLU**	**CB**	**27.461**	**0.04**
257	**32**	**GLU**	**CG**	**34.324**	**0**
258	**32**	**GLU**	**H**	**7.533**	**0.008**
259	**32**	**GLU**	**HA**	**4.184**	**0.011**
260	**32**	**GLU**	**N**	**117.494**	**0.031**
261	**32**	**GLU**	**QB**	**2.187**	**0.006**
262	**32**	**GLU**	**QG**	**2.373**	**0**
263	**33**	**VAL**	**C**	**178.106**	**0.011**
264	**33**	**VAL**	**CA**	**64.284**	**0.047**
265	**33**	**VAL**	**CB**	**28.795**	**0.004**
266	**33**	**VAL**	**CG1**	**22.359**	**0**
267	**33**	**VAL**	**H**	**7.877**	**0.006**
268	**33**	**VAL**	**HA**	**3.604**	**0.002**
269	**33**	**VAL**	**HB**	**2.057**	**0.004**
270	**33**	**VAL**	**N**	**116.562**	**0.022**
271	**33**	**VAL**	**QG1**	**1.133**	**0.009**
272	**33**	**VAL**	**QG2**	**0.925**	**0.015**
273	**34**	**MET**	**C**	**177.631**	**0.01**

274	34	MET	CA	54.632	0.018
275	34	MET	CB	28.481	0.013
276	34	MET	CG	30.709	0
277	34	MET	H	8.218	0.003
278	34	MET	HA	4.256	0.005
279	34	MET	HB2	2.062	0
280	34	MET	HB3	1.992	0.002
281	34	MET	N	116.612	0.082
282	35	ASN	C	175.192	0.012
283	35	ASN	CA	50.821	0.053
284	35	ASN	CB	36.386	0
285	35	ASN	H	7.494	0.007
286	35	ASN	HA	4.741	0.03
287	35	ASN	HB2	2.844	0
288	35	ASN	HB3	2.782	0
289	35	ASN	N	117.387	0.033
290	36	PHE	C	174.289	0
291	36	PHE	CA	53.156	0.043
292	36	PHE	CB	38.215	0.003
293	36	PHE	H	8.419	0.005
294	36	PHE	HA	4.678	0
295	36	PHE	N	118.552	0.045
296	36	PHE	QB	2.623	0
297	37	TRP	C	175.287	0.021
298	37	TRP	CA	54.598	0.036
299	37	TRP	CB	27.023	0.004
300	37	TRP	H	8.2	0
301	37	TRP	HA	4.615	0.001
302	37	TRP	HE1	10.079	0
303	37	TRP	N	121.793	0
304	37	TRP	NE1	127.66	0
305	37	TRP	QB	3.215	0
306	38	LYS	C	175.494	0.011
307	38	LYS	CA	53.373	0.041
308	38	LYS	CB	30.925	0.016
309	38	LYS	CD	27.834	0
310	38	LYS	CE	38.458	0
311	38	LYS	H	7.784	0.002
312	38	LYS	HA	4.187	0.005
313	38	LYS	N	123.225	0.032
314	38	LYS	QB	1.745	0.032
315	38	LYS	QD	1.615	0.012
316	38	LYS	QE	3.018	0.073
317	38	LYS	QG	1.261	0.002

318	39	GLU	C	175.603	0.067
319	39	GLU	CA	53.903	0.036
320	39	GLU	CB	27.928	0.031
321	39	GLU	CG	33.713	0
322	39	GLU	H	8.197	0.003
323	39	GLU	HA	4.434	0.003
324	39	GLU	HB2	2.108	0
325	39	GLU	HB3	1.977	0.011
326	39	GLU	N	122.801	0.026
327	39	GLU	QG	2.241	0
328	40	GLY	C	173.287	0.003
329	40	GLY	CA	43.148	0.273
330	40	GLY	H	8.103	0.001
331	40	GLY	HA2	4.083	0.011
332	40	GLY	HA3	3.683	0.063
333	40	GLY	N	116.8	0.04
334	41	TYR	C	174.92	0
335	41	TYR	CA	55.305	0
336	41	TYR	CB	37.791	0
337	41	TYR	H	7.395	0.007
338	41	TYR	HA	4.494	0.003
339	41	TYR	N	120.576	0.019
340	41	TYR	QB	2.683	0.001
341	42	VAL	C	174.06	0.013
342	42	VAL	CA	58.235	0.079
343	42	VAL	CB	31.119	0.016
344	42	VAL	CG2	18.306	0
345	42	VAL	H	8.071	0
346	42	VAL	HA	3.982	0.001
347	42	VAL	HB	1.759	0.002
348	42	VAL	N	128.049	0
349	42	VAL	QQG	0.72	0.002
350	43	LEU	C	176.06	0.016
351	43	LEU	CA	53.481	0.039
352	43	LEU	CB	40.12	0.058
353	43	LEU	H	7.495	0.006
354	43	LEU	HA	4.221	0.008
355	43	LEU	N	127.513	0.039
356	43	LEU	QB	1.616	0.058
357	43	LEU	QG	1.666	0
358	44	THR	C	175.086	0.011
359	44	THR	CA	59.334	0.07
360	44	THR	CB	68.785	0.08
361	44	THR	CG2	18.766	0

362	**44**	**THR**	**H**	**9.048**	**0.008**
363	**44**	**THR**	**HA**	**4.403**	**0.008**
364	**44**	**THR**	**HB**	**4.171**	**0.007**
365	**44**	**THR**	**N**	**111.03**	**0.03**
366	**44**	**THR**	**QG2**	**1.031**	**0.007**
367	**45**	**SER**	**C**	**174.66**	**0.005**
368	**45**	**SER**	**CA**	**55.634**	**0.038**
369	**45**	**SER**	**CB**	**61.988**	**0.043**
370	**45**	**SER**	**H**	**8.95**	**0.012**
371	**45**	**SER**	**HA**	**4.497**	**0.014**
372	**45**	**SER**	**HB2**	**4.262**	**0.008**
373	**45**	**SER**	**HB3**	**4.026**	**0.008**
374	**45**	**SER**	**N**	**121.124**	**0.033**
375	**46**	**ARG**	**C**	**178.316**	**0.006**
376	**46**	**ARG**	**CA**	**57.417**	**0.018**
377	**46**	**ARG**	**CB**	**26.905**	**0.065**
378	**46**	**ARG**	**CD**	**40.436**	**0**
379	**46**	**ARG**	**CG**	**26.265**	**0**
380	**46**	**ARG**	**H**	**8.989**	**0.006**
381	**46**	**ARG**	**HA**	**3.694**	**0.007**
382	**46**	**ARG**	**N**	**125.09**	**0.021**
383	**46**	**ARG**	**QB**	**1.948**	**0.005**
384	**46**	**ARG**	**QG**	**1.698**	**0.003**
385	**47**	**GLU**	**C**	**178.461**	**0.01**
386	**47**	**GLU**	**CA**	**58.818**	**0.012**
387	**47**	**GLU**	**CB**	**26.103**	**0.059**
388	**47**	**GLU**	**CG**	**35.269**	**0**
389	**47**	**GLU**	**H**	**9.099**	**0.006**
390	**47**	**GLU**	**HA**	**3.861**	**0.012**
391	**47**	**GLU**	**N**	**116.148**	**0.024**
392	**47**	**GLU**	**QB**	**1.83**	**0.005**
393	**47**	**GLU**	**QG**	**2.163**	**0.002**
394	**48**	**ALA**	**C**	**179.776**	**0.009**
395	**48**	**ALA**	**CA**	**52.85**	**0.023**
396	**48**	**ALA**	**CB**	**15.824**	**0.034**
397	**48**	**ALA**	**H**	**8.092**	**0.005**
398	**48**	**ALA**	**HA**	**4.036**	**0.003**
399	**48**	**ALA**	**N**	**122.702**	**0.026**
400	**48**	**ALA**	**QB**	**1.58**	**0.005**
401	**49**	**GLY**	**C**	**175.804**	**0.011**
402	**49**	**GLY**	**CA**	**45.888**	**0.043**
403	**49**	**GLY**	**H**	**7.616**	**0.005**
404	**49**	**GLY**	**HA2**	**3.526**	**0.007**
405	**49**	**GLY**	**HA3**	**4.116**	**0.003**

406	**49**	**GLY**	**N**	**105.002**	**0.026**
407	**50**	**CYS**	**C**	**177.262**	**0.012**
408	**50**	**CYS**	**CA**	**52.527**	**0.05**
409	**50**	**CYS**	**CB**	**32.765**	**0.025**
410	**50**	**CYS**	**H**	**8.806**	**0.007**
411	**50**	**CYS**	**HA**	**4.644**	**0.071**
412	**50**	**CYS**	**HB2**	**2.97**	**0.024**
413	**50**	**CYS**	**HB3**	**2.538**	**0.01**
414	**50**	**CYS**	**N**	**119.395**	**0.053**
415	**51**	**ALA**	**C**	**178.318**	**0.009**
416	**51**	**ALA**	**CA**	**53.404**	**0.034**
417	**51**	**ALA**	**CB**	**15.947**	**0.026**
418	**51**	**ALA**	**H**	**8.431**	**0.006**
419	**51**	**ALA**	**HA**	**3.672**	**0.007**
420	**51**	**ALA**	**N**	**123.628**	**0.036**
421	**51**	**ALA**	**QB**	**1.642**	**0.006**
422	**52**	**ILE**	**C**	**177.255**	**0.011**
423	**52**	**ILE**	**CA**	**62.004**	**0.049**
424	**52**	**ILE**	**CB**	**34.584**	**0.05**
425	**52**	**ILE**	**CD1**	**16.625**	**0**
426	**52**	**ILE**	**CG1**	**27.773**	**0**
427	**52**	**ILE**	**H**	**7.811**	**0.005**
428	**52**	**ILE**	**HA**	**3.576**	**0.009**
429	**52**	**ILE**	**HB**	**1.961**	**0.003**
430	**52**	**ILE**	**N**	**118.349**	**0.021**
431	**52**	**ILE**	**QG1**	**0.766**	**0.005**
432	**53**	**LEU**	**C**	**179.966**	**0.005**
433	**53**	**LEU**	**CA**	**55.34**	**0.068**
434	**53**	**LEU**	**CB**	**39.748**	**0.056**
435	**53**	**LEU**	**CG**	**27.355**	**0**
436	**53**	**LEU**	**H**	**8.102**	**0.006**
437	**53**	**LEU**	**HA**	**3.913**	**0.037**
438	**53**	**LEU**	**HG**	**1.433**	**0.012**
439	**53**	**LEU**	**N**	**120.301**	**0.064**
440	**53**	**LEU**	**QB**	**2.989**	**0.02**
441	**54**	**CYS**	**C**	**177.555**	**0.007**
442	**54**	**CYS**	**CA**	**57.381**	**0.009**
443	**54**	**CYS**	**CB**	**39.207**	**0.049**
444	**54**	**CYS**	**H**	**8.436**	**0.008**
445	**54**	**CYS**	**HA**	**4.043**	**0.003**
446	**54**	**CYS**	**HB2**	**2.734**	**0.045**
447	**54**	**CYS**	**HB3**	**2.787**	**0.044**
448	**54**	**CYS**	**N**	**119.62**	**0.022**
449	**55**	**LEU**	**C**	**178.482**	**0.012**

450	**55**	**LEU**	**CA**	**56.017**	**0.037**
451	**55**	**LEU**	**CB**	**39.664**	**0.015**
452	**55**	**LEU**	**CG**	**26.147**	**0**
453	**55**	**LEU**	**H**	**9.022**	**0.006**
454	**55**	**LEU**	**HA**	**3.636**	**0.003**
455	**55**	**LEU**	**N**	**125.314**	**0.029**
456	**55**	**LEU**	**QB**	**2.065**	**0.02**
457	**55**	**LEU**	**QD2**	**0.409**	**0.001**
458	**55**	**LEU**	**QD1**	**0.632**	**0.001**
459	**55**	**LEU**	**HG**	**1.094**	**0.002**
460	**56**	**SER**	**CA**	**58.764**	**0.026**
461	**56**	**SER**	**CB**	**60.6**	**0.05**
462	**56**	**SER**	**H**	**8.018**	**0.003**
463	**56**	**SER**	**HA**	**4.064**	**0.008**
464	**56**	**SER**	**HB2**	**3.99**	**0.003**
465	**56**	**SER**	**HB3**	**3.891**	**0.009**
466	**56**	**SER**	**N**	**112.79**	**0.016**
467	**57**	**SER**	**C**	**174.757**	**0.029**
468	**57**	**SER**	**CA**	**56.206**	**0.054**
469	**57**	**SER**	**CB**	**62.105**	**0.054**
470	**57**	**SER**	**H**	**7.811**	**0.008**
471	**57**	**SER**	**HA**	**4.716**	**0.009**
472	**57**	**SER**	**HB2**	**4.234**	**0.013**
473	**57**	**SER**	**HB3**	**4.093**	**0.073**
474	**57**	**SER**	**N**	**115.695**	**0.019**
475	**58**	**LYS**	**C**	**177.654**	**0.02**
476	**58**	**LYS**	**CA**	**50.882**	**0.093**
477	**58**	**LYS**	**CB**	**28.912**	**0.072**
478	**58**	**LYS**	**CD**	**26.252**	**0**
479	**58**	**LYS**	**CE**	**45.576**	**0**
480	**58**	**LYS**	**CG**	**21.678**	**0**
481	**58**	**LYS**	**H**	**7.889**	**0.007**
482	**58**	**LYS**	**HA**	**4.884**	**0.007**
483	**58**	**LYS**	**HB2**	**2.048**	**0.008**
484	**58**	**LYS**	**HB3**	**1.774**	**0.02**
485	**58**	**LYS**	**N**	**123.704**	**0.048**
486	**58**	**LYS**	**HD2**	**0.851**	**0**
487	**58**	**LYS**	**HD3**	**0.679**	**0**
488	**58**	**LYS**	**QE**	**2.785**	**0**
489	**59**	**LEU**	**C**	**177.896**	**0.008**
490	**59**	**LEU**	**CA**	**55.846**	**0.086**
491	**59**	**LEU**	**CB**	**37.633**	**0.042**
492	**59**	**LEU**	**CG**	**27.575**	**0**
493	**59**	**LEU**	**H**	**7.786**	**0.004**

494	**59**	**LEU**	**HA**	**3.809**	**0.005**
495	**59**	**LEU**	**HG**	**1.394**	**0.002**
496	**59**	**LEU**	**N**	**119.774**	**0.067**
497	**59**	**LEU**	**QB**	**1.821**	**0.019**
498	**59**	**LEU**	**QD1**	**0.851**	**0**
499	**59**	**LEU**	**QD2**	**0.679**	**0**
500	**60**	**ASN**	**C**	**175.858**	**0.002**
501	**60**	**ASN**	**CA**	**52.583**	**0.073**
502	**60**	**ASN**	**CB**	**34.249**	**0.022**
503	**60**	**ASN**	**H**	**8.555**	**0.01**
504	**60**	**ASN**	**HA**	**4.438**	**0.009**
505	**60**	**ASN**	**HB2**	**2.793**	**0.011**
506	**60**	**ASN**	**HB3**	**2.748**	**0.012**
507	**60**	**ASN**	**HD21**	**6.614**	**0**
508	**60**	**ASN**	**HD22**	**7.398**	**0**
509	**60**	**ASN**	**N**	**113.293**	**0.045**
510	**60**	**ASN**	**ND2**	**112.611**	**0**
511	**61**	**LEU**	**C**	**177.983**	**0.009**
512	**61**	**LEU**	**CA**	**53.18**	**0.035**
513	**61**	**LEU**	**CB**	**39.795**	**0.037**
514	**61**	**LEU**	**CG**	**28.277**	**0**
515	**61**	**LEU**	**H**	**7.435**	**0.006**
516	**61**	**LEU**	**HA**	**4.272**	**0.004**
517	**61**	**LEU**	**HG**	**1.411**	**0.006**
518	**61**	**LEU**	**N**	**117.844**	**0.033**
519	**61**	**LEU**	**QB**	**1.858**	**0.002**
520	**61**	**LEU**	**QQD**	**0.933**	**0.002**
521	**62**	**LEU**	**C**	**176.769**	**0**
522	**62**	**LEU**	**CA**	**54.414**	**0.186**
523	**62**	**LEU**	**CB**	**39.651**	**0.161**
524	**62**	**LEU**	**CG**	**27.318**	**0**
525	**62**	**LEU**	**H**	**7.377**	**0.008**
526	**62**	**LEU**	**HA**	**3.95**	**0**
527	**62**	**LEU**	**HG**	**1.378**	**0**
528	**62**	**LEU**	**N**	**116.98**	**0.04**
529	**62**	**LEU**	**QB**	**1.673**	**0.012**
530	**62**	**LEU**	**QD2**	**0.719**	**0**
531	**64**	**PRO**	**C**	**177.925**	**0.01**
532	**64**	**PRO**	**CA**	**61.631**	**0.024**
533	**64**	**PRO**	**CB**	**29.631**	**0.011**
534	**64**	**PRO**	**HA**	**4.576**	**0**
535	**64**	**PRO**	**QB**	**2.027**	**0.013**
536	**64**	**PRO**	**QG**	**2.427**	**0.009**
537	**65**	**GLU**	**C**	**178.249**	**0.011**

538	65	GLU	CA	54.702	0.055
539	65	GLU	CB	27.103	0.014
540	65	GLU	CG	34.313	0
541	65	GLU	H	8.002	0.009
542	65	GLU	HA	4.261	0.005
543	65	GLU	N	116.093	0.033
544	65	GLU	QB	2.291	0.006
545	65	GLU	QG	2.361	0.028
546	66	GLY	C	175.171	0.013
547	66	GLY	CA	43.951	0.08
548	66	GLY	H	7.925	0.005
549	66	GLY	HA2	4.046	0
550	66	GLY	HA3	4.043	0.003
551	66	GLY	N	106.489	0.041
552	67	THR	C	175.235	0.069
553	67	THR	CA	61.139	0.031
554	67	THR	CB	66.427	0.046
555	67	THR	H	8.34	0.004
556	67	THR	HA	4.361	0.077
557	67	THR	HB	1.261	0
558	67	THR	HG2	1.262	0.001
559	67	THR	N	115.352	0.043
560	68	LEU	CA	52.66	0.065
561	68	LEU	CB	39.791	0.032
562	68	LEU	H	8.893	0.008
563	68	LEU	HA	3.928	0
564	68	LEU	N	125.001	0.083
565	68	LEU	QB	1.647	0.011
566	68	LEU	QD1	0.867	0
567	68	LEU	QD2	0.424	0.003
568	69	HIS	CB	27.715	0
569	69	HIS	H	7.879	0.004
570	69	HIS	N	123.968	0.031
571	70	ARG	C	176.627	0.003
572	70	ARG	CA	53.679	0.026
573	70	ARG	CB	28.207	0.013
574	70	ARG	CD	40.683	0
575	70	ARG	CG	24.48	0
576	70	ARG	H	7.765	0
577	70	ARG	HA	4.332	0.002
578	70	ARG	HB2	1.866	0
579	70	ARG	HB3	1.812	0.049
580	70	ARG	N	118.396	0
581	70	ARG	QD	3.185	0

582	**70**	**ARG**	**QG**	**1.619**	**0.002**
583	**71**	**GLY**	**C**	**173.775**	**0.007**
584	**71**	**GLY**	**CA**	**42.754**	**0.07**
585	**71**	**GLY**	**H**	**8.39**	**0.008**
586	**71**	**GLY**	**HA**	**3.971**	**0.004**
587	**71**	**GLY**	**N**	**109.468**	**0.065**
588	**72**	**ASN**	**C**	**175.317**	**0.008**
589	**72**	**ASN**	**CA**	**50.701**	**0.071**
590	**72**	**ASN**	**CB**	**36.498**	**0.053**
591	**72**	**ASN**	**H**	**8.336**	**0.003**
592	**72**	**ASN**	**HA**	**4.804**	**0.002**
593	**72**	**ASN**	**HB2**	**2.846**	**0.003**
594	**72**	**ASN**	**HB3**	**2.775**	**0.005**
595	**72**	**ASN**	**HD21**	**6.3**	**0**
596	**72**	**ASN**	**HD22**	**7.303**	**0**
597	**72**	**ASN**	**N**	**118.604**	**0.088**
598	**72**	**ASN**	**ND2**	**108.596**	**0**
599	**73**	**THR**	**C**	**174.2**	**0.004**
600	**73**	**THR**	**CA**	**59.298**	**0.101**
601	**73**	**THR**	**CB**	**67.182**	**0.011**
602	**73**	**THR**	**CG2**	**19.003**	**0**
603	**73**	**THR**	**H**	**8.125**	**0.009**
604	**73**	**THR**	**HA**	**4.358**	**0.002**
605	**73**	**THR**	**HB**	**4.21**	**0.002**
606	**73**	**THR**	**HG2**	**1.175**	**0.003**
607	**73**	**THR**	**N**	**114.569**	**0.039**
608	**74**	**VAL**	**C**	**175.475**	**0.01**
609	**74**	**VAL**	**CA**	**59.532**	**0.057**
610	**74**	**VAL**	**CB**	**30.249**	**0.028**
611	**74**	**VAL**	**CG1**	**18.653**	**0**
612	**74**	**VAL**	**CG2**	**17.602**	**0**
613	**74**	**VAL**	**H**	**8.062**	**0.006**
614	**74**	**VAL**	**HA**	**4.137**	**0.004**
615	**74**	**VAL**	**HB**	**1.998**	**0.005**
616	**74**	**VAL**	**N**	**121.792**	**0.065**
617	**74**	**VAL**	**QG1**	**0.841**	**0**
618	**74**	**VAL**	**QG2**	**0.851**	**0.009**
619	**75**	**GLU**	**C**	**174.961**	**0.007**
620	**75**	**GLU**	**CA**	**53.784**	**0.022**
621	**75**	**GLU**	**CB**	**28.167**	**0.008**
622	**75**	**GLU**	**CG**	**33.652**	**0**
623	**75**	**GLU**	**H**	**8.254**	**0.005**
624	**75**	**GLU**	**HA**	**4.285**	**0.009**
625	**75**	**GLU**	**HB2**	**1.986**	**0.002**

626	**75**	**GLU**	**HB3**	**1.827**	**0.007**
627	**75**	**GLU**	**HG2**	**2.196**	**0.015**
628	**75**	**GLU**	**HG3**	**2.132**	**0.001**
629	**75**	**GLU**	**N**	**124.495**	**0.066**
630	**76**	**PHE**	**C**	**180.136**	**0**
631	**76**	**PHE**	**CA**	**56.436**	**0**
632	**76**	**PHE**	**CB**	**37.767**	**0**
633	**76**	**PHE**	**H**	**7.675**	**0.01**
634	**76**	**PHE**	**HA**	**4.424**	**0**
635	**76**	**PHE**	**HB2**	**3.164**	**0**
636	**76**	**PHE**	**HB3**	**2.95**	**0.012**
637	**76**	**PHE**	**N**	**125.704**	**0.05**
638	**77**	**ALA**	**C**	**181.06**	**0.001**
639	**77**	**ALA**	**CA**	**52.904**	**0.002**
640	**77**	**ALA**	**CB**	**14.736**	**0.03**
641	**77**	**ALA**	**H**	**8.717**	**0**
642	**77**	**ALA**	**HA**	**4.306**	**0**
643	**77**	**ALA**	**N**	**122.484**	**0**
644	**77**	**ALA**	**QB**	**1.413**	**0**
645	**78**	**LYS**	**C**	**179.67**	**0.005**
646	**78**	**LYS**	**CA**	**56.137**	**0.019**
647	**78**	**LYS**	**CB**	**29.319**	**0.023**
648	**78**	**LYS**	**CE**	**41.052**	**0**
649	**78**	**LYS**	**H**	**8.215**	**0.008**
650	**78**	**LYS**	**HA**	**4.47**	**0.004**
651	**78**	**LYS**	**N**	**119.264**	**0.029**
652	**78**	**LYS**	**QB**	**2.026**	**0.003**
653	**78**	**LYS**	**QE**	**2.657**	**0**
654	**79**	**GLN**	**C**	**176.39**	**0.003**
655	**79**	**GLN**	**CA**	**54.483**	**0.035**
656	**79**	**GLN**	**CB**	**25.611**	**0**
657	**79**	**GLN**	**CG**	**31.147**	**0**
658	**79**	**GLN**	**H**	**7.928**	**0.033**
659	**79**	**GLN**	**HA**	**4.103**	**0.007**
660	**79**	**GLN**	**HB2**	**2.198**	**0.009**
661	**79**	**GLN**	**HB3**	**1.921**	**0.001**
662	**79**	**GLN**	**HE21**	**6.696**	**0**
663	**79**	**GLN**	**HE22**	**7.294**	**0**
664	**79**	**GLN**	**HG2**	**2.431**	**0.001**
665	**79**	**GLN**	**HG3**	**2.29**	**0.019**
666	**79**	**GLN**	**N**	**118.724**	**0.076**
667	**79**	**GLN**	**NE2**	**110.613**	**0**
668	**80**	**HIS**	**C**	**174.963**	**0.023**
669	**80**	**HIS**	**CA**	**53.205**	**0.012**

670	80	HIS	CB	25.812	0.032
671	80	HIS	H	7.565	0.008
672	80	HIS	HA	4.539	0.002
673	80	HIS	HB2	3.458	0.006
674	80	HIS	HB3	2.548	0.005
675	80	HIS	N	115.884	0.021
676	81	GLY	C	174.862	0.015
677	81	GLY	H	7.772	0.003
678	81	GLY	HA2	4.29	0.001
679	81	GLY	HA3	3.732	0.002
680	81	GLY	N	106.911	0.051
681	81	GLY	CA	42.749	0.021
682	82	SER	C	173.634	0
683	82	SER	CA	55.965	0.046
684	82	SER	CB	61.338	0.076
685	82	SER	HA	4.541	0.025
686	82	SER	HB2	3.63	0.009
687	82	SER	HB3	3.4	0.006
688	82	SER	N	116.296	0.039
689	82	SER	H	8.307	0.002
690	83	ASP	C	176.299	0.06
691	83	ASP	CA	52.02	0.057
692	83	ASP	CB	38.449	0.036
693	83	ASP	H	8.423	0.006
694	83	ASP	HA	4.624	0.012
695	83	ASP	HB2	2.729	0
696	83	ASP	HB3	2.654	0.003
697	83	ASP	N	122.009	0.086
698	84	ASP	C	177.831	0
699	84	ASP	CA	55.639	0.098
700	84	ASP	CB	38.873	0.058
701	84	ASP	H	8.129	0.004
702	84	ASP	HA	4.559	0.002
703	84	ASP	N	120.518	0.051
704	84	ASP	QB	2.668	0.005
705	85	ALA	C	180.915	0.007
706	85	ALA	CA	52.803	0.07
707	85	ALA	CB	15.426	0.039
708	85	ALA	H	8.256	0.008
709	85	ALA	HA	4.226	0.01
710	85	ALA	N	122.122	0.047
711	85	ALA	QB	1.496	0.003
712	86	MET	C	177.762	0.002
713	86	MET	CA	55.861	0.071

714	**86**	**MET**	**CB**	**30.333**	**0.054**
715	**86**	**MET**	**CE**	**27.775**	**0**
716	**86**	**MET**	**H**	**8.509**	**0.009**
717	**86**	**MET**	**HA**	**4.122**	**0.007**
718	**86**	**MET**	**HB2**	**1.967**	**0.011**
719	**86**	**MET**	**HB3**	**1.885**	**0.012**
720	**86**	**MET**	**HG2**	**2.74**	**0.006**
721	**86**	**MET**	**HG3**	**2.655**	**0.007**
722	**86**	**MET**	**N**	**119.743**	**0.023**
723	**87**	**ALA**	**C**	**179.064**	**0.006**
724	**87**	**ALA**	**CA**	**52.531**	**0.072**
725	**87**	**ALA**	**CB**	**16.903**	**0.029**
726	**87**	**ALA**	**H**	**8.093**	**0.009**
727	**87**	**ALA**	**HA**	**3.963**	**0.032**
728	**87**	**ALA**	**N**	**120.527**	**0.021**
729	**87**	**ALA**	**QB**	**1.44**	**0.002**
730	**88**	**HIS**	**C**	**176.933**	**0.028**
731	**88**	**HIS**	**CA**	**57.32**	**0.017**
732	**88**	**HIS**	**CB**	**25.734**	**0.023**
733	**88**	**HIS**	**H**	**7.968**	**0.008**
734	**88**	**HIS**	**HA**	**4.188**	**0.004**
735	**88**	**HIS**	**HB2**	**3.384**	**0.004**
736	**88**	**HIS**	**HB3**	**3.217**	**0.003**
737	**88**	**HIS**	**N**	**112.784**	**0.03**
738	**89**	**GLN**	**C**	**178.118**	**0.004**
739	**89**	**GLN**	**CA**	**56.716**	**0.087**
740	**89**	**GLN**	**CB**	**26.025**	**0.073**
741	**89**	**GLN**	**CG**	**31.558**	**0**
742	**89**	**GLN**	**H**	**8.306**	**0.005**
743	**89**	**GLN**	**HA**	**4.262**	**0.017**
744	**89**	**GLN**	**HB2**	**2.428**	**0**
745	**89**	**GLN**	**HB3**	**2.361**	**0.001**
746	**89**	**GLN**	**HE21**	**6.508**	**0.002**
747	**89**	**GLN**	**HE22**	**6.991**	**0**
748	**89**	**GLN**	**N**	**120.522**	**0.062**
749	**89**	**GLN**	**NE2**	**109.308**	**0**
750	**89**	**GLN**	**QG**	**2.438**	**0.002**
751	**90**	**LEU**	**C**	**178.474**	**0.025**
752	**90**	**LEU**	**CA**	**57.226**	**0.024**
753	**90**	**LEU**	**CB**	**39.4**	**0.031**
754	**90**	**LEU**	**H**	**8.297**	**0.006**
755	**90**	**LEU**	**HA**	**4.072**	**0.005**
756	**90**	**LEU**	**HG**	**1.401**	**0**
757	**90**	**LEU**	**N**	**119.456**	**0.072**

758	**90**	**LEU**	**QB**	**2.065**	**0**
759	**90**	**LEU**	**QD1**	**0.88**	**0**
760	**90**	**LEU**	**QD2**	**0.708**	**0**
761	**91**	**VAL**	**C**	**177.386**	**0.004**
762	**91**	**VAL**	**CA**	**64.337**	**0.047**
763	**91**	**VAL**	**CB**	**28.834**	**0.035**
764	**91**	**VAL**	**CG1**	**19.529**	**0**
765	**91**	**VAL**	**H**	**7.359**	**0.007**
766	**91**	**VAL**	**HA**	**3.38**	**0.003**
767	**91**	**VAL**	**HB**	**2.125**	**0.005**
768	**91**	**VAL**	**N**	**116.975**	**0.022**
769	**91**	**VAL**	**QQG**	**0.875**	**0.001**
770	**92**	**ASP**	**C**	**179.988**	**0.004**
771	**92**	**ASP**	**CA**	**55.01**	**0.069**
772	**92**	**ASP**	**CB**	**37.423**	**0.028**
773	**92**	**ASP**	**H**	**8.077**	**0.007**
774	**92**	**ASP**	**HA**	**4.416**	**0.005**
775	**92**	**ASP**	**HB2**	**2.947**	**0.007**
776	**92**	**ASP**	**HB3**	**2.771**	**0.024**
777	**92**	**ASP**	**N**	**120.39**	**0.031**
778	**93**	**ILE**	**C**	**178.526**	**0.005**
779	**93**	**ILE**	**CA**	**62.956**	**0.047**
780	**93**	**ILE**	**CB**	**35.516**	**0.021**
781	**93**	**ILE**	**CG1**	**27.002**	**0**
782	**93**	**ILE**	**CG2**	**16.561**	**0**
783	**93**	**ILE**	**H**	**8.339**	**0.008**
784	**93**	**ILE**	**HA**	**3.845**	**0.004**
785	**93**	**ILE**	**HB**	**2.202**	**0.003**
786	**93**	**ILE**	**N**	**121.761**	**0.038**
787	**93**	**ILE**	**QG1**	**1.288**	**0**
788	**93**	**ILE**	**QG2**	**1.011**	**0.009**
789	**94**	**VAL**	**C**	**177.834**	**0.007**
790	**94**	**VAL**	**CA**	**65.851**	**0.066**
791	**94**	**VAL**	**CB**	**28.706**	**0.111**
792	**94**	**VAL**	**CG1**	**21.264**	**0**
793	**94**	**VAL**	**CG2**	**17.936**	**0**
794	**94**	**VAL**	**H**	**7.947**	**0.006**
795	**94**	**VAL**	**HA**	**3.347**	**0.011**
796	**94**	**VAL**	**HB**	**2.049**	**0.011**
797	**94**	**VAL**	**N**	**119.911**	**0.051**
798	**94**	**VAL**	**QQG**	**0.739**	**0.006**
799	**95**	**HIS**	**C**	**177.648**	**0.01**
800	**95**	**HIS**	**CA**	**54.839**	**0.03**
801	**95**	**HIS**	**CB**	**25.734**	**0.01**

802	**95**	**HIS**	**H**	**8.029**	**0.005**
803	**95**	**HIS**	**HA**	**4.763**	**0.012**
804	**95**	**HIS**	**HB2**	**3.36**	**0**
805	**95**	**HIS**	**HB3**	**3.305**	**0.002**
806	**95**	**HIS**	**N**	**115.5**	**0.054**
807	**96**	**ALA**	**C**	**181.906**	**0.002**
808	**96**	**ALA**	**CA**	**52.797**	**0.063**
809	**96**	**ALA**	**CB**	**15.348**	**0.018**
810	**96**	**ALA**	**H**	**8.697**	**0.007**
811	**96**	**ALA**	**HA**	**4.278**	**0.008**
812	**96**	**ALA**	**N**	**123.747**	**0.018**
813	**96**	**ALA**	**QB**	**1.607**	**0.004**
814	**97**	**CYS**	**C**	**176.679**	**0.005**
815	**97**	**CYS**	**CA**	**55.31**	**0.074**
816	**97**	**CYS**	**CB**	**36.326**	**0.002**
817	**97**	**CYS**	**H**	**8.756**	**0.006**
818	**97**	**CYS**	**HA**	**4.741**	**0.009**
819	**97**	**CYS**	**HB2**	**3.46**	**0.013**
820	**97**	**CYS**	**HB3**	**2.963**	**0.02**
821	**97**	**CYS**	**N**	**119.19**	**0.021**
822	**98**	**GLU**	**C**	**177.927**	**0.006**
823	**98**	**GLU**	**CA**	**57.134**	**0.023**
824	**98**	**GLU**	**CB**	**27.058**	**0.01**
825	**98**	**GLU**	**CG**	**34.672**	**0**
826	**98**	**GLU**	**H**	**8.462**	**0.005**
827	**98**	**GLU**	**HA**	**3.926**	**0.002**
828	**98**	**GLU**	**HB2**	**2.295**	**0.002**
829	**98**	**GLU**	**HB3**	**2.119**	**0.005**
830	**98**	**GLU**	**HG2**	**2.661**	**0.023**
831	**98**	**GLU**	**HG3**	**2.46**	**0.004**
832	**98**	**GLU**	**N**	**119.228**	**0.038**
833	**99**	**LYS**	**C**	**177.457**	**0.011**
834	**99**	**LYS**	**CA**	**54.602**	**0.024**
835	**99**	**LYS**	**CB**	**30.536**	**0.017**
836	**99**	**LYS**	**CD**	**26.249**	**0**
837	**99**	**LYS**	**CE**	**42.933**	**0**
838	**99**	**LYS**	**CG**	**22.465**	**0**
839	**99**	**LYS**	**H**	**7.414**	**0.005**
840	**99**	**LYS**	**HA**	**4.364**	**0.006**
841	**99**	**LYS**	**HB2**	**2.01**	**0.007**
842	**99**	**LYS**	**HB3**	**1.95**	**0.001**
843	**99**	**LYS**	**N**	**115.275**	**0.016**
844	**99**	**LYS**	**QD**	**1.614**	**0.051**
845	**99**	**LYS**	**QE**	**2.958**	**0.058**

846	**99**	**LYS**	**QG**	**1.631**	**0.057**
847	**100**	**SER**	**C**	**174.87**	**0.004**
848	**100**	**SER**	**CA**	**57.038**	**0.052**
849	**100**	**SER**	**CB**	**61.643**	**0.048**
850	**100**	**SER**	**H**	**7.973**	**0.005**
851	**100**	**SER**	**HA**	**4.455**	**0.003**
852	**100**	**SER**	**HB2**	**4.033**	**0.007**
853	**100**	**SER**	**HB3**	**3.917**	**0.012**
854	**100**	**SER**	**N**	**114.257**	**0.05**
855	**101**	**VAL**	**C**	**173.358**	**0**
856	**101**	**VAL**	**CA**	**57.812**	**0**
857	**101**	**VAL**	**CB**	**29.628**	**0**
858	**101**	**VAL**	**H**	**7.553**	**0.005**
859	**101**	**VAL**	**HA**	**4.332**	**0.008**
860	**101**	**VAL**	**HB**	**2.065**	**0.013**
861	**101**	**VAL**	**N**	**124.244**	**0.039**
862	**101**	**VAL**	**QQG**	**1.014**	**0.007**
863	**103**	**PRO**	**C**	**176.197**	**0.008**
864	**103**	**PRO**	**CA**	**60.749**	**0.034**
865	**103**	**PRO**	**CB**	**29.364**	**0.083**
866	**103**	**PRO**	**CD**	**45.357**	**0**
867	**103**	**PRO**	**CG**	**24.696**	**0**
868	**103**	**PRO**	**HA**	**4.312**	**0.002**
869	**103**	**PRO**	**HD2**	**3.64**	**0.001**
870	**103**	**PRO**	**HD3**	**3.412**	**0**
871	**103**	**PRO**	**QB**	**2.268**	**0**
872	**103**	**PRO**	**QG**	**1.872**	**0**
873	**104**	**ASN**	**C**	**174.386**	**0.001**
874	**104**	**ASN**	**CA**	**50.256**	**0.031**
875	**104**	**ASN**	**CB**	**39.648**	**0.028**
876	**104**	**ASN**	**H**	**8.002**	**0.006**
877	**104**	**ASN**	**HA**	**4.681**	**0.006**
878	**104**	**ASN**	**HB2**	**2.659**	**0.013**
879	**104**	**ASN**	**HB3**	**2.556**	**0.006**
880	**104**	**ASN**	**N**	**122.086**	**0.063**
881	**104**	**ASN**	**QD2**	**6.933**	**0**
882	**105**	**GLU**	**C**	**175.802**	**0.003**
883	**105**	**GLU**	**CA**	**55.303**	**0.041**
884	**105**	**GLU**	**CB**	**27.168**	**0.026**
885	**105**	**GLU**	**CG**	**33.748**	**0**
886	**105**	**GLU**	**H**	**8.764**	**0.005**
887	**105**	**GLU**	**HA**	**4.118**	**0.005**
888	**105**	**GLU**	**HB2**	**2.104**	**0.006**
889	**105**	**GLU**	**HB3**	**1.986**	**0.004**

890	**105**	**GLU**	**N**	**124.584**	**0.017**
891	**105**	**GLU**	**QG**	**2.318**	**0.002**
892	**106**	**ASP**	**C**	**176.52**	**0.008**
893	**106**	**ASP**	**CA**	**50.252**	**0.013**
894	**106**	**ASP**	**CB**	**39.048**	**0.003**
895	**106**	**ASP**	**H**	**8.798**	**0.006**
896	**106**	**ASP**	**HA**	**4.72**	**0.007**
897	**106**	**ASP**	**HB2**	**3.003**	**0.005**
898	**106**	**ASP**	**HB3**	**2.543**	**0.01**
899	**106**	**ASP**	**N**	**119.688**	**0.034**
900	**107**	**ASN**	**C**	**178.217**	**0.006**
901	**107**	**ASN**	**CA**	**53.412**	**0.022**
902	**107**	**ASN**	**CB**	**35.063**	**0.053**
903	**107**	**ASN**	**H**	**8.847**	**0.005**
904	**107**	**ASN**	**HA**	**4.385**	**0.006**
905	**107**	**ASN**	**HB2**	**2.877**	**0.007**
906	**107**	**ASN**	**HB3**	**2.798**	**0.003**
907	**107**	**ASN**	**N**	**125.522**	**0.033**
908	**108**	**CYS**	**C**	**176.508**	**0.007**
909	**108**	**CYS**	**CA**	**58.526**	**0.019**
910	**108**	**CYS**	**CB**	**41.589**	**0.079**
911	**108**	**CYS**	**H**	**8.453**	**0.006**
912	**108**	**CYS**	**HA**	**4.119**	**0.007**
913	**108**	**CYS**	**HB2**	**3.88**	**0.007**
914	**108**	**CYS**	**HB3**	**3.04**	**0.007**
915	**108**	**CYS**	**N**	**118.907**	**0.047**
916	**109**	**LEU**	**C**	**180.971**	**0.01**
917	**109**	**LEU**	**CA**	**54.584**	**0.056**
918	**109**	**LEU**	**CB**	**38.103**	**0.034**
919	**109**	**LEU**	**CD1**	**24.555**	**0**
920	**109**	**LEU**	**H**	**7.56**	**0.008**
921	**109**	**LEU**	**HA**	**4.13**	**0.003**
922	**109**	**LEU**	**HG**	**1.484**	**0.038**
923	**109**	**LEU**	**N**	**119.646**	**0.02**
924	**109**	**LEU**	**QB**	**1.867**	**0.005**
925	**110**	**MET**	**C**	**177.435**	**0.012**
926	**110**	**MET**	**CA**	**55.758**	**0.049**
927	**110**	**MET**	**CB**	**29.027**	**0.072**
928	**110**	**MET**	**H**	**8.197**	**0.005**
929	**110**	**MET**	**HA**	**4.36**	**0.008**
930	**110**	**MET**	**HB2**	**2.243**	**0.004**
931	**110**	**MET**	**HB3**	**2.175**	**0.005**
932	**110**	**MET**	**HG2**	**2.676**	**0.007**
933	**110**	**MET**	**HG3**	**2.632**	**0.001**

934	**110**	**MET**	**N**	**121.689**	**0.038**
935	**111**	**ALA**	**C**	**180.608**	**0.003**
936	**111**	**ALA**	**CA**	**52.914**	**0.041**
937	**111**	**ALA**	**CB**	**15.565**	**0.031**
938	**111**	**ALA**	**H**	**7.738**	**0.005**
939	**111**	**ALA**	**HA**	**4.003**	**0.006**
940	**111**	**ALA**	**N**	**120.78**	**0.017**
941	**111**	**ALA**	**QB**	**1.444**	**0.001**
942	**112**	**LEU**	**C**	**178.606**	**0.013**
943	**112**	**LEU**	**CA**	**55.824**	**0.066**
944	**112**	**LEU**	**CB**	**40.244**	**0.05**
945	**112**	**LEU**	**CG**	**27.144**	**0**
946	**112**	**LEU**	**H**	**7.824**	**0.004**
947	**112**	**LEU**	**HA**	**3.884**	**0.007**
948	**112**	**LEU**	**HG**	**1.506**	**0.001**
949	**112**	**LEU**	**N**	**119.288**	**0.024**
950	**112**	**LEU**	**QB**	**1.946**	**0.002**
951	**112**	**LEU**	**QD1**	**0.896**	**0**
952	**112**	**LEU**	**QD2**	**0.674**	**0**
953	**113**	**GLY**	**C**	**177.612**	**0.007**
954	**113**	**GLY**	**CA**	**44.849**	**0.033**
955	**113**	**GLY**	**H**	**8.224**	**0.007**
956	**113**	**GLY**	**HA**	**3.902**	**0.015**
957	**113**	**GLY**	**N**	**106.725**	**0.037**
958	**114**	**ILE**	**C**	**177.581**	**0.081**
959	**114**	**ILE**	**CA**	**63.614**	**0.08**
960	**114**	**ILE**	**CB**	**35.472**	**0.098**
961	**114**	**ILE**	**CD1**	**13.787**	**0**
962	**114**	**ILE**	**CG1**	**27.029**	**0**
963	**114**	**ILE**	**H**	**8.879**	**0.007**
964	**114**	**ILE**	**HA**	**3.581**	**0.006**
965	**114**	**ILE**	**HB**	**1.862**	**0.018**
966	**114**	**ILE**	**N**	**124.077**	**0.029**
967	**114**	**ILE**	**QD1**	**0.757**	**0**
968	**114**	**ILE**	**QG1**	**0.909**	**0.009**
969	**114**	**ILE**	**QG2**	**1.843**	**0**
970	**115**	**SER**	**C**	**176.441**	**0.009**
971	**115**	**SER**	**CA**	**60.22**	**0.032**
972	**115**	**SER**	**CB**	**64.265**	**0**
973	**115**	**SER**	**H**	**8.199**	**0.005**
974	**115**	**SER**	**HA**	**4.569**	**0.013**
975	**115**	**SER**	**HB2**	**4.05**	**0.013**
976	**115**	**SER**	**HB3**	**3.885**	**0.005**
977	**115**	**SER**	**N**	**116.394**	**0.021**

978	**116**	**MET**	**C**	**178.988**	**0.004**
979	**116**	**MET**	**CA**	**53.87**	**0.08**
980	**116**	**MET**	**CB**	**29.883**	**0.063**
981	**116**	**MET**	**CG**	**29.558**	**0**
982	**116**	**MET**	**H**	**8.276**	**0.002**
983	**116**	**MET**	**HA**	**4.686**	**0.013**
984	**116**	**MET**	**HB2**	**2.272**	**0**
985	**116**	**MET**	**HB3**	**2.039**	**0**
986	**116**	**MET**	**N**	**120.508**	**0.035**
987	**116**	**MET**	**QG**	**2.274**	**0.006**
988	**117**	**CYS**	**C**	**175.191**	**0.003**
989	**117**	**CYS**	**CA**	**57.475**	**0.031**
990	**117**	**CYS**	**CB**	**34.763**	**0.025**
991	**117**	**CYS**	**H**	**8.019**	**0.006**
992	**117**	**CYS**	**HA**	**4.405**	**0.002**
993	**117**	**CYS**	**HB2**	**3.458**	**0.002**
994	**117**	**CYS**	**HB3**	**3.21**	**0.007**
995	**117**	**CYS**	**N**	**123.861**	**0.031**
996	**118**	**PHE**	**C**	**176.203**	**0.015**
997	**118**	**PHE**	**CA**	**58.992**	**0.057**
998	**118**	**PHE**	**CB**	**37.281**	**0.02**
999	**118**	**PHE**	**H**	**8.897**	**0.007**
1000	**118**	**PHE**	**HA**	**4.021**	**0.006**
1001	**118**	**PHE**	**N**	**122.917**	**0.035**
1002	**118**	**PHE**	**QB**	**3.056**	**0.005**
1003	**118**	**PHE**	**QD**	**7.189**	**0**
1004	**119**	**LYS**	**C**	**177.986**	**0.007**
1005	**119**	**LYS**	**CA**	**57.569**	**0.025**
1006	**119**	**LYS**	**CB**	**30.274**	**0.034**
1007	**119**	**LYS**	**CD**	**26.912**	**0**
1008	**119**	**LYS**	**CE**	**43.036**	**0**
1009	**119**	**LYS**	**CG**	**22.555**	**0**
1010	**119**	**LYS**	**H**	**8.555**	**0.006**
1011	**119**	**LYS**	**HA**	**3.853**	**0.004**
1012	**119**	**LYS**	**HB2**	**1.909**	**0.002**
1013	**119**	**LYS**	**HB3**	**1.792**	**0.002**
1014	**119**	**LYS**	**N**	**118.467**	**0.027**
1015	**119**	**LYS**	**QD**	**1.792**	**0**
1016	**119**	**LYS**	**QG**	**1.244**	**0.005**
1017	**120**	**THR**	**C**	**176.908**	**0.017**
1018	**120**	**THR**	**CA**	**64.083**	**0.011**
1019	**120**	**THR**	**CB**	**66.424**	**0.025**
1020	**120**	**THR**	**CG2**	**22.162**	**0**
1021	**120**	**THR**	**H**	**8.002**	**0.006**

1022	120	THR	HA	3.908	0.004
1023	120	THR	HB	4.442	0.006
1024	120	THR	N	114.124	0.008
1025	120	THR	QG2	1.252	0.003
1026	121	GLU	C	176.064	0.022
1027	121	GLU	CA	56.133	0.01
1028	121	GLU	CB	27.383	0.033
1029	121	GLU	CG	36.361	0
1030	121	GLU	H	8.305	0.004
1031	121	GLU	HA	4.682	0.013
1032	121	GLU	N	120.499	0.023
1033	121	GLU	QB	2.336	0.019
1034	121	GLU	QG	3.015	0.013
1035	122	ILE	C	178.966	0.007
1036	122	ILE	CA	59.581	0.1
1037	122	ILE	CB	59.747	0
1038	122	ILE	H	8.478	0.007
1039	122	ILE	HA	3.586	0.002
1040	122	ILE	HB	1.912	0
1041	122	ILE	N	119.516	0.039
1042	122	ILE	QD1	0.362	0
1043	122	ILE	QG1	1.071	0
1044	122	ILE	QG2	0.743	0
1045	123	HIS	C	177.17	0.014
1046	123	HIS	CA	56.148	0.063
1047	123	HIS	CB	26.809	0.029
1048	123	HIS	H	7.973	0.008
1049	123	HIS	HA	4.479	0.003
1050	123	HIS	N	119.284	0.053
1051	123	HIS	QB	3.367	0.008
1052	124	LYS	C	177.376	0.013
1053	124	LYS	CA	55.628	0.03
1054	124	LYS	CB	30.351	0.033
1055	124	LYS	CD	27.04	0
1056	124	LYS	CE	39.717	0
1057	124	LYS	CG	22.722	0
1058	124	LYS	H	7.411	0.019
1059	124	LYS	HA	4.077	0.005
1060	124	LYS	HB2	1.961	0.017
1061	124	LYS	HB3	1.851	0.019
1062	124	LYS	N	117.298	0.133
1063	124	LYS	QD	1.442	0.002
1064	124	LYS	QG	1.69	0.005
1065	125	LEU	C	176.985	0.005

1066	**125**	**LEU**	**CA**	**52.084**	**0.021**
1067	**125**	**LEU**	**CB**	**39.083**	**0.06**
1068	**125**	**LEU**	**H**	**7.316**	**0.004**
1069	**125**	**LEU**	**HA**	**4.171**	**0.003**
1070	**125**	**LEU**	**HB2**	**1.825**	**0.335**
1071	**125**	**LEU**	**HB3**	**1.592**	**0.363**
1072	**125**	**LEU**	**HG**	**1.962**	**0**
1073	**125**	**LEU**	**N**	**117.857**	**0.051**
1074	**125**	**LEU**	**QD1**	**0.826**	**0.119**
1075	**125**	**LEU**	**QD2**	**0.356**	**0**
1076	**126**	**ASN**	**C**	**174.844**	**0.021**
1077	**126**	**ASN**	**CA**	**51.333**	**0.026**
1078	**126**	**ASN**	**CB**	**34.889**	**0.043**
1079	**126**	**ASN**	**H**	**8.097**	**0.01**
1080	**126**	**ASN**	**HA**	**4.539**	**0.003**
1081	**126**	**ASN**	**HB2**	**3.03**	**0.014**
1082	**126**	**ASN**	**HB3**	**2.759**	**0.011**
1083	**126**	**ASN**	**N**	**116.667**	**0.036**
1084	**128**	**ALA**	**C**	**176.134**	**0**
1085	**128**	**ALA**	**CA**	**51.312**	**0**
1086	**128**	**ALA**	**CB**	**17.466**	**0**
1087	**128**	**ALA**	**H**	**8.076**	**0.007**
1088	**128**	**ALA**	**HA**	**4.155**	**0**
1089	**128**	**ALA**	**N**	**131.733**	**0.022**
1090	**129**	**PRO**	**C**	**176.661**	**0.008**
1091	**129**	**PRO**	**CA**	**60.49**	**0.025**
1092	**129**	**PRO**	**CB**	**29.4**	**0.023**
1093	**129**	**PRO**	**CD**	**47.8**	**0**
1094	**129**	**PRO**	**CG**	**24.826**	**0**
1095	**129**	**PRO**	**HA**	**4.447**	**0.003**
1096	**129**	**PRO**	**QB**	**2.294**	**0**
1097	**129**	**PRO**	**QD**	**2.82**	**0**
1098	**129**	**PRO**	**QG**	**1.941**	**0**
1099	**130**	**ASP**	**C**	**175.248**	**0.006**
1100	**130**	**ASP**	**CA**	**50.792**	**0.035**
1101	**130**	**ASP**	**CB**	**36.399**	**0.02**
1102	**130**	**ASP**	**H**	**8.371**	**0.003**
1103	**130**	**ASP**	**HA**	**4.667**	**0.004**
1104	**130**	**ASP**	**N**	**118.029**	**0.03**
1105	**130**	**ASP**	**QB**	**2.804**	**0.007**
1106	**131**	**HIS**	**C**	**176.074**	**0.01**
1107	**131**	**HIS**	**CA**	**53.875**	**0.042**
1108	**131**	**HIS**	**CB**	**27.6**	**0.06**
1109	**131**	**HIS**	**H**	**8.376**	**0.009**

1110	**131**	**HIS**	**HA**	**4.376**	**0.006**
1111	**131**	**HIS**	**HB2**	**2.24**	**0.014**
1112	**131**	**HIS**	**HB3**	**1.933**	**0.01**
1113	**131**	**HIS**	**N**	**121.147**	**0.03**
1114	**132**	**GLU**	**C**	**176.063**	**0.068**
1115	**132**	**GLU**	**CA**	**54.123**	**0.057**
1116	**132**	**GLU**	**CB**	**27.744**	**0.068**
1117	**132**	**GLU**	**CG**	**33.85**	**0**
1118	**132**	**GLU**	**H**	**8.519**	**0.003**
1119	**132**	**GLU**	**HA**	**4.251**	**0.01**
1120	**132**	**GLU**	**N**	**122.204**	**0.152**
1121	**132**	**GLU**	**QB**	**1.963**	**0**
1122	**132**	**GLU**	**QG**	**2.469**	**0.209**
1123	**133**	**LEU**	**C**	**175.945**	**0.022**
1124	**133**	**LEU**	**CA**	**52.492**	**0.067**
1125	**133**	**LEU**	**CB**	**39.468**	**0.058**
1126	**133**	**LEU**	**H**	**8.177**	**0.002**
1127	**133**	**LEU**	**HA**	**4.376**	**0.001**
1128	**133**	**LEU**	**N**	**123.523**	**0.026**
1129	**133**	**LEU**	**QB**	**1.626**	**0.005**
1130	**133**	**LEU**	**QQD**	**0.861**	**0**
1131	**134**	**LEU**	**C**	**176.165**	**0**
1132	**134**	**LEU**	**CA**	**54.072**	**0**
1133	**134**	**LEU**	**CB**	**40.954**	**0**
1134	**134**	**LEU**	**H**	**7.673**	**0.006**
1135	**134**	**LEU**	**HA**	**4.209**	**0**
1136	**134**	**LEU**	**N**	**128.855**	**0.029**
1137	**134**	**LEU**	**QB**	**1.576**	**0**
1138	**134**	**LEU**	**QQD**	**0.886**	**0**
1139	**135**	**LEU**	**H**	**8.217**	**0**
1140	**135**	**LEU**	**N**	**119.407**	**0**
1141	**136**	**GLU**	**H**	**8.452**	**0.001**
1142	**136**	**GLU**	**N**	**119.585**	**0.008**
1143	**137**	**GLU**	**C**	**177.599**	**0.023**
1144	**137**	**GLU**	**CA**	**56.552**	**0.056**
1145	**137**	**GLU**	**CB**	**25.444**	**0.041**
1146	**137**	**GLU**	**CG**	**31.6**	**0**
1147	**137**	**GLU**	**H**	**8**	**0**
1148	**137**	**GLU**	**HA**	**3.7**	**0.001**
1149	**137**	**GLU**	**N**	**119.157**	**0**
1150	**137**	**GLU**	**QB**	**2.488**	**0**
1151	**138**	**MET**	**C**	**178.206**	**0.01**
1152	**138**	**MET**	**CA**	**56.592**	**0.098**
1153	**138**	**MET**	**CB**	**30.074**	**0.014**

1154	**138**	**MET**	**H**	**7.986**	**0.007**
1155	**138**	**MET**	**HA**	**3.517**	**0.344**
1156	**138**	**MET**	**N**	**118.487**	**0.012**
1157	**138**	**MET**	**QB**	**1.912**	**0.007**
1158	**139**	**MET**	**C**	**177.541**	**0.043**
1159	**139**	**MET**	**CA**	**56.989**	**0.089**
1160	**139**	**MET**	**CB**	**30.606**	**0.162**
1161	**139**	**MET**	**H**	**8.227**	**0.005**
1162	**139**	**MET**	**HA**	**3.835**	**0.009**
1163	**139**	**MET**	**N**	**116.061**	**0.028**
1164	**139**	**MET**	**QB**	**2.149**	**0.01**
1165	**139**	**MET**	**QG**	**2.937**	**0.039**
1166	**140**	**ALA**	**C**	**177.799**	**0.013**
1167	**140**	**ALA**	**CA**	**51.527**	**0.125**
1168	**140**	**ALA**	**CB**	**16.121**	**0.02**
1169	**140**	**ALA**	**H**	**7.739**	**0.007**
1170	**140**	**ALA**	**HA**	**4.044**	**0.009**
1171	**140**	**ALA**	**N**	**118.69**	**0.032**
1172	**140**	**ALA**	**QB**	**1.298**	**0.003**
1173	**141**	**GLU**	**C**	**176.7**	**0.005**
1174	**141**	**GLU**	**CA**	**52.828**	**0.021**
1175	**141**	**GLU**	**CB**	**28.019**	**0.024**
1176	**141**	**GLU**	**CG**	**33.06**	**0**
1177	**141**	**GLU**	**H**	**7.33**	**0.006**
1178	**141**	**GLU**	**HA**	**4.373**	**0.01**
1179	**141**	**GLU**	**N**	**115.489**	**0.031**
1180	**141**	**GLU**	**QB**	**2.008**	**0.02**
1181	**141**	**GLU**	**QG**	**2.353**	**0.01**
1182	**142**	**MET**	**C**	**175.52**	**0**
1183	**142**	**MET**	**CA**	**54.161**	**0**
1184	**142**	**MET**	**CB**	**31.376**	**0**
1185	**142**	**MET**	**H**	**7.538**	**0.008**
1186	**142**	**MET**	**HA**	**4.128**	**0**
1187	**142**	**MET**	**N**	**119.523**	**0.045**
1188	**143**	**LYS**	**C**	**175.674**	**0.007**
1189	**143**	**LYS**	**CA**	**54.209**	**0.036**
1190	**143**	**LYS**	**CB**	**30.186**	**0.028**
1191	**143**	**LYS**	**CD**	**27.998**	**0**
1192	**143**	**LYS**	**CE**	**39.555**	**0**
1193	**143**	**LYS**	**CG**	**26.544**	**0**
1194	**143**	**LYS**	**H**	**7.877**	**0**
1195	**143**	**LYS**	**HA**	**4.22**	**0.002**
1196	**143**	**LYS**	**N**	**126.923**	**0.025**
1197	**143**	**LYS**	**QB**	**1.845**	**0**

1198	**143**	**LYS**	**QD**	**2.131**	**0**
1199	**143**	**LYS**	**QE**	**2.88**	**0**
1200	**143**	**LYS**	**QG**	**1.497**	**0.001**
1201	**144**	**GLN**	**C**	**180.643**	**0**
1202	**144**	**GLN**	**CA**	**54.394**	**0**
1203	**144**	**GLN**	**CB**	**31.35**	**0**
1204	**144**	**GLN**	**H**	**7.876**	**0.005**
1205	**144**	**GLN**	**HA**	**4.132**	**0.006**
1206	**144**	**GLN**	**HE21**	**6.7**	**0**
1207	**144**	**GLN**	**HE22**	**7.291**	**0**
1208	**144**	**GLN**	**HG2**	**2.253**	**0.007**
1209	**144**	**GLN**	**HG3**	**2.11**	**0.009**
1210	**144**	**GLN**	**N**	**126.87**	**0.044**
1211	**144**	**GLN**	**NE2**	**111.075**	**0**
1212	**144**	**GLN**	**QB**	**1.849**	**0.003**

Appendix Table A-3 Summary of NMR experiments and protein samples prepared for the studies in this dissertation

	Experiments	Protein sample
Backbone assignments	3D-HN(CO)CACB,3D-HNCACB 3D-HNCO, 3D-HN(CA)CO, 3D-HNCA, 3D-HN(CO)CA	{ ^{13}C, ^{15}N}-double labeled protein
Side-chain Assignment	3D-TOCSY-HSQC 3D-HCCH-TOCSY, 3D-HCC(CO)NH , 3D-H(CCCO)NH	{ ^{13}C, ^{15}N}-double labeled protein sample
NOE Assignment	3D- NOESY-HSQC, 3D-^{13}C-NOESY-HSQC (Aliphatic Region) 3D-^{13}C-NOESY-HSQC (Aromatic)	{ ^{13}C, ^{15}N}-double labeled protein sample
NMR titration studies	2D ^{15}N HSQC	^{15}N- labeled protein

VITA

Salik Ram Dahal

Candidate for the Degree of Doctor of Philosophy

Dissertation: STRUCTURE AND FUNCTIONAL CHARACTERIZATION OF THE PHEROMONE BINDING PROTEIN 2 FROM *OSTRINIA FURNACALIS*

Major Field: Chemistry

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Chemistry at Oklahoma State University, Stillwater, Oklahoma in May, 2021.

Completed the requirements for the Master of Science in Chemistry at Tribhuvan University, Nepal in 2013.

Completed the requirements for the Bachelor of Science Chemistry at Tribhuvan University, Nepal, in 2009.

Experience:

Teaching Assistant: Oklahoma State University 2015-2021

Teaching Faculty: Tribhuvan University, Nepal, 2013-2015

Cambridge University A-Level Chemistry Lecturer: Kathmandu University, 2012/13

Professional Memberships:

Member of Nepal Chemical Society, 2011

General Secretory, Electrochemical society, Oklahoma Chapter, 2019

General Member of American Chemical Society, 2019