OVEREXPRESSION, CHARACTERIZATION, AND STRUCTURE-FUNCTION STUDIES OF PHEROMONE BINDING PROTEINS FROM EUROPEAN CORN BORER, OSTRINIA NUBILALIS

By

OMAR I. AL-DANOON

Bachelor of Science in Chemistry
University of Mosul
Mosul, Iraq
2007

Master of Science in Chemistry
Florida Institute of Technology
Melbourne, Florida
2014

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
July, 2021

OVEREXPRESSION, CHARACTERIZATION, AND STRUCTURE-FUNCTION STUDIES OF PHEROMONE BINDING PROTEINS FROM EUROPEAN CORN BORER, OSTRINIA NUBILALIS

Dissertation Approved:

Dr. Smita Mohanty

Dissertation Adviser

Dr. Richard Bunce

Dr. Barry Lavine

Dr. Ziad El Rassi

Dr. Junpeng Deng

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Smita Mohanty for the guidance, and support during my research. Thanks also to my other committee members: Dr. Richard Bunce, Dr. Barry Lavine, Dr. Ziad El Rassi, and Dr. Junpeng Deng. My appreciation is also given to Dr.Trever Holland for proofreading of the dissertation. I would like to thank Dr. Suman Mazumder and Dr. Bharat Chaudhary for their aid during the research. Special thanks to my mother, brothers, and sister for their encouragement and support.

Name: OMAR I. AL-DANOON

Date of Degree: JULY, 2021

Title of Study: OVEREXPRESSION, CHARACTERIZATION, AND STRUCTURE-FUNCTION STUDIES OF PHEROMONE BINDING PROTEINS FROM EUROPEAN CORN BORER, OSTRINIA NUBILALIS

Major Field: CHEMISTRY

Abstract: *Ostrinia nubilalis* (European corn borer) has a huge impact on production of economically important crops in the USA and Europe. This species, similar to other lepidopterans, relies on olfaction to find their mating partners. The moth male is invited by the female through a chemical signal started with releasing sex pheromones. Pheromone binding proteins (PBPs) are a key component to bind and deliver the received sex pheromones to olfactory receptors in male moth antennae. The detailed structural and mechanistic studies will aid to biorationally design pheromone mimetics for an effective pest management. OnubPBP3 is one of the *Ostrinia nubilalis* PBPs and it is expressed at higher levels in the antennae of males than in the females. Here, we report the overexpression, refolding, purification, and biophysical characterization of the recombinant OnubPBP3 by using different techniques including fluorescence spectroscopy, circular dichroism (CD), small angle X-ray scattering (SAXS), and high-resolution solution NMR. We determined the binding affinity of the recombinant OnubPBP3 to both *Ostrinia nubilalis* pheromones, *E*11-tetradecenyl acetate and *Z*11-tetradecenyl acetate. The recombinant protein showed nanomolar affinity to each isomer of the *Ostrinia nubilalis* pheromones. Ligand-induce conformational change was observed in OnubPBP3 by using NMR spectroscopy. The high-resolution three-dimensional structure of OnubPBP3 was determined by using solution NMR. The structure of OnubPBP3 consists of eight α-helices with residues 3–9 (α1a), 15–23 (α1b), 27–35 (α2), 48–57 (α3), 72–79 (α4), 84–98 (α5), 114–116 (α6a), and 120–123 (α6b), held together by three disulfide bridges: 19–54, 50–107 and 97–116 forming a large hydrophobic pocket inside the protein. The pheromone binding site of the protein was detected by molecular docking. Molecular dynamics simulation study demonstrates ligand-induced conformational and flexibility changes in the protein with both *E* and *Z*-pheromone. Characterization of pH-dependent conformational changes of OnubPBP3 via CD, NMR spectroscopy, and SAXS were shown a distinguishable behavior as compared to the well-investigated lepidopteran PBPs at the same pH. Our findings suggest that OnubPBP3 in acidic conditions is in a molten globule state.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

.

CHAPTER I

LITERATURE REVIEW

1.1 Nuclear magnetic resonance (NMR) spectroscopy

1.1.1 Introduction

Nuclear magnetic resonance spectroscopy (NMR) is one of the most important tools used in biology, chemistry, and physics for determining high-resolution molecular structures and studying molecular dynamics. It was discovered individually by two groups of scientists, Felix Bloch and coworkers at Stanford University and Edward Purcell and coworkers at Harvard University.[1] Together, they were awarded a Nobel Prize in physics in 1952. NMR spectroscopy depends on the interaction of atomic nuclei of specific isotopes subjected to a high external magnetic field with radio-frequency electromagnetic radiation. The chemical environment of NMR-active isotopes of elements plays a crucial role in the structural determination of micro and macromolecules. NMR-active isotopes at the same external magnetic field and in different chemical environments produce different resonance signals. In the early years, NMR spectroscopy was used to characterize small organic compounds. In 1957, the first NMR spectrum of bovine pancreatic ribonuclease was collected showing four poorly resolved signals.[2] The low sensitivity of early NMR spectrometers was problematic. This problem was approached in the 1960s first by using superconducting magnets and second by getting a homogeneous and stable magnetic field.[3] In 1966, physical chemists Ernst with his supervisor Anderson at Varian Associates in Californian were able to improve the sensitivity by introducing Fourier transformed (FT) NMR which increases the signal

to noise ratio.[4] The development of FT-NMR enabled scientists to collect proton NMR spectra of small proteins with a few hundred resonances. However, these NMR spectra were severely crowded and the assignment of their resonances was a challenge.[5] The idea of multidimensional NMR spectra was raised by physicist, Jean Jeener through an exercise given to students in a summer school course in 1971. Nonetheless, Ernst was captivated by the idea of applying a two-pulse experiment to produce 2D-NMR[5]. In 1975, Ernst's group published the first article about two-dimensional NMR.[6] In 1991, the Nobel Foundation awarded Richard R. Ernst the Nobel Prize in chemistry for his achievements in the field of NMR spectroscopy. Later, NMR spectroscopy was developed to perform two, three, four, and five dimensions. The multidimensional NMR experiments can be homonuclear where all the spectrum axes correspond to the same type of nuclei or heteronuclear where the spectrum axes represent different nuclei.

NMR spectroscopy has been actively employed to determine the atomic-resolution structure of biomacromolecules. It is one of the most powerful techniques in the field of structural biology. NMR spectroscopy is considered a complementary approach to X-ray crystallography. NMR techniques need noncrystalline protein samples while crystalline samples are required for X-ray crystallography. Thus, NMR spectroscopy can be used to confirm the three-dimensional structures of proteins determined via X-ray crystallography. Crystalline samples are difficult to prepare for some proteins; therefore, NMR spectroscopy is used to study these kinds of proteins. Protein interactions with another molecule can be studied by NMR under different solution conditions such as temperature, pH, salt concentration, and buffer. Moreover, NMR spectroscopy is utilized to characterize the internal dynamics of protein structures.[7]

1.1.2 Basics of NMR

The fundamental requirements of NMR spectroscopy are similar to other kinds of spectroscopy such as UV-Visible spectroscopy. In spectroscopy, electromagnetic radiation is absorbed by a molecule producing a transition from the ground state to the excited state. In the case of UV-Visible spectroscopy, an electron absorbs visible or ultraviolet light to move to the excited state. However, NMR spectroscopy needs electromagnetic radiation in the range of radiofrequency to promote a nuclear spin to the excited state. The main difference between NMR spectroscopy and other spectroscopies is that the former requires an external magnetic field to generate the nuclear ground and excited states. In the case of NMR, the two energy levels, called ground and excited states, are generated as a result of an interaction between a nuclear magnetic dipole moment and the external magnetic field. The nuclear magnetic dipole moment is formed in some nuclei containing spin angular momentum. This spin exists in nuclei with an odd mass number (e.g. ^{1}H, ^{13}C, ^{15}N) due to an unpaired proton. The quantum mechanical description of NMR refers to the maximum observable component of the angular momentum of a nucleus possessing a spin is $Ih/2\pi$, where I is the nuclear spin quantum number and it can be a half-integer or integer and h is Planck's constant. Every nucleus with $I \neq 0$ has a magnetic moment (μ), which is proportional to its spin.

$$\mu = \gamma I h / 2\pi \qquad (1.1)$$

In this expression γ is the gyromagnetic ratio, a characteristic constant for a given nucleus. The properties of the most used nuclei for biomacromolecular studies are listed in Table 1.1.[8]

In the presence of an external magnetic field, the magnetic moment is quantized along $2I + 1$ possible states or orientations. The energy difference between two states, also known as the transition energy, at an applied magnetic field B_0 can be calculated from Eq. 1.2.

$$\Delta E = h\nu = \mu B_0 / I \qquad (1.2)$$

Where ν is the radiation frequency of a nucleus, called the Larmor frequency, which depends only on the applied magnetic field and the nature of the nucleus.

Table 1.1: Properties of some selected isotopes, with non-zero spin, of biological NMR importance. Adapted from reference[8]

Nucleus	Spin	Gyromagnetic ratio, $\gamma(Ts^{-1})$	Natural abundance (%)
^{1}H	1/2	$2.675 * 10^{8}$	99.99
^{2}H	1	$4.107 *10^{7}$	0.012
^{13}C	1/2	$6.728 *10^{7}$	1.07
^{14}N	1	$1.934 *10^{7}$	99.63
^{15}N	1/2	$-2.713 *10^{7}$	0.37
^{17}O	1/2	$-3.628 *10^{7}$	0.038
^{19}F	1/2	$2.518 *10^{8}$	100
^{31}P	1/2	$1.084*10^{8}$	100

Nuclei with spin ½, such as ^{1}H, ^{13}C, possess two possible states in the presence of the external magnetic field, shown in Figure 1.1. The magnetic moments of the nuclei in a magnetic field will align with the field (ground state) or against the field (excited state). The populations of both states are described by the Boltzmann distribution Eq. 1.3.

$$N^{+}/N^{-} = e^{-(E/KT)} \quad (1.3)$$

Where N^+ is the number of nuclei in the upper level and N^- is the number of nuclei in the lower level, K is the Boltzmann constant, T is the temperature.

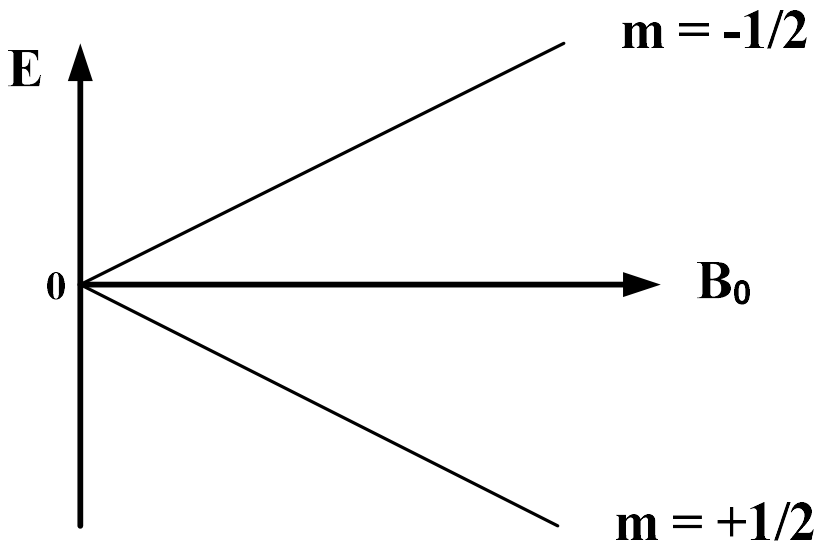

Figure 1.1: The nuclear spin energy levels of a nucleus with spin ½ in an external magnetic field. Adapted from reference 7.

The energy gap ΔE between these states is comparatively small. Although the energy from thermal collisions is enough to promote a number of nuclei into higher energy spin states, the population of nuclei in the ground state will be much higher. The nuclei in the low energy state can be excited to the high energy state by applying radiofrequency radiation. The energy difference between the low and high spin states determines the convenient radiation frequency for excitation to take place.

The Larmor frequency is also known as the precession frequency due to the precession of the spinning nuclear axis around the external magnetic field. The Larmor frequency can be determined for a nucleus at a precise applied magnetic field by Eq. 1.3.

$$\nu = \gamma B_0/2\pi \quad (1.3)$$

The higher the applied magnetic field, the higher the energy gap and the resonance frequency, which is known as the MHz value for different spectrometer types[9].

Chemical shift

The resonance frequency is proportional to the applied magnetic field B_0. This magnetic field is modified by the local chemical environments before it reaches the nuclei. Therefore, the resonance frequency of a particular nucleus in distinct surroundings is different. The position of the resonance frequency of a specific nucleus is known as chemical shift. Chemical shifts of different NMR active nuclei have been listed and used by organic chemists to characterize small molecules. However, that is not the same case with proteins, which are very large molecules and give hundreds to thousands of chemical shifts. Today, chemical shifts are very important in biophysical studies. They are used to calculate backbone and side-chain torsion angles, predict secondary structures, measure protein flexibility, and determine three-dimensional structures of proteins.[10]

The chemical shift is the resonance frequency, the observed frequency, of a nucleus reported relative to a standard in a magnetic field, and it is often denoted as δ which is defined in terms of a standard reference. Tetramethylsilane (TMS) is a common reference utilized for the ^{1}H and ^{13}C nuclei, while liquid NH_3 is used for the ^{15}N nucleus. The chemical shift is conventionally provided as a fraction of the applied magnetic field, in parts per million (ppm).[11]

Nucleus–nucleus interactions

Two nuclei may interact in two ways with each other in NMR spectroscopy. The first is an indirect interaction between two spins through bonds, mediated by electrons. This interaction is known as

scalar coupling or J-coupling. The second way is a direct interaction between two spins through space, and it is called dipolar coupling.

Scalar couplings are important to determine the chemical connectivity between atoms since they occur as a result of spin-spin interactions through bonding electrons. Scalar couplings are designated as nJAX, where A and X are nuclei interacting with each other through an n number of bonds. This coupling can take place through 1, 2, 3, 4, or 5 bonds. The magnitude of the coupling constant is independent of the external magnetic field. However, the coupling constant, measured in Hz, depends on the number of bonds connecting the coupled spins, the gyromagnetic ratio of the coupled nuclei, and the conformation of bonds involved in multiple bond couplings. Therefore, the scalar coupling is used for backbone and side chain assignments of proteins.[12] Typical values for the scalar coupling in proteins are listed in Table 1.2

Table 1.2: Typical values for the scalar coupling in proteins. Adapted from reference[12]

J-coupled spin pair	Coupling value	J-coupled spin pair	Coupling value
$^{2}J_{HH}$	9-15 Hz	$^{1}J_{NCO}$	15 Hz
$^{3}J_{HH}$	0-14 Hz	$^{1}J_{C\alpha CO}$	55 Hz
$^{1}J_{NH}$	90 Hz	$^{1}J_{C\alpha C\beta}$	35 Hz
$^{1}J_{NC\alpha}$	7-11 Hz	$^{1}J_{CH}$ (aliphatic)	130-150 Hz
$^{2}J_{NC\alpha}$	4-9 Hz	$^{1}J_{CH}$ (aromatic)	160 Hz

Dipolar coupling is another type of nuclear interaction, which occurs between two spins through space. In contrast to scalar coupling, the dipolar coupling can possess a magnitude as large as

thousands of Hertz. Cross relaxation and the nuclear Overhauser effect result from dipole-dipole interaction between nuclear spins in close proximity. Close nuclear spins, within 5 angstroms, experience the magnetic dipole moment of each other. NMR pulse techniques are able to measure the effect of one spin on the other, which is called the nuclear Overhauser effect (NOE). The transmission rate of this effect is known as the cross relaxation rate. The rate is inversely proportional to r^6, where r is the distance between two nuclear spins. Consequently, the determination of the cross relaxation rate helps to calculate the distance between the interacting nuclei. The NOE is particularly important in the structure calculation of proteins by NMR spectroscopy.[13]

Relaxation

As mentioned earlier, the population of the nuclear spins in the lower and higher energy states follows the Boltzman distribution. At room temperature, the number of spins in the lower energy state is higher than the number in the upper state and the system is in equilibrium. Appling radio frequency (RF) radiation will increase the spin population in the higher state. When the source of radiation energy is stopped, the system undergoes relaxation processes where the system spends a finite time to return to the original equilibrium condition. Two nuclear relaxations are involved in this process: spin-lattice relaxation and spin-spin relaxation.

Spin-lattice relaxation, also referred to as longitudinal relaxation or T_1, occurs as a result of energy exchange between the spins and the lattice. By this relaxation, the net magnetization returns to the z-axis and aligns with the applied magnetic field. The recovery of longitudinal magnetization is described by an exponential curve (Eq 1.4).

$$\mathbf{M_z = M_0\,[1-\exp^{(-t/T_1)}]} \qquad \textbf{(Eq. 1.4)}$$

Spin-spin relaxation, also known as transverse relaxation or T_2, takes place due to the dephasing of the transverse components of magnetization (M_{xy}) in the xy plane. In this process, the magnetic fields interact with each other, without energy transfer to the lattice, and their interactions cause a loss in phase leading to transverse magnetization decay. The decay of transverse magnetization follows an exponential curve (Eq 1.5).

$$\mathbf{M_{xy} = M_0 \exp^{(-t/T_2)}} \qquad \textbf{(Eq. 1.5)}$$

Both longitudinal and transverse relaxations can be determined for proteins by utilizing particular NMR experiments.

1.1.3 Multidimensional NMR

Multidimensional NMR spectroscopy is very useful for structural studies on biomacromolecules such as proteins in which 1D-NMR is insufficient due to overlapping. Multidimensional NMR experiments produce a spectrum where the location of the peaks or spectral lines is defined by more than one frequency. For instance, a 2D-NMR spectrum has two frequency axes representing chemical shifts of peaks. The appearance of these peaks is a result of nuclear spins interacting with each other via scalar or dipolar coupling. Multidimensional NMR experiments have a different principle compared to 1D experiments. In a 1D experiment, the pulse sequence consists of two basic elements: preparation and detection periods (Figure 1.2A). During preparation, the spin system is allowed to return to, or near, thermodynamic equilibrium and it usually ends with a single 90° pulse. During the detection period, the resulting signal is recorded. Nonetheless, the pulse sequence of the 2D-NMR experiment includes four elements: preparation, evolution, mixing, and detection (Figure 1.2B). During the evolution period, the magnetization is labeled with the chemical shift of the first nucleus and referred to as the indirectly detected dimension, because of the indirect detection of the excited state of the spin by the receiver coil. During the mixing period, the magnetization is transferred from the first spin to the second. The mixing can be evoked by either

scalar or dipolar coupling. In addition to the four basic elements of the 2D-NMR pulse sequence, 3D-NMR experiments have two more periods, evolution and mixing, in the pulse sequence[14-15](Figure 1.2C).

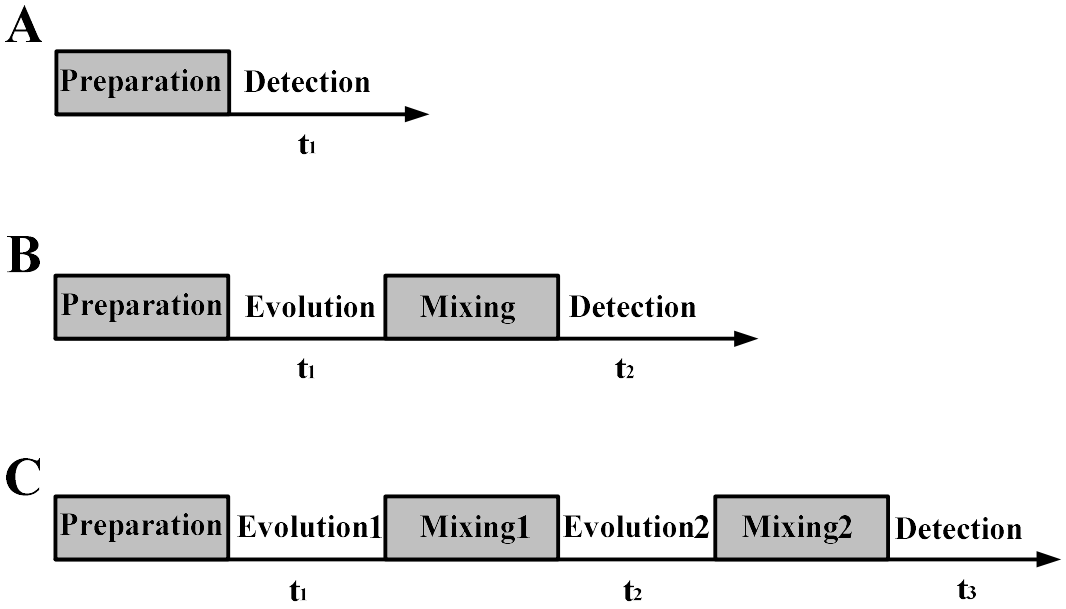

Figure 1.2: Schematic representation of the periods in typical NMR experiments: (**A**)1D-NMR (**B**) 2D-NMR (**C**) 3D-NMR.

Multidimensional NMR experiments can be classified into two main categories: homonuclear and heteronuclear experiment.[16] Homonuclear experiments provide the chemical shift of the same nucleus (which is a proton) on spectrum axes. These experiments are impractical for proteins, especially for large ones, because it increases the number of possible correlations without increasing the chemical shifts scale. Therefore, the analysis of homonuclear spectra of proteins are too complicated. On the other hand, heteronuclear experiments include two or three different nuclei on spectrum axes. For instance, the 3D NMR experiments commonly used for protein backbone assignment produce chemical shifts which are related to protons, carbons, and nitrogens. The major advantage of these experiments is the different nucleus (e.g. nitrogen) in other dimensions lowers the number of correlations while increasing the chemical shift scale resulting in high resolution spectra.[16] The chemical shifts for ^{13}C, ^{15}N nuclei in protein molecules are spread out over a much

wider range than for protons. For example, The window for ^{13}C is up to 200 ppm but for ^{1}H it is up to 12 ppm. Thus, the signal from each carbon in a protein can often be observed as a distinct peak, without the overlapping that is encountered in ^{1}H spectra.

1.1.4 Protein NMR

In the field of structural biology, NMR is one of the most important methods employed for the determination of protein structures at atomic resolution.[17] In general, the determination of an NMR solution structure of a protein can be divided into five major parts: (1) sample preparation, (2) NMR data acquisition and processing, (3) backbone and side-chain assignments, (4) collection of structural restraints, and (5) tertiary structure calculation and refinement.

The first step for a protein structural study using NMR spectroscopy is to prepare a sample with good quality and proper isotopic labeling. A quality protein sample reduces the time for collecting and analyzing the NMR data by giving good NMR spectra. Although ^{15}N and ^{13}C isotopes are present in proteins, the natural abundance of these isotopes is very low (table 1.1). ^{15}N and ^{13}C-labeled samples are typically sufficient to produce high quality NMR spectra for structural studies.[18] To enrich NMR samples with ^{13}C and ^{15}N, the expression of the proteins is usually done in minimal growth media using bacterial expression systems. The minimal media include ^{13}C-glucose and ^{15}N-ammonium chloride as the sole carbon and nitrogen sources, respectively.[19]

One of the most important 2D-NMR experiments used for protein studies is HSQC (Heteronuclear Single Quantum Coherence).[20] In an HSQC experiment, magnetization is transferred via the scalar coupling from the ^{1}H nucleus to the directly attached ^{13}C or ^{15}N nucleus. The chemical shift is evolved on the heteronuclei and the magnetization is then transferred back to the hydrogen for detection.[21] The resultant spectrum possesses one axis for a 1H chemical shift and the other for a heterogeneous (^{13}C or ^{15}N) (Figure 1.3). The ^{15}N-HSQC spectrum is considered a fingerprint of a protein at specific conditions such as temperature, pH, solvents, salt, etc.[22] Thus, any change in

protein structure due to mutation(s), ligand binding, or changes in the conditions mentioned previously reflects on the HSQC spectrum. ^{15}N-HSQC is usually the first heteronuclear experiment performed on proteins. All H-N correlations, especially the backbone amide groups, are visible in this experiment.

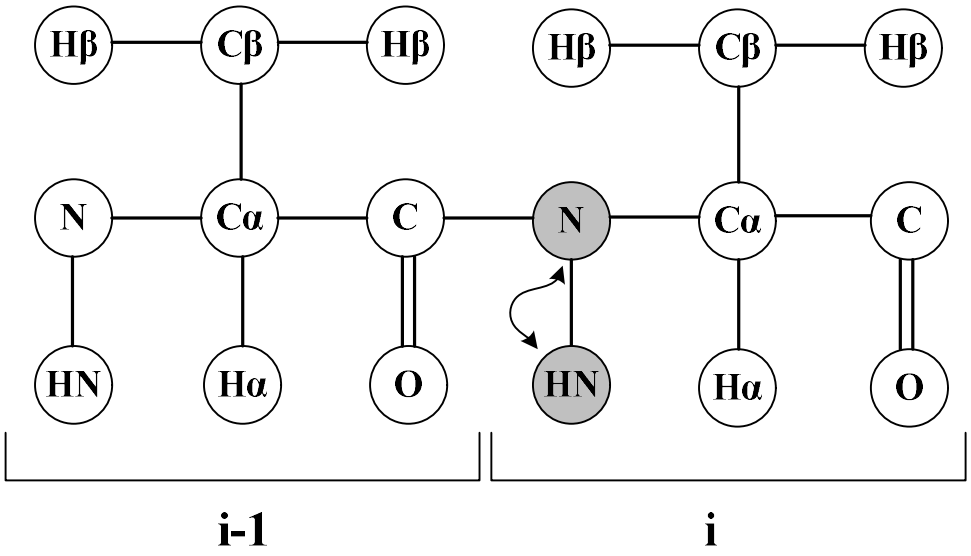

Figure 1.3: Schematic representation of the magnetization transfer pathway between hydrogen and the attached ^{15}N nuclei of the i residue in the HSQC experiment.

In applications to proteins, where it is common to have over a hundred resonance signals in a ^{15}N-HSQC spectrum, it is common to have more than one resonance signal or peak at the same chemical shift. Three-demintional NMR experiments significantly reduce peak overlapping by spreading out the information contained in an HSQC spectrum into a cube to generate a three-dimensional data set (Figure 1.4).[23] Before the structure calculation step, the protein sequence-specific resonance assignments must be conducted. The assignments generally can be divided into two sections: the sequential assignment of the amino acids in the protein sequence (also known as the backbone assignment) and the assignment of the amino acid side chains.[24]

One of the well-known approaches to achieve sequential assignment is the use of triple resonance NMR experiments; e.g. 3D HNCACB, 3D CBCA(CO)NH, 3D HNCA, 3D HN(CO)CA, 3D HNCO, and 3D HN(CA)CO (Figure 1.5). In these experiments, three different nuclei, such as ^{1}H, ^{13}C, and ^{15}N are correlated. The sequential correlations via ^{1}J and ^{2}J couplings (through bond interactions) between ^{15}N and ^{13}C nuclei and between these nuclei and their attached protons are employed to establish connectivities between amino acids. Their spectra are often analyzed in pairs with one spectrum including both intra- and inter-residue peaks and the second including only inter-residue peaks.

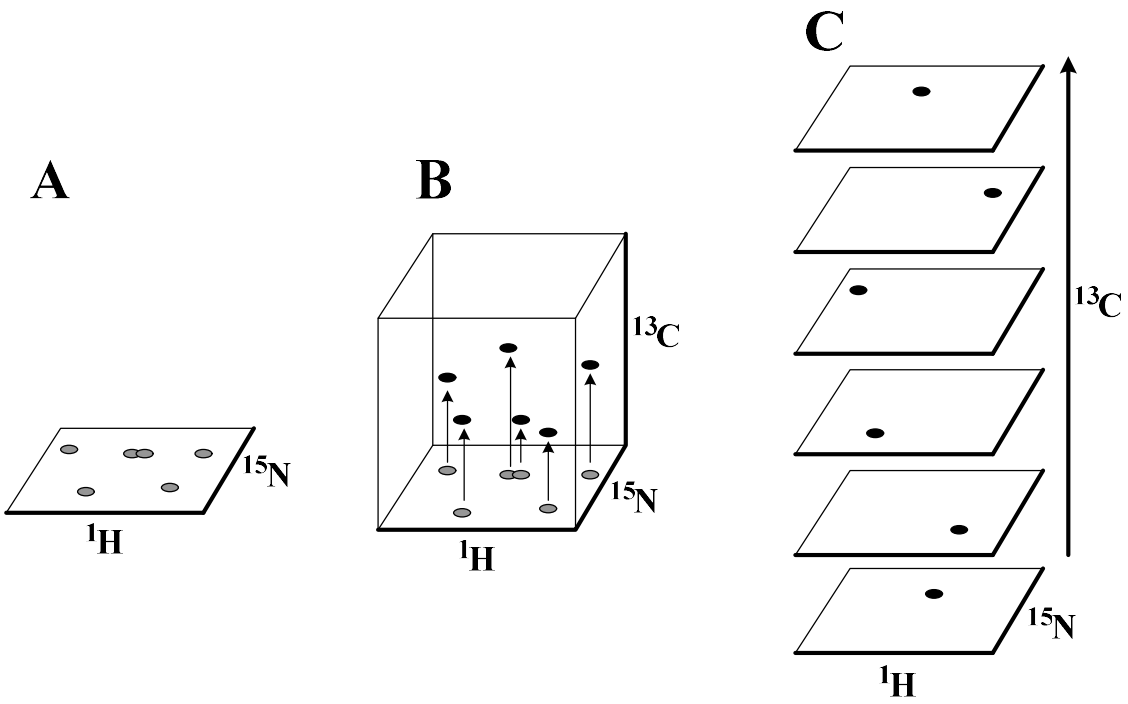

Figure 1.4: Schematic representation of 3D-NMR development from 2D-NMR, (**A**) [^{1}H,^{15}N]-HSQC (**B**) [^{1}H,^{15}N,^{13}C]-NMR, (**C**) multiple [^{1}H,^{15}N]-HSQC views at precise ^{13}C chemical shifts obtained from [^{1}H,^{15}N,^{13}C]-NMR spectrum.

All of the triple resonance experiments mentioned above for protein backbone assignments provide an [^{1}H-^{15}N]-HSQC plane expanded to a ^{13}C third dimension. Among these, the 3D HNCACB correlates the ^{1}HN and ^{15}N chemical shifts with the ^{13}Cα and ^{13}Cβ chemical shifts of its own residue (residue i) and the preceding residue in the sequence (residue i-1). This experiment gives four cross peaks in the ^{13}C dimension for each residue (except glycine which does not have Cβ): two from the intra-residual and two from the inter-residual ^{13}Cα and ^{13}Cβ atom. Although the 3D HNCACB experiment gives the highest number of ^{13}C signals, ^{13}Cα and ^{13}Cβ peaks are easily distinguished by color since they have opposite phase signals. On the other hand, the 3D CBCA(CO)NH correlates the ^{1}HN and ^{15}N chemical shifts only with the ^{13}Cα and ^{13}Cβ chemical shifts of the previous residue (residue i-1). Interestingly, the magnetization in the latter experiment is transferred by the CO atom. Therefore, only two cross peaks are observed in the spectrum, which belong to the inter-residual ^{13}Cα and ^{13}Cβ atoms. Sequential assignments can then be accomplished by matching the chemical shifts of the individual spin systems with each other. The 3D HNCACB and 3D 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 shift. 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. The 3D HNCO and 3D HN(CA)CO work in a similar way, just with the carbonyl carbons rather than Cα and Cβ atoms, and the 3D HNCA and the 3D HN(CO)CA contain the chemical shifts of only ^{13}Cα atoms.

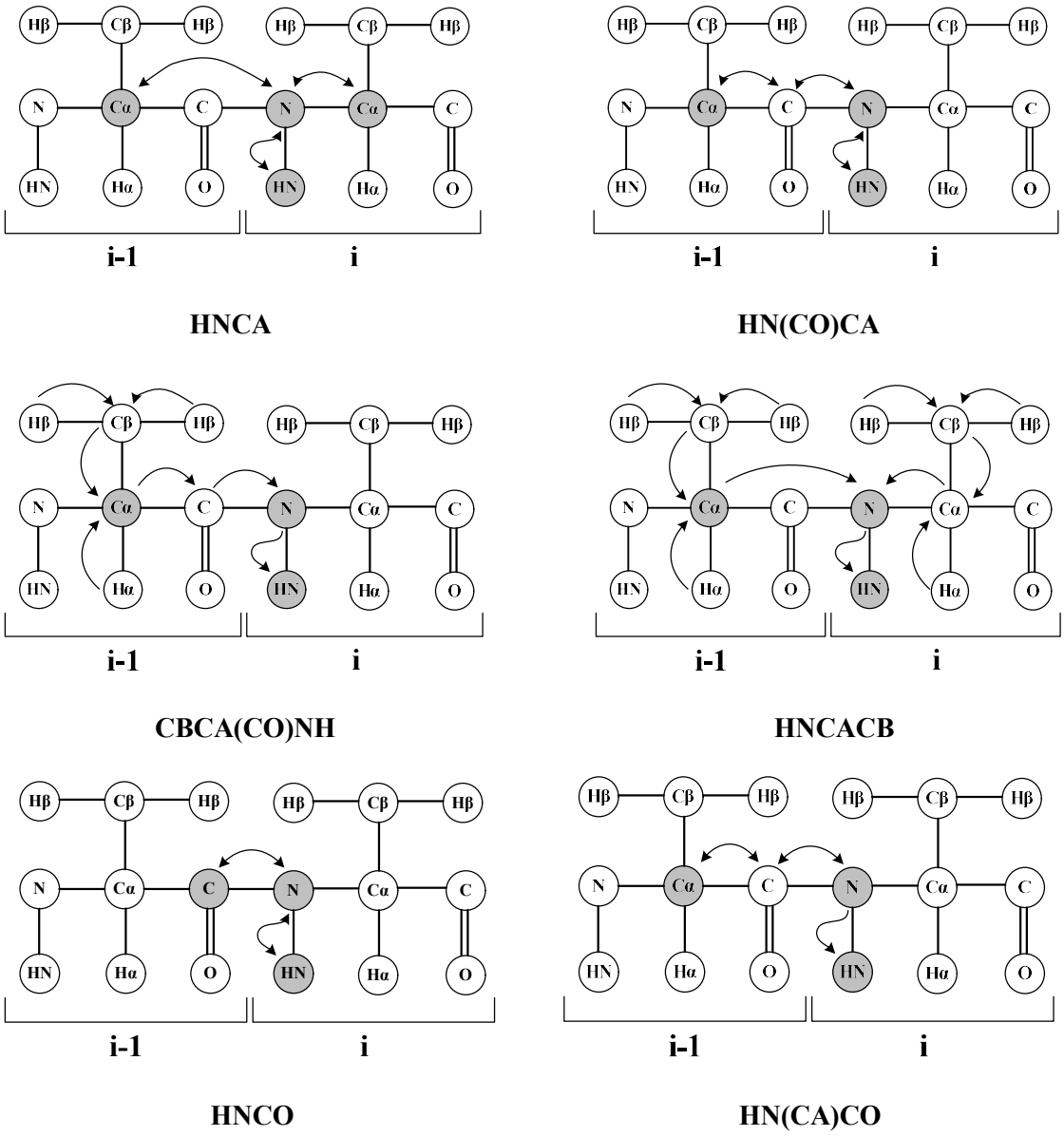

Figure 1.5: Schematic representation of the magnetization transfer pathways of most commonly used triple-resonance NMR experiments for protein backbone assignment: HNCA, HN(CO)CA, CBCA(CO)NH, HNCACB, HNCO, and HN(CA)CO

The assignment of the amino acid side chains can be done via one or two of the following experiments: 3D ^{15}N-TOCSY-HSQC, 3D HCCH-TOCSY, or 3D H(CCO)NH (Figure 1.6). A ^{15}N-labeled protein is sufficient to achieve a 3D ^{15}N-TOCSY-HSQC experiment. Here, the magnetization is transferred between all of the aliphatic proton nuclei. The magnetization is then transferred to the amide ^{15}N nucleus of its own residue (residue i) and back to proton nuclei for

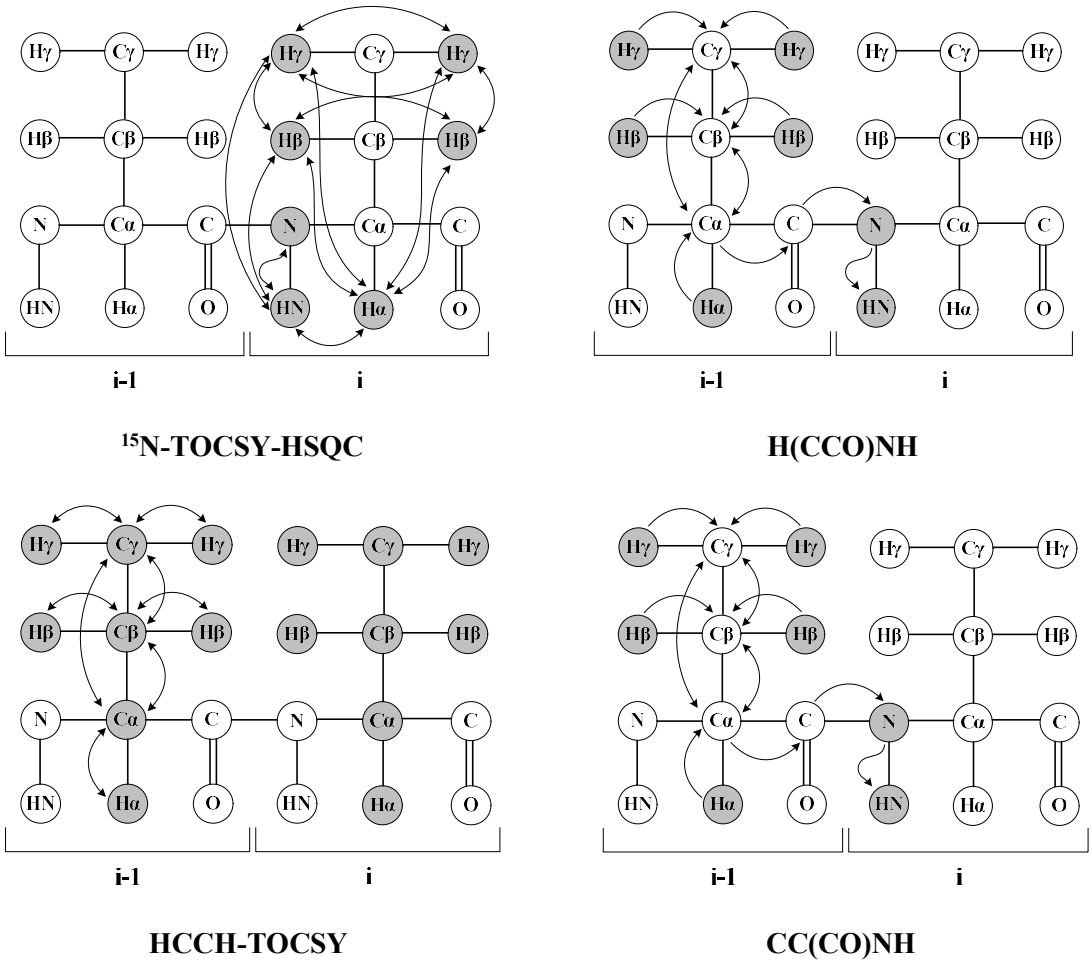

Figure 1.6: Schematic representation of the magnetization transfer pathways of some commonly used 3D-NMR experiments for protein side-chain experiments: ^{15}N-TOCSY-HSQC, H(CCO)NH, HCCH-TOCSY, and CC(CO)NH.

detection. Indeed, this experiment helps to identify amino acid residue types. 3D HCCH-TOCSY and 3D H(CCO)NH experiments require a ^{15}N, ^{13}C-labeled protein sample. 3D HCCH-TOCSY provides chemical shifts of side chain protons of residues i and i-1. 3D H(CCO)NH, alternatively,

provides a spectrum where chemical shifts of the side chain protons of residue i-1 are visible.[25] In addition to the proton side chain assignments, the chemical shifts of side chain carbons are necessary to be assigned. The 3D CC(CO)NH experiment is used to obtain carbon side-chain assignments. In this experiment, a magnetization transfer occurs from the side chain protons of residue i-1 to their attached ^{13}C nuclei. Magnetization is then transferred to the amide nitrogen through the carbonyl carbon. Finally, the magnetization is passed to the amide hydrogen for detection (Figure 1.6). The 3D CC(CO)NH experiment provides peaks for all the aliphatic side chain carbons of the residue i-1.

In protein NMR structure determination, interproton distance- and torsion angle-constraints are the source of geometric information. The interproton distance-constraints are obtained from NOE measurements, while the torsion angle-constraints are generated from backbone chemical shifts.[26] The basic concept of NOE has been discussed previously. 3D NOE experiments are needed to have an adequate number of distance constraints for obtaining precise and accurate structural results. These experiments label proximal proton pairs with the chemical shift of ^{15}N as in 3D ^{15}N-NOESY-HSQC or ^{13}C as in 3D ^{13}C-NOESY-HSQC. The 3D ^{15}N-NOESY-HSQC experiment can be done using a ^{15}N-labeled sample, but a ^{13}C-NOESY-HSQC needs a ^{15}N and ^{13}C-labeled sample to be accomplished. A comparison of the chemical shifts of peaks in the NOESY spectra with those of the triple resonance experiments (backbone and side chains) is applied to achieve NOE assignments. The structure calculation is carried out by utilizing a computer program, such as CYANA,[27] XPLOR-NIH,[28] ARIA,[29], etc. These programs used input files containing all the empirically generated distance and angular constraints. Their calculations produce an ensemble of conformers, which will converge if the data is sufficient to indicate a certain fold.

1.2 Pheromone binding Proteins (PBPs)

1.2.1 Introduction

The physiological reception of chemical stimuli by organisms is known as chemoreception. This process can be classified into two classes: olfaction and gustation. Olfaction, also known as smell, is the sensory detection of chemicals in the gaseous state; gustation, also called taste, is the sense of substances in solutions.

Chemoreception is an essential process for the regulation of insect behaviors, such as mating, feeding, oviposition, and predator avoidance. It detects particular olfactory stimuli (semiochemicals) and translates these into nerve impulses. The semiochemical is a term derived from the Greek word semeon, which means "a signal".[30] Diverse proteins are located inside the antenna of insects, in order to receive signals, including odorant binding proteins (OBPs), chemosensory proteins (CSPs), odorant receptors (ORs), and sensory neuron membrane proteins. OBPs are divided upon their target ligands to pheromone binding proteins (PBPs) and general odorant binding proteins (GOBPs).[31] The OBPs that bind to pheromones are known as pheromone binding proteins (PBPs) which are more abundant in male antennae.[32-33] The PBPs have high sequence variability when compared among species.[34] The GOBPs binds to general odorants like plant volatiles and present at similar level in antenna of both female and male insect. In contrast to the PBPs, GOBPs are highly conserved when compared among species.[34-35]

The PBPs are classified based on the length of their C-terminal segment into three structural categories: long, medium, and short.[36-37] *Antheraea polyphemus* (ApolPBP), *Amyelois transitella* (AtraPBP1) and *Bombyx mori* (BmorPBP) are lepidopteran PBPs which belong to the long-length C-terminus subclass. The *Apis mellifera* (AmelPBP) from honey bees is considered as a PBP with a medium-length C-terminus.[38] The *Leucophaea maderae* (LmadPBP) from cockroach is one of the PBPs that classified as short-length C-terminus subclass.[39]

Pheromone binding proteins (PBPs) are tremendously selective to differing pheromones.[40] These proteins are present in the space between the dendritic membrane of sensory neurons and the cuticular wall of sensory hairs. Pheromones, a class of semiochemicals, work as chemical stimuli in intraspecies communication. The term pheromone is a combination of two Greek words; pherein means “to carry” and hormone means “to excite”.[41] This term was introduced in 1959 by German biochemist Peter Karlson and Swiss biologist Martin Lüscher.[42] Pheromones are small volatile hydrophobic molecules that carry information from one member to another member of the same species. In lepidopteran moths, sex pheromones are produced and emitted by female moths to trigger mating behavior.[43] The airborne signal released by the females may consist of a single kind of pheromone or a specific blend of pheromones. Males receive the chemical message through a relatively large antenna where pheromone receptors are enclosed by an aqueous solution called the sensillum lymph.[44] To reach the receptors, pheromones are encapsulated by binding to a hydrophobic pocket of the PBPs. [45]

PBPs are acidic and water-soluble proteins with a small molecular mass (15–20 kDa). All insect PBPs contain six conserved cysteine residues in their amino acid sequences.[46] Trichogen and tormogen are olfactory accessory cells that produce PBPs.[47] The mature PBPs are secreted copiously into the sensillum lymph of the trichoid sensilla after cleaving off a signal peptide of about 20 amino acids.[31] By binding to PBPs close to the cuticular pores, pheromones are solubilized, concentrated in the sensillum lymph, protected from enzymatic degradation[48], and transferred to olfactory receptors.

Three-dimensional structures of proteins can be determined, as mentioned previously, by X-ray crystallography or NMR spectroscopy. Three-dimensional structures of different PBPs have been determined by at least one of the two techniques. The tertiary structure of the PBP in the silkworm moth *Bombyx mori* (BmorPBP) in the bound form with the species-specific pheromone was calculated by X-ray crystallography.[49] This crystal structure was the first calculated structure of a six-cysteine insect PBP. After that, the NMR structure of free (Figure 1.7A) and ligand-bound BmorPBP were determined.[50-51] The PBP structure of the giant silk moth *Antheraea polyphemus* (ApolPBP) was solved by NMR spectroscopy[46] (Figure 1.7B). Furthermore, NMR spectroscopy was employed to determine the PBP of the navel orange worm, Amyelois transitella (AtraPBP1).[52] The PBP structure of honeybee *Apis mellifera* L (AmelASP1) was solved by X-ray crystallography.[38] Recently, the structure of the Gypsy Moth *Lymantria dispar* PBP1 (LdisPBP1) was elucidated by NMR.[53] Some of the lepidopteran PBPs which their three-dimensional structures have been determined are listed in Table 1.1.

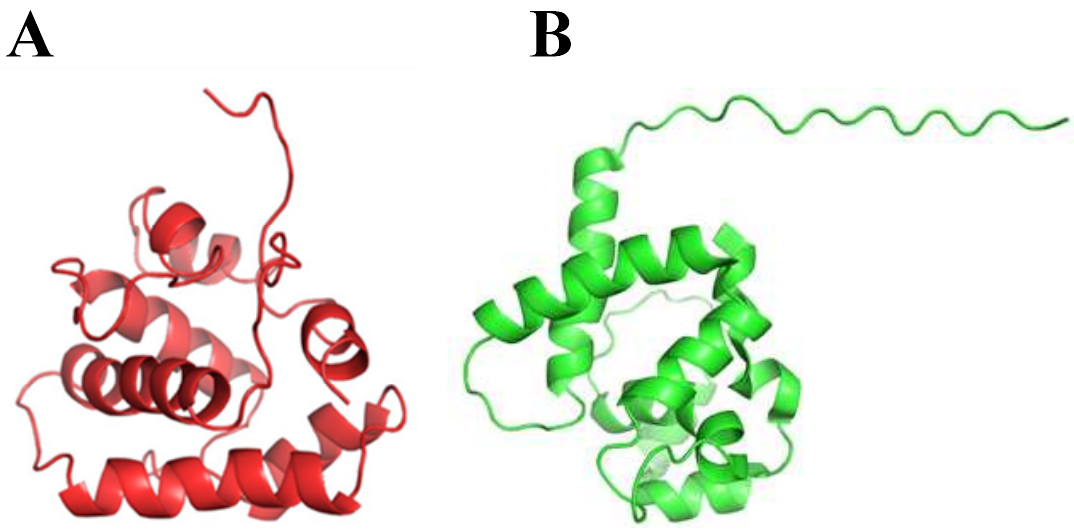

Figure 1.7: NMR structure of two free pheromone binding proteins: (**A**) BmorPBP, (**B**) ApolPBP

Table 1.3: List of the high-resolution structures of lepidopteran PBPs determined by different techniques thus far with their PDB code.

Protein	Method	PDB ID	Reference
ApolPBP1 (pH 6.3)	NMR	1QWV	[46]
ApolPBP1 (pH 5.2)	NMR	1TWO	[54]
ApolPBP1 (pH 4.5)	NMR	2JPO	[55]
AtraPBP1 (pH 4.5)	NMR	2KPH	[52]
AtraPBP1 (pH 6.5)	X-ray	4INW	[56]
BmorPBP (pH 7.5)	X-ray	2FJY	[57]
BmorPBP (pH 6.5)	NMR	1LS8	[51]
BmorPBP (pH 4.5)	NMR	1GM0	[50]
BmorPBP (pH 8.2)	X-ray	1DQE	[58]
LdisPBP1 (pH 4.5)	NMR	6UM9	[53]
EposPBP3 (pH 6.5)	X-ray	6VQ5	[59]

1.2.2 Well-studied lepidopteran pheromone binding proteins

The pheromone binding protein was discovered for the first time by isolating ApolPBP1 protein from the antenna of *Antheraea polyphemus*, the giant silk moth.[60] Since that time, extensive investigation has been performed of several lepidopteran pheromone binding proteins, such as *Antheraea polyphemus* (ApolPBP1), *Amyelois transitella* (AtraPBP1), *Bombyx mori* (BmorPBP), and *Lymantria dispar* (LdisPBP2). All of the mentioned PBPs are α-helical and contain six conserved cysteine residues creating disulfide bonds, which stabilize the helices that form a hydrophobic pocket. It has been reported that these moth PBPs undergo a pH-driven mechanism for ligand binding and release. They bind to their pheromones (ligands) at a pH above 6.0 and

release them at a pH lower than 4.5. These proteins have two unique conformations: open (or bound or PBP^B) and closed (or free or PBP^A). Changing between the two conformations depends on the pH of the solution and the presence/absence of a ligand. Interestingly, the structure of the C-terminus is influenced by the pH switch and ligand binding. The C-terminus of the bound PBPs at neutral pH in the open conformation is unstructured, flexible, and extended out to the solution. However, the C-terminus of the free PBPs in the closed conformation is α-helical and occupies the hydrophobic pocket. In addition to the C-terminus structural change, two histidine residues (H70 and H95) play an important role in the pH-dependent conformational switch. It has been described that the C-terminus and the two histidines are acting as biological gates at the ends of the hydrophobic cavity. The pH-driven conformational switch is controlled by two biological gates. At a pH above 6.0, the two histidine residues are neutral and act as a gate to one end of the hydrophobic cavity, whereas the C-terminus acts as a gate on the other end of the cavity. Both gates work together at pH higher than 6.0 to maintain the pheromone in the hydrophobic pocket (cavity). At acidic pH, the repulsion between the protonated histidines creates an exit for the pheromone. On the other side of the pocket and at the acidic pH, the C-terminus enters the pocket as an α-helix evicting the pheromone.[61-63] It has been reported that the membrane potential in the vicinity of olfactory receptor neurons (ORNs) has a reduced pH that triggers the conformational change

discussed above for pheromone release. Olfactory systems in the Lepidoptera include thousands of olfactory receptor neurons housed in an array of sensory hairs located on the surface of their antennae (Figure 1.8A). Olfactory receptors are ligand-gated ion channels known as ionotropic receptors located on the dendritic membrane of sensory neurons. These receptors are highly selective and extremely sensitive to low concentrations of pheromones. Once pheromones enter the sensory hairs through cuticular pores where the pH is higher than 6.0, they bind to PBPs. Pheromones are transferred by PBPs to the olfactory receptors where the pH is lower than 5.0. At acidic pH, the pheromones are released and bind to the olfactory receptors (Figure 1.8B, C).

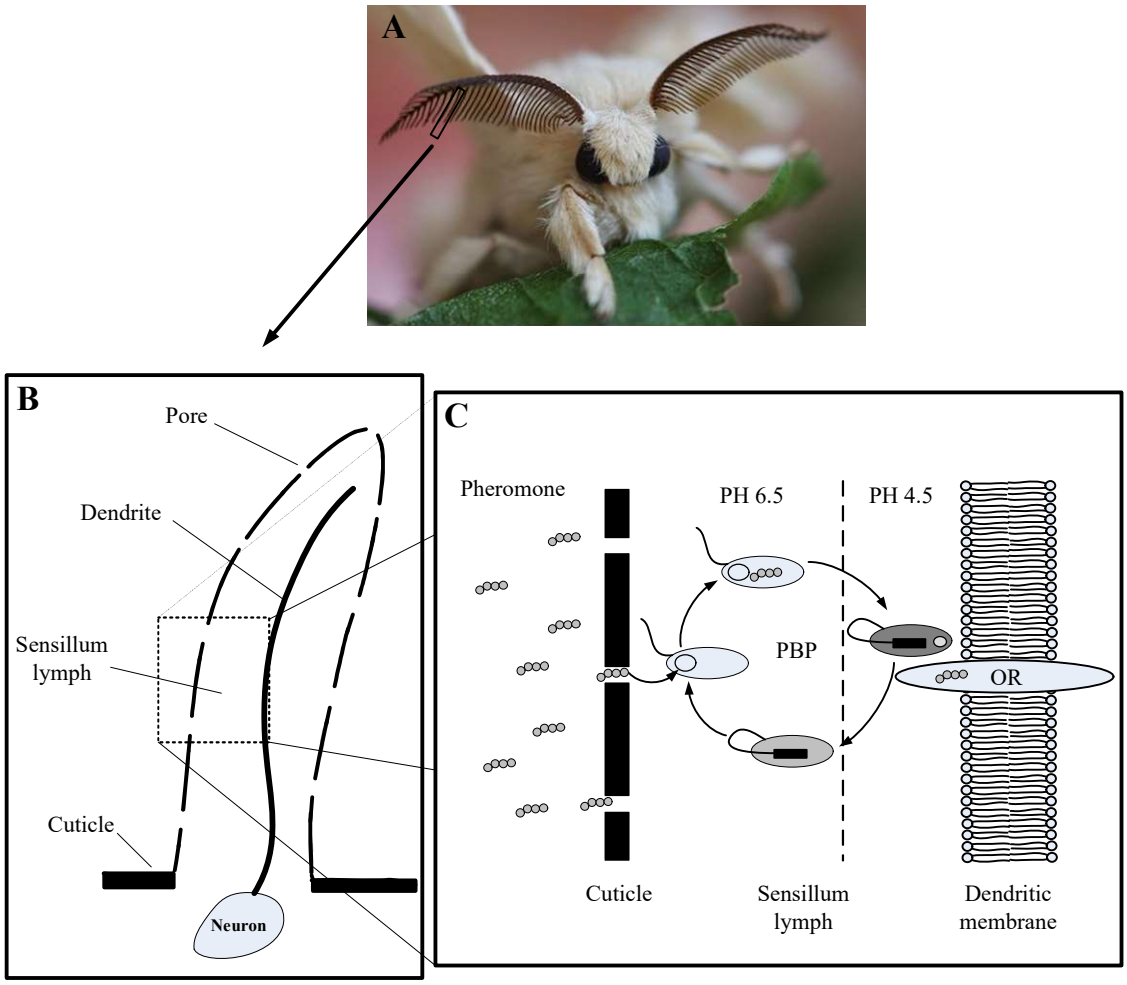

Figure 1.8: Moth antennae with olfactory sensilla: (**A**) Male moth of Bombyx mori. (**B**) Scheme of an olfactory sensillum (**C**) Schematic representation of the mechanisms of ligand binding and release of moth pheromone-binding proteins (PBPs) in the sensillum lymph.

1.2.3 The genus *Ostrinia*

The genus *Ostrinia* is a group of moths that belong to the Crambidae family (Lepidoptera: Pyraloidea). *Ostrinia* was described by the German entomologist Jacob Hübner in 1825. *Ostrinia* has 20 species classified into three groups based on uncus morphology. Ten species in one of the groups contain a trilobed uncus and they appear highly similar. However, the morphology of their male mid-tibia is different and used to divide the trilobed uncus species into three subgroups: small, medium, and massive mid-tibia. The small mid-tibia subgroup includes *Ostrinia nubilalis*, *Ostrinia furnacalis*, *Ostrinia orientalis*, and *Ostrinia dorsivittata*. *Ostrinia nubilalis (O. nubilalis)*, and

Ostrinia furnacalis (O. furnacalis) are widespread moths in the world and are described as invasive pests which damage a variety of crops.[64]

1.2.4 Ostrinia nubilalis

Ostrinia nubilalis, also known as European Corn Borer (ECB), is a serious pest in Europe and North America.[65-66] It damages a variety of agricultural crops such as corn, potato, pepper, tomato, apple, lima bean, oat, buckwheat, hop, millet, and soybean. The European continent is the original home of this species. However, it was unintentionally introduced to the USA at the beginning of the last century. *O. nubilalis* has disseminated massively throughout the nation, particularly in the northeastern part. At least one billion dollars in crops is lost annually in the USA as a result of the *O. nubilalis* invasion.[67] It is also present in Canada, Central Asia, and Northern Africa.

The life cycle of *O. nubilalis* consists of four stages beginning with eggs laid on the underside of leaves which typically take four to nine days to hatch. In the second stage, larvae are released from the eggs and move to the whorls, ears, and stalks of corn, or to the stem and the fruit of another host plant for feeding. Larvae start the pupal stage wherever they are in the host. Finally, the pupae turn into adult moths that live around 20 days. The larval stage is the one that causes damage to crops. This happens due to the larvae eating leaves and fruit and chewing tunnels through several parts of the plants, such as the stalk and main stem, resulting in breakage.[67]

Different techniques have been conducted to reduce the economic loss caused by the *O. nubilalis*, which dramatically impacts crop producers. Biological and chemical management, along with transgenic crops, have been used to control the pest population and reduce the magnitude of the crop yield loss.

Crop protection is mainly insecticide based although insecticides are harmful to human health and the environment. The number of worldwide deaths and chronic disease cases caused by pesticide (44% insecticides) poisoning is about 1 million per year.[68-69] It has been reported that insecticides

contaminate water (surface and groundwater), air, soil, and vegetation. Insecticides are not only killing pests but also other organisms including fish, beneficial insects, birds, and non-target plants. Furthermore, insecticides reduce the number of beneficial soil microorganisms and change their biochemical activity.[69-70] Dichlorodiphenyltrichloroethane, known as DDT, is one of the well-known insecticides which was synthesized by Swiss scientist Paul Müller. Consequently, he was awarded the Nobel Prize in physiology or medicine in 1948. Nevertheless, the use of DDT was discontinued in 1972 as a result of observing the carcinogenic properties of DDT and its harm to wildlife.[71-72] Control of the *O. nubilalis* infestation is not easy for two reasons: (1) great fluctuation in the population of the moths from season to season and (2) the feeding and survival behavior of the moths. To get the benefit of insecticides, they should be applied to the field within one to three days after the eggs of *O. nubilalis* are hatched when the larvae are still on the surface of plants. However, once larvae (borers) have tunneled into the plants, the insecticides are no longer effective.[73]

To avoid the risk of insecticides, biological control has been used against *O. nubilalis*. In this type of insect control, natural predators, parasites, or pathogens of a pest are used to reduce the insect's population and minimize their influence on economic and environmental practices. A complex of predators can inhibit the population growth of *O. nubilalis*. This technique does not provide a predictable or economic level of control. It has been reported that the effectiveness of predators changes widely from year to year. *Trichogramma spp.* are small wasps that are endoparasites of lepidopteran eggs. These species have been used to control the *O. nubilalis* population in the egg stage. The small wasps lay their eggs in the moth eggs stopping the development of the moth embryos before they hatch. In this way, the crops are protected from the damaging effects of the larvae. The drawback of *Trichogramma spp.* is that their eggs are a food supply for ants and get consumed quickly. In the case of an overrun infestation, the wasps are not capable of laying enough eggs to be a useful biological control.[74-75]

Transgenic crop species have been used in agriculture for the *O. nubilalis* management. These crops are genetically modified by inserting genes from unrelated species to introduce desirable traits in plants. Transgenic crops were developed commercially in the mid-1990s.[76] *Bacillus thuringiensis* (Bt) crops are plants genetically modified to be able to produce insecticidal proteins in the form of crystals.[77] One strain of Bt has been used to create an insecticidal protein for controlling *O. nubilalis*. The disadvantages of Bt crops are that pests might develop a resistance to the produced toxin and the crops are very costly. Although Bt corn minimizes the *O. nubilalis* damage compared to non-Bt, it seldom enhances profits due to high-priced seeds and inconsistent yield increase and therefore is not very effective.[67]

1.2.5 Pheromone binding protein of *Ostrinia nubilalis* (OnubPBPs)

Many insects express more than one PBP in the antenna. This phenomenon was observed for the first time in the male antenna of *L.dispar*. Two PBPs, known as LdisPBP1 and LdisPBP2, were reported to be encoded by two different genes in the male antenna of this species.[78] Similarly, five distinguishable PBPs have been reported to be present in the male antennas of *O. nubilalis* species: PBP1, PBP2, PBP3, PBP4, and PBP5. Their binding and release mechanism and binding affinities to the pheromones have yet to be studied. PBP2 and PBP3 are expressed in male *O. nubilalis* antenna at high levels compared to the female. *O. nubilais* has a sibling species, as mentioned earlier, called *O. furnacalis* which is a common agricultural pest in Asia and Oceania. The male antennas of *O. furnacalis* also have five distinct PBPs.[40] The primary structure of each PBP of both *O. nubilais* and *O. furnacalis* has six conserved cysteine residues. These residues may create disulfide bonds which are very important for stabilizing the α-helices that construct a hydrophobic pocket in the interior of the protein similar to well-studied lepidopteran PBPs. However, a protein sequence comparison of *O. nubilalis* PBPs to their counterparts in *O. furnacalis* shows a different identity percentage for each pair: PBP1 98.592%, PBP2 97.222%, PBP3 93.056%, PBP4 97.183%, and PBP5 97.887%. Based on the protein sequence, PBP3 has been suggested to be a key

component in the selection of the right pheromones in each species due to the highest difference they have among the other PBPs (Figure 1.9).

```
OnubPBP1  -SQDVMKQMTINFGKALDTCRKELDLPDSINADFYNFWKEGYELSNRQTGCAIMCLSSKL 59
OfurPBP1  -SQDVMKQMTINFGKALDTCRKELDLPDSINADFYNFWKEGYELSNRHTGCAIMCLSSKL 59
OnubPBP2  -SQAVMKDMTKNFIKAYEVCAKEYNLPEAAGAELMNFWKEGYVVTSREAGCAILCLSSKL 59
OfurPBP2  -SQAVMKDMTKNFIKAYEVCAKEYNLPEAAGAEVMNFWKEGYVLTSREAGCAILCLSSKL 59
OnubPBP3  -SQTVMREMTRNFIKAYEVCAKEYNLPEATGSELINFWKEGHELTTREAGCAILCMSTKL 59
OfurPBP3  -SQTVMGEMTKNFIKAYEVCAKELNLSEATGLQLINFWKEGHELTTRETGCAILCMSTEL 59
OnubPBP4  -SEEVMTKMGVTFFNVLEECKKELKVTTNINEGLVRFWSQGAA-PERELGCVFLCMAHKK 58
OfurPBP4  -SEELMTKMGVTFFNVLEECKKELKVTTNINEGLVRFWSQGAA-PERELGCVFLCMAHKK 58
OnubPBP5  MVPEAMKQLTGGFLKVLDQCKKELNLSDGVISDLYHLWKEEYDQISRDAGCVIHCMSQKL 60
OfurPBP5  MVPEAMKQLTGGFLKVLDQCKKELNLSDGVISDLYHLWKEEYDQISRDAGCVIHCMSQKL 60
               * .:   * :. : * ** .:       . .:*.:      *. **.: *:: :

OnubPBP1  DLVDPEGKLHHGNTHEFAKKHGADDSMAKQLVELIHKCEGSVADDPDACMKVLDIAKCFK 119
OfurPBP1  DLVDPEGKLHHGNTHEFAKKHGADDSMAKQLVELIHKCEGSVADDPDACMKVLNIAKCFK 119
OnubPBP2  NLLDPEGTLHRGNTVEFAKQHGSDDAMAHQLVDIVHACEKSVPPNEDNCLMALGISMCFK 119
OfurPBP2  NLLDPEGTLHRGNTVEFAKQHGSDDAMAHQLVDIVHACEKSVPPNEDNCLMALGISMCFK 119
OnubPBP3  NLLDVQGSVHRGNTVEFAKHHGSDDAMAHQVVDILHACEKAT-PNEDKCMLALSIAMCFK 118
OfurPBP3  NLLDVQGSVHRGNTVEFAKHHGSDDAMAHQVVDILHACEKAT-PNEDKCMLALSIAMCFK 118
OnubPBP4  DLLEDQKRLHHENAHQFARGHGADDDKATEIVSLLRECEQQFITITDDCSRALEVARCFQ 118
OfurPBP4  DLLEDQKRIHHENAHQFARGHGAEDDKATEIVSLLRECEQQFITITDDCLRALEVARCFQ 118
OnubPBP5  ELVGGDGKMHHVNIKDFALKHGAGDEIATQLVTLAHECEKQKASIEDDCERTLEMSKCFR 120
OfurPBP5  ELLGGDGRMHHVNIKDFALKHGAGDEIATQLVTLAHECEKQKAAIEDDCERTLEMSKCFR 120
          :*:  :  :*: *  :**  **: *  * ::* : : **       * *  .* :: **:

OnubPBP1  AEIHKLNWAPSMDLIVAEVLAEV--- 142
OfurPBP1  AEIHKLNWAPSMDLIVAEVLAEV--- 142
OnubPBP2  TEIHKLNWAPNHELMLEEMMAEMKQ- 144
OfurPBP2  TEIHKLNWAPDHELLLEEMMAEMKQ- 144
OnubPBP3  AEIHKLDWAPNHELMFEELVSDMWNS 144
OfurPBP3  AEIHKLDWAPNNELMFEELVLDMWNS 144
OnubPBP4  AHMQRLQWAPSMEVMVEEILAGMA-- 142
OfurPBP4  AHMQRLQWAPSMEVMVEEILAGMA-- 142
OnubPBP5  SDVKQVDWTPKMEVIITEVIEV---- 142
OfurPBP5  SDVKQVDWTPKMEVIITEVIEV---- 142
          :.::::*:*. :::. *::
```

Figure 1.9: Sequence alignment of OnubPBPs and OfurPBPs. The conserved cysteine residues are in red and highlighted in yellow. The symbols bellow the protein sequences are: (*) conserved sequence, (:) conservative substitution, (.) semi-conservative substitution, ( ) and non-conservative substitution.

Despite the fact that both *O.nubilalis* and *O. furnacalis* are related to the same subgroup called trilobed uncus, their males respond to different sex pheromones. Odorant receptors of *O. nubilalis* males are stimulated by a blend of *E* and *Z* isomers known as 11-tetradecenyl acetate (*E/Z*-11-14:OAc). On the other hand, a blend of *E* and *Z* isomers known as 12-tetradecenyl acetate (*E/Z*-12-14:OAc) (Figure 1.10) stimulate the odorant receptors of *O. furnacalis* males. *O.nubilalis* is known to be polymorphic with regard to the component ratio of the pheromone blend. Some females

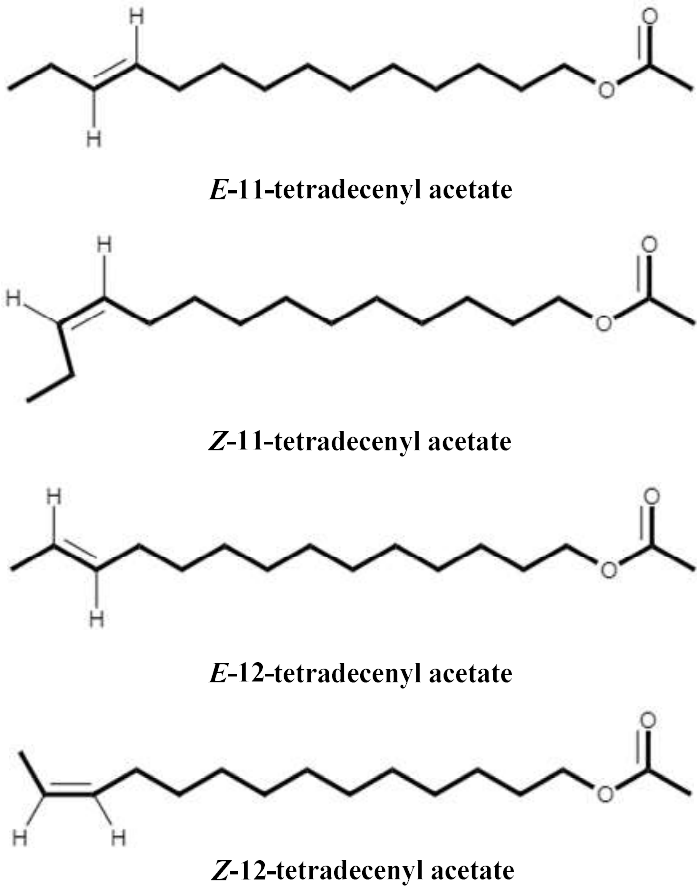

Figure 1.10: Chemical structures of *O. nubilalis* and *O. furnacalis* pheromones.

produce an *E*:*Z* blend with a ratio equal to 99:1 and some others produce the blend with a ratio of 3:97. Along with the production of two distinct pheromone blends by *O.nubilalis* females, the response of *O.nubilalis* males is different depending on the received blend. Thus, *O.nubilalis* can

be classified into two strains: *E*- and *Z*-strain. Males of the *E*-strain are attracted to a blend with 99% *E*-pheromone while males of the *Z*-strain are attracted to the other blend which contains 97% *Z*-pheromone.[79] The PBPs of *E*- and *Z*-strains have an identical protein sequence, suggesting that PBP is not a key factor for the male to discriminate the strain. Hansson and coworkers have reported that the nervous systems of both strains at the males' antenna contain the same neuronal type. However, they have found differences in abundance and sensitivity, which provide the ability for the *O. nubilalis* males to communicate with the females of their own strain.[80] Sex pheromone polymorphism is also present in *O. furnacalis*. Females of this species produce their *E*:*Z* pheromone blend with a ratio varying from 1:1 to 2:3 depending on the geographic population.[81]

The OnubPBP3 sequence was compared to the sequence of the well-studied lepidopteran PBPs, mentioned earlier, ApolPBP1, AtraPBP1, BmorPBP, and LdisPBP2 (Figure 1.11). As we described earlier, the investigated lepidopteran PBPs undergo a pH-driven conformational switch to bind and release a ligand. The sequence of OnubPBP3 is over 50% identical to the lepidopteran PBPs; however, some essential differences are noticed at the two biological gates. The C-terminus contains two additional charged residues compared to the well-characterized PBPs and the His70 is substituted with an Arg residue. Similar differences were observed in the OfurPBP2 sequence. In the protein NMR pH titration experiments, OfurPBP2 behaved differently than what has already been reported for these well-studied PBPs. OfurPBP2 at a pH above 6.0 has one predominant conformation similar to the studied lepidopteran PBPs. However, OfurPBP2 at pH 4.5 is present in different conformations instead of one (PBP^{A}) as in the other investigated PBPs.[82]

```
OnubPBP3  SQTVMREMTRNFIKAYEVCAKEYNLPEATGSELINFWKEGHELTTREAGCAILCMSTKLN 60
BmorPBP   SQEVMKNLSLNFGKALDECKKEMTLTDAINEDFYNFWKEGYEIKNRETGCAIMCLSTKLN 60
ApolPBP1  SPEIMKNLSNNFGKAMDQCKDELSLPDSVVADLYNFWKDDYVMTDRLAGCAINCLATKLD 60
AtraPBP1  SPEIMKDLSINFGKALDTCKKELDLPDSINEDFYKFWKEDYEITNRLTGCAIKCLSEKLE 60
LdisPBP2  SKDVMHQMALKFGKPIKLCQQELGADDSVVKEFLDFWKDGYVMKDRQTGCMLICMAMKLE 60
          *  :*:::: :* *  . * .*    ::   :: .***:.: :. * :** : *:: **:

OnubPBP3  LLDVQGSVHRGNTVEFAKHHGSDDAMAHQVVDILHACEKAT-P--NEDKCMLALSIAMCF 117
BmorPBP   MLDPEGNLHHGNAMEFAKKHGADETMAQQLIDIVHGCEKSTPA--NDDKCIWTLGVATCF 118
ApolPBP1  VVDPDGNLHHGNAKDFAMKHGADETMAQQLVDIIHGCEKSAPP--NDDKCMKTIDVAMCF 118
AtraPBP1  MVDADGKLHHGNAREFAMKHGADDAMAKQLVDLIHGCEKSIPP--NDDRCMEVLSIAMCF 118
LdisPBP2  LLDSAMEIHHGSTFAFAKAHGADEAMAQQIIDIVHGCTTTYPAAETNDPCQRAVNVAMCF 120
          ::*   .:*:*.:  **  **:*::**:*::*::*.* .:     .:* *  .:.:* **

OnubPBP3  KAEIHKLDWAPNHELMFEELVSDMWNS  144
BmorPBP   KAEIHKLNWAPSMDVAVGEILAEV---  142
ApolPBP1  KKEIHKLNWVPNMDLVIGEVLAEV---  142
AtraPBP1  KKEIHNLKWAPNMEVVVGEVLAEV---  142
LdisPBP2  KADVHKLNWAPDVELLVADFLAESQ--  145
          * ::*:*.*.*. :: . :.:::
```

Figure 1.11: Sequence alignment of different PBPs of moths: *Ostrinia nubilalis*, *Bombyx mori*, *Antheraea polyphemus*, *Amyelois transitella*, *Lymantria dispar*. The C-terminus (H70, R70 in case of OnubPBP3, and H95) are highlighted in green. Every charged residue in the C-terminus is written in red. All the conserved cysteine residues are written in red and highlighted with yellow. The symbols below the protein sequences are: (*) conserved sequence, (:) conservative substitution, (.) semi-conservative substitution, ( ) and non-conservative substitution.

1.3 The objective of the Study

O. nubilalis moths, as we mentioned earlier, have an enormous impact on the production of economically important agricultural crops in North America and Europe. *O. nubilalis*, similar to other lepidopterans, relies on olfaction to find their mating partners, in addition to other fundamental behaviors. The communication between female and male with the species-specific pheromones, released by females and received by males, plays an essential role in moth reproduction. Pheromone binding proteins (PBPs) present in the male moth antennae bind to the sex pheromones, transport them across the aqueous sensillum lymph, and deliver them to the olfactory receptor neurons. *O. nubilalis* PBP3 (OnubPBP3) is expressed at higher levels in the antennae of males than in the females. Although OnubPBP3 shares over 50% sequence identity with several well-characterized lepidopteran PBPs, there are critical differences at the two biological gates (histidine and C-terminus gates). The His70 is substituted with an Arg residue, while the C-terminus has two additional charged residues, compared with other well-studied PBPs. The detailed structural and mechanistic studies of OnubPBP3 will aid in biorational design of pheromone mimetics for effective pest management. The structure-based design of a mimetic can terminate the communication between the males and the females leading to suppress ion of *O. nubilalis* populations harmlessly.

OnubPBP3 has not been overexpressed and biophysically characterized. Therefore, the first objective of my work is to overexpress, and purify a recombinant OnubPBP3 (isotope-labeled and -unlabeled) by using *Escherichia coli* as a cell host. The second objective is to determine the binding affinity of the species-specific pheromones (*E*- and *Z*-11-tetradecenyl acetate) to the recombinant OnubPBP3 after delipidation by using fluorescence spectroscopy. Furthermore, we want to understand the influence of delipidation and pheromone binding on the structural conformation of OnubPBP3 by utilizing NMR spectroscopy. The third objective of this work is to determine the three-dimensional structure of OnubPBP3 by solution-state NMR spectroscopy. The

fourth objective is to gain insight into pH effects on the conformation of OnubPBP3 by different techniques including circular dichroism (CD), NMR, and small angle X-ray scattering (SAXS).

CHAPTER II

PRODUCTION, DELIPIDATION, AND LIGAND TITRATION OF ONUBPBP3

2.1 Introduction

Ostrina nubilalis is an invasive pest affecting agricultural crops in North America and Europe. Pheromone binding proteins (PBPs), present in the male antenna of this species, are crucial to establish the signaling process leading to mating. OnubPBP3, one of five distinct PBPs present in the antenna of *O. nubilalis*, is expressed in male *O. nubilalis* antenna at high levels compared to the female. To date, the structure and the mechanism of the function of OnubPBP3 have not been studied. To understand the influence of OnubPBP3 on chemosignals between female and male moths of this species, several studies have to be conducted *in vitro*.

For many years, recombinant proteins have increasingly replaced the corresponding native proteins. Recombinant protein is generated through recombinant DNA which is the protein-encoded gene cloned in a vector. The vector is used to enable a host cell to replicate, transcript, and translate the inserted the protein-encoded gene. A variety of proteins involved in different biological activities have been produced through heterologous systems using recombinant DNA technology to determine their high-resolution structures and to understand their mechanisms of actions. Different host cells can be tried to produce a recombinant protein, such as bacteria, insect, yeast, and mammalian cells. Overexpression of a protein in the *Escherichia coli* (*E. coli*) is always the preferable option over others because it is easy, cheap, and often provides a high yield of protein.

However, this protein might be obtained as inclusion bodies instead of soluble protein. Inclusion bodies (IBs) are an extremely aggregated form of the expressed protein formed as a result of protein misfolding. The misfolded recombinant protein can be reprocessed to the correct fold of the corresponding native protein through a process called refolding. This process involves several steps starting with denaturation of the inclusion body with guanidine hydrochloride or urea. In order to dissolve the aggregated proteins into a monomeric chain, a high concentration of a denaturant is used. Concentration reduction of the denaturant is a key in protein refolding and it can be achieved through various approaches including a step-wise dialysis technique to obtain active proteins.[46, 83] In general, lepidopteran PBPs, as described previously, contain disulfide bonds in their structures. To form the disulfide bonds in the recombinant OnubPBP3, the denatured protein should be refolded using a redox system, such as glutathione[84] or cysteine/cystine redox system.[85]

Recombinant proteins should be tested for the ability to bind their own ligands after purification to ensure that the proteins are active and also to determine the binding affinity to their ligands. Although *E. coli* has been used successfully to express a variety of recombinant PBPs, these proteins cannot be directly used for a ligand binding assay. It has been reported that a lipid molecule that belongs to the *E. coli* expression system occupies the hydrophobic pocket of the pure recombinant protein.[61, 86] Therefore, the recombinant OnubPBP3 must undergo a delipidation process to remove the lipid endogenous to the *E. coli* before the binding assay.

The ligand binding affinity of different PBPs has been investigated by ligand titration studies with NMR and also using fluorescence ligand binding assays.[40, 53, 61, 63, 82] In NMR ligand titration studies, a series of [^{1}H-^{15}N]-HSQC spectra of a protein are collected in the absence and presence of varying concentrations of the ligand. By overlaying all HSQC spectra collected during the ligand titration, protein-ligand binding is monitored. In this experiment, three cases may be observed: fast exchange, slow exchange, and intermediate exchange. In the fast exchange case, a single peak is observed at the average chemical shift of the free and ligand-bound protein in the NMR spectrum

representing the population-averaged value of both states. The fast exchange is observed as a result of weak interactions between protein and ligand. In contrast, slow exchange produces distinct signals for free and bound protein. By increasing the ligand concentration during the titration, the signal intensities of the bound state grow and the free state diminish. Strong protein-ligand interactions result in slow exchange.[87] In case of intermediate exchange, signal broadening is observed with the increase of ligand concentration. [88]

Fluorescence spectroscopy is a highly sensitive method used extensively for protein analysis. Extrinsic fluorescent probes have been employed to determine ligand binding affinity to PBPs.[61-63] Fluorescent probes interact noncovalently with the hydrophobic pocket of the proteins. *N*-phenyl-1-naphthylamine (1-NPN), an extrinsic fluorescent dye, was utilized as an external probe to investigate the hydrophobic pocket of many proteins.[40] 1-NPN exhibits a very low fluorescence intensity in an aqueous solution while the intensity is amplified as it binds to the hydrophobic cavity of the protein.[40, 89-91]

Herein, we carried out the expression, refolding, purification, and delipidation of the recombinant OnubPBP3. In addition, we studied OnubPBP3 interactions with its own pheromones, *E*- and *Z*-11-14:OAc, by performing NMR titration and fluorescence ligand binding assays.

2.2 Materials and methods

2.2.1 Cloning and overexpression

The OnubPBP3 gene was cloned into the NdeI and BamHI sites of the pET21a vector (Novagen/EMD Millipore) previously in Dr. Smita Mohanty's laboratory. The recombinant OnubPBP3 protein was expressed in *E. coli* Origami 2 cells using pET21a vector. The constructed OnubPBP3-pE21a plasmid was transformed into Origami 2 cells using the heat shock method.[92]

To conduct the plasmid transformation, samples of Origami 2 and OnubPBP3-pE21a plasmid were thawed in ice. The transformation was started with a mixture of chemically competent Origami 2 bacteria and 2 µl OnubPBP3-pE21a plasmid and incubated on ice for 30 min. After the incubation in ice, the mixture was placed in a water bath at 42 °C for 45 seconds to give a heat shock and then placed back on ice for 30 min. Super Optimal broth with Catabolite repression (SOC) media (200 µl) was added into the transformed cells and the mixture was then incubated at 37 °C for 30 min with shaking. After growing the transformed cells in the SOC media, 20 µl of the mixture was plated on a Lysogeny broth (LB) agar plate containing ampicillin and tetracycline. The plate was incubated at 37 °C overnight to obtain a few separated colonies of the transformed cells. Saturated overnight LB bacterial culture containing ampicillin and tetracycline started from a single colony was diluted (1:70, v/v) in LB media. 1 µl of each of 100 mg/ml Ampicillin and 10 mg/ml tetracycline were added to each ml of the culture to grow Origami 2 cells containing the OnubPBP3 gene. The fresh LB bacterial culture was shaken at 37 °C for growing until the optical density at 600 nm (OD_{600}) became 0.5-0.6, where OD was monitored by UV/Visible spectrophotometer (Ultrospec 2100 pro). Expression was induced by adding 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG), then the culture was incubated at 30 °C with shaking for 6 h. Bacterial cells were harvested by centrifugation (Sorvall LYNX 4000 centrifuge) at 9000 rpm and 4 °C for 30 min. To express ^{15}N labeled OnubPBP3, the saturated overnight bacterial culture was diluted (1:50 v/v) in M9 minimal medium. The medium contained 1.2 g/L [^{15}N] ammonium chloride, 1 mg/L thiamine, 10 µg/ml tetracycline, 100 µg/ml ampicillin, 100 µM trace elements, 2 mM $MgSO_4$, and 50 µM $CaCl_2$. The cells were allowed to grow at 37 °C until OD_{600} reached to 0.5-0.6. Expression induction, incubation conditions, and cell harvesting were the same as mentioned above for the expression of the unlabeled OnubPBP3, except the time increased to 16 h.

2.2.2 Refolding of OnubPBP3

The harvested bacterial cells from 0.5 L of medium were resuspended in 15 ml of Bacterial Protein Extraction Reagent (B-PER, Thermo-Scientific) containing 1 mM EDTA and lysed by sonication under ice-cold conditions (10 cycles of 5 s each with a 1 min gap) followed by centrifugation at 12000 rpm and 4 °C for 30 min. The inclusion bodies (IBs) were collected and washed two times by sonication with dilute B-PER solution. Refolding of the protein was performed by step dialysis at 4 ºC overnight to obtain an active protein. Approximately 0.6 g of wet IBs was solubilized in 25 ml of 50 mM Tris-HCl buffer of pH 8.0 containing 6 M guanidine hydrochloride (GuHCl) and 10 mM DTT followed by overnight incubation at room temperature. The sample was centrifuged at 12000 rpm and 4 °C for 30 min to remove undissolved particles. The supernatant was diluted with an equal volume of a solution containing 50 mM Tris-HCl and 2 M GuHCl. The protein solution was dialyzed at 4 ºC in 4 L of buffer 1 (DB #1) containing 50 mM Tris-HCl and 2 M GuHCl. Subsequently, the solution was dialyzed in 2 L of buffer 2 (DB #2) containing 1 M GuHCl, 0.8 M arginine, 0.9 mM oxidized glutathione and 3 mM reduced glutathione in 50 mM Tris-HCl. After that, the solution dialyzed in 2 L of buffer 3 (DB #3) containing 0.5 M GuHCl, 0.4 M arginine, and 0.45 mM oxidized glutathione, 1.5 µM reduced glutathione in 50 mM Tris-HCl. Before the last dialysis step, the protein solution was collected and centrifuged at 12000 rpm and 4 °C for 30 min to remove any precipitate and suspended particles. Finally, the clear supernatant was dialyzed in 2 L of buffer 4 (DB #4) containing 50 mM Tris-HCl, 0.1 M arginine, 250 mM NaCl, 0.9 mM oxidized glutathione, and 3 mM reduced glutathione. The flowchart in Figure 2.1 summarizes the whole process of protein refolding.

2.2.3 Purification

The recombinant OnubPBP3 was purified by three techniques: dialysis, anion exchange chromatography, and size exclusion chromatography. The refolded protein (50 ml) was dialyzed

using a 3000 Da cutoff membrane in 4 L of 20 mM Tris-HCl solution, pH 8.0 at 4 ºC overnight. Dialyzed refolded protein was centrifuged, filtered by 0.2 µm syringe filter, and subsequently purified by anion exchange chromatography using a HiPrep DEAE FF 16/10 column equilibrated with 20 mM Tris-HCl, at pH 8.0. The separated protein was eluted by using a NaCl gradient. The protein of interest was collected from specific fractions and concentrated to 4 ml via a Millipore ultrafiltration concentrator (capacity 15 ml, molecular weight cutoff (MWCO) 3000). The concentrated protein sample was then purified by size exclusion chromatography (SEC) using a Superdex 75 column fitted to an AKTA FPLC (GE Healthcare). A buffer containing 20 mM sodium phosphate, pH 6.5 with 150 mM NaCl, 1 mM EDTA, and 0.01% sodium azide was used to elute the separated protein. The pure monomeric protein was collected from specific fractions and stored at 4 ºC for further downstream applications.

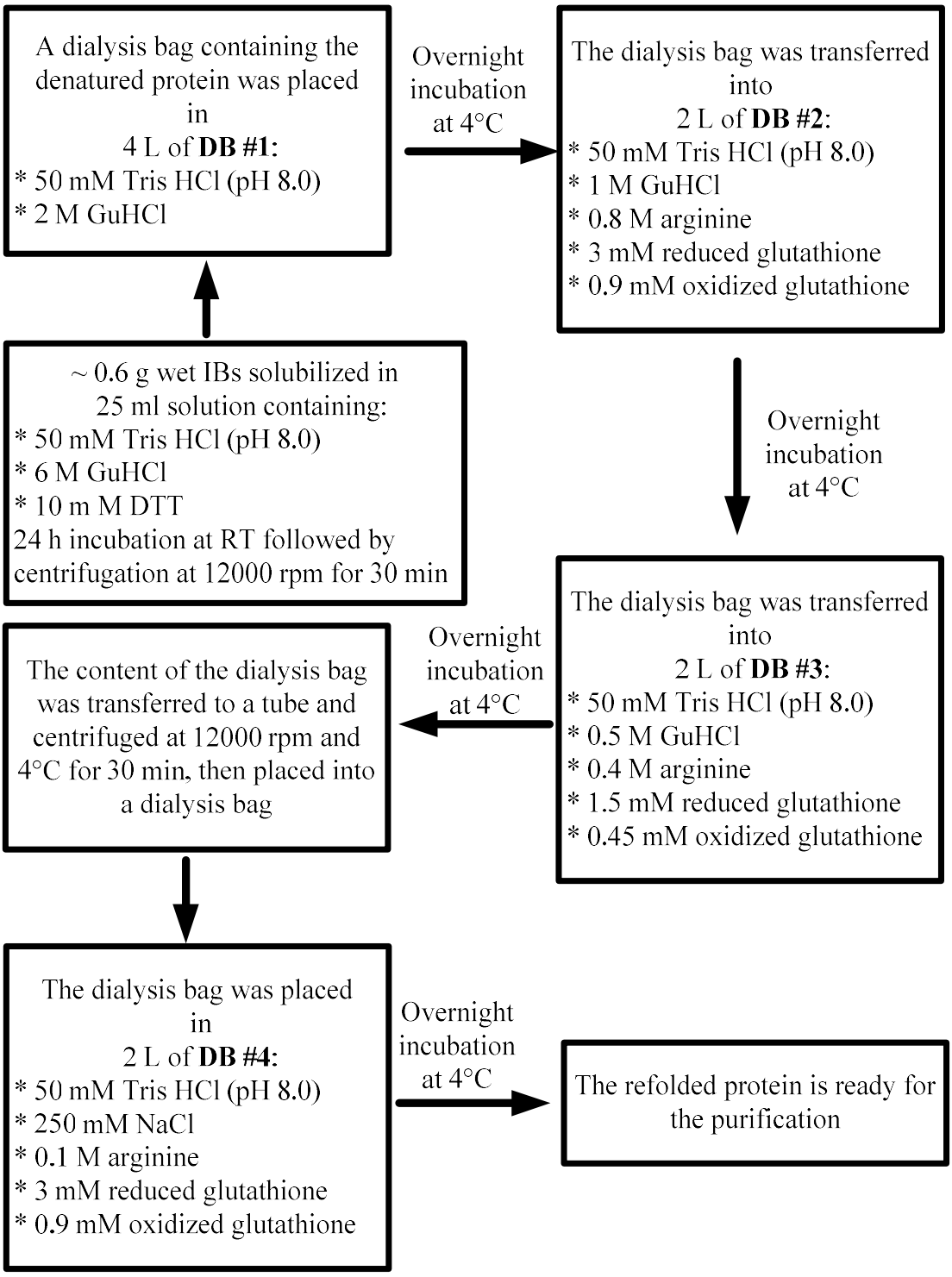

Figure 2.1: Refolding protocol using step dialysis to refold OnubPBP3: the protocol includes five steps using four different dialysis buffers. The entire process was performed at 4 ºC.

2.2.4 Protein delipidation

OnubPBP3 was delipidated by LipidiexTM-1000 resin (PerkinElmer) loaded in a column. This protocol was modified by Dr. Smita Mohanty from protocol used for ApolPBP1 purification.[61] The height of the packed Lipidiex and the diameter inside the column were 5.7 cm and 1.4 cm, respectively. The packed column was washed thoroughly with nanopure water to remove any trace of methanol used to store the Lipidiex. Sodium citrate buffer (50 mM, pH 6.5), named buffer A, was then used to equilibrate the column. Meanwhile, OnubPBP3 was loaded to a Millipore ultrafiltration concentrator (capacity 15 ml, MWCO 3000) and buffer exchanged to buffer A. After concentrating to 1.5 ml, the protein sample was transferred to the Lipidiex column and incubated for 30 min at 37 °C. Next, the free protein was eluted with buffer A directly to the Millipore ultrafiltration concentrator. The elution process was monitored by a UV/Visible spectrophotometer (Ultrospec 2100 pro). Finally, the free protein was concentrated, and buffer-exchanged with 15 mM sodium phosphate buffer (pH 6.5) to be ready for fluorescence binding and competitive displacement assays. To prepare the NMR sample, the delipidated ^{15}N-labeled protein was exchanged with a buffer containing 50 mM phosphate buffer at pH 6.5, 95% H_2O, 5% D_2O, 1 mM EDTA, and 0.01% (w/v) NaN_3 by the Millipore ultrafiltration concentrator.

2.2.5 Fluorescence spectroscopy

Fluorecence binding assy

Fluorescence binding assay of OnubPBP3 was conducted with *N*-phenyl-1-naphthylamine (1-NPN), a fluorescent probe using an LS-55 Fluorescence Spectrometer (PerkinElmer). The sample for fluorescence measurement was prepared in 3ml solution containing 15 mM sodium phosphate buffer (pH 6.5), 0.3% methanol, and 1 μM delipidated OnubPBP3 at 22 °C. A 2 mM stock solution of 1-NPN in methanol was also prepared for the titration experiment. The mixture was equilibrated for 10 min at room temperature in a quartz cuvette (1 cm light path length) before each

measurement. Emission and excitation slit widths were set to 7 nm and 4.7 nm, respectively. The fluorescence spectra were recorded at an excitation wavelength of 337 nm, an emission wavelength of 400-600 nm, and a speed of 500 nm/min. The binding of 1-NPN to delipidated OnubPBP3 was studied by monitoring the increase in the 1-NPN fluorescence at 409 nm. To the 1 μM protein solution, aliquots were added successively from a 1- NPN stock solution to a final concentration of 0–25.6 μM. All experiments were performed in triplicate and the spectra were corrected by employing proper controls. The k_{NPN} value was determined using the single site binding equation from Origin 2019 (Eq. 2.1).

$$y = B\,x / (k + x) \tag{2.1}$$

where B is the maximum fluorescence intensity which represents the maximum binding capacity, k is the dissociation constant, x is the pheromone concentration, and y is the fluorescence intensity at the specific ligand concentration.

Table 2.1: Fluorecence binding assay. Ratio of Protein:Ligand and volume of the 1-NPN added at each step of the titration with delipidated OnubPBP3.

(P:L) ratio	Ligand added (μl)	Total ligand added (μl)
1:0	0	0
1:0.1	0.15	0.15
1:0.2	0.15	0.30
1:0.3	0.15	0.45
1:0.4	0.15	0.60
1:0.5	0.15	0.75
1:0.6	0.15	0.90
1:0.8	0.30	1.20
1:1.0	0.30	1.50
1:1.4	0.60	2.1
1:1.8	0.60	2.7
1:2.4	0.9	3.6
1:3.2	1.20	4.8
1:4.0	1.20	6.0
1:5.8	2.7	8.7
1:7.6	2.7	11.4
1:10.6	4.5	15.9
1:13.6	4.5	20.4

Competitive displacement assay

Competitive displacement assay was carried out to determine the affinity of *E*-11-14:OAc and *Z*-11-14:OAc pheromone to OnubPBP3. This assay involved the displacement of 1-NPN from the OnubPBP3:1-NPN complex with each pheromone separately while monitoring the fluorescence of 1-NPN. Briefly, 3 ml solution containing 15 mM sodium phosphate buffer (pH 6.5), 0.3% methanol, and 2 μM delipidated OnubPBP3 was equilibrated overnight with 2 μM 1-NPN at 4 °C in the dark. Emission and excitation slit widths were set to 7 nm and 4.7 nm, respectively. The fluorescence spectra were recorded at an excitation wavelength of 337 nm, an emission wavelength of 400-600 nm, and a speed of 500 nm/min. Aliquots of pheromone were added from a 1 mM stock solution followed by a 10 min incubation at room temperature before recording the spectrum. Control experiments were performed for each addition of pheromone to a 2 μM 1-NPN solution in the absence of the protein. The IC_{50} values were determined using the equation below from Origin 2019 (Eq. 2.2).

$$\mathbf{y = 1- x / (k + x)} \qquad \mathbf{(2.2)}$$

where k is the IC_{50}, x is the pheromone concentration, and y is the fluorescence intensity at the specific ligand concentration.

To calculate the K_d of delipidated OnubPBP3 with each pheromone the $K_{1\text{-}NPN}$ and IC_{50} values were substituted in the following equation (Eq. 2.3).

$$\mathbf{K_d = [IC_{50}]/(1 + [1\text{-}NPN]/K_{1\text{-}NPN})} \qquad \mathbf{(2.3)}$$

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 2.2 : Fluorescence competitive displacement assay. Ratio of Protein:Ligand and volume of the ligand added at each step of the titration of the pheromone (E-11-14:OAc or Z-11-14:OAc) with delipidated OnubPBP3.

(P:L) ratio	Ligand added (µl)	Total ligand added (µl)
1:0.0	0	0
1:0.025	0.15	0.15
1:0.050	0.15	0.30
1:0.075	0.15	0.45
1:0.090	0.15	0.60
1:0.115	0.15	0.75
1:0.150	0.15	0.90
1:0.200	0.30	1.20
1:0.250	0.30	1.50
1:0.300	0.30	1.80
1:0.400	0.60	2.40
1:0.500	0.60	3.00
1:0.700	1.20	4.20
1:0.900	1.20	5.40
1:1.100	1.20	6.60
1:1.400	1.80	8.40
1:1.800	2.40	10.8
1:2.400	3.60	14.4

2.2.6 NMR experiment and data analysis

NMR samples consisted of 460 µM uniformly ^{15}N-labeled OnubPBP3 in 50 mM phosphate buffer at pH 6.5, 95% H_2O, 5% D_2O, 1 mM EDTA, and 0.01% (w/v) NaN_3 in a shaped tube. NMR data were collected at 35 °C on Bruker Neo 800 MHz spectrometer. The two dimensional-[^{1}H,^{15}N] heteronuclear single quantum coherence (HSQC) experiment was performed on the undelipidated and delipidated OnubPBP3 samples at pH 6.5. The HSQC spectra were collected using the Bruker pulse sequence hsqcetgpsi2. The transmitter frequency offset was set to 117 ppm in the ^{15}N dimension and 4.7 ppm in the ^{1}H dimension. The spectral width was 12500 Hz and 2919.708 Hz for ^{1}H and ^{15}N dimensions, respectively. The data were collected with 8 scans and 2048 complex points in the ^{1}H dimension and 300 complex points in the ^{15}N dimension. For ligand titration

experiments, the uniformly ^{15}N-labeled OnubPBP3 was placed in two shaped tubes; the concentration and volume of the protein sample in each tube were 460 µM and 270 µl, respectively. One sample was titrated with *E*-11-14:OAc pheromone and the other sample was titrated with *Z*-11-14:OAc pheromone. During the titration, the concentration of the pheromone was increased gradually (0–2.2 mM). Aliquots of pheromone dissolved in methanol were added from a 50 mM stock solution followed by 70 min incubation before recording the spectrum (Table 2.1). In both titrations, the corresponding two-dimensional HSQC spectra were recorded. The collected data were processed and analyzed by NMRPipe[93] and NMRFAM-SPARKY,[94] respectively.

Table 2.3: NMR ligand titration. Concentrations and ratio of Protein:Ligand and volume of the ligand added at each step of the titration of the pheromone (E-11-14:OAc or Z-11-14:OAc) with delipidated ^{15}N-labeled OnubPBP3.

(P:L) µM	(P:L) ratio	Ligand added (µl)	Total ligand added (µl)
460:0	1:0.0	0	0
460:92	1:0.2	0.5	0.5
460:276	1:0.6	1.0	1.5
460:460	1:1.0	1.0	2.5
460:1380	1:3.0	5.0	7.5
460:2300	1:5.0	4.9	12.4

2.3 Results and Discussion

2.3.1 Expression, refolding, and purification

The overexpression of OnubPBP3 was successfully accomplished by using a pET21-OnubPBP3 plasmid and *E. Coli* Origami 2 cells. Optimization of the expression conditions including IPTG concentration and growth temperature was carried out to obtain soluble recombinant OnubPBP3.

However, all the attempts under the attempted conditions with Origami 2 cells failed to avoid the production of IBs and produced instead the soluble protein (Figure 2.2A and B). Therefore, the IBs were solubilized by GuHCl, and the denatured protein was refolded to produce a biologically active recombinant protein. The resulting refolded protein was purified by anion exchange chromatography using a diethylamino ethanol (DEAE) column fitted to an AKTA Pure (GE

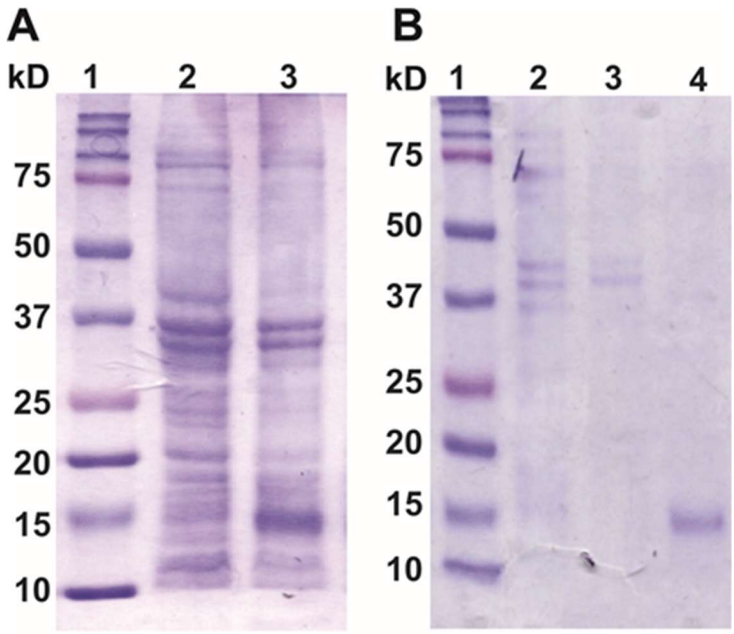

Figure 2.2: Coomassie-stained SDS-PAGE of the recombinant OnubPBP3. (**A**) SDS-PAGE of OnubPBP3; Lane 1: protein molecular weight marker; Lane 2: Origami cells before protein induction; Lane 3: Origami cells harvested after 6 h protein induction, (**B**) SDS-PAGE of OnubPBP3; Lanes 1: protein molecular weight marker; Lanes 2: Origami cells before protein induction; Lane 3: supernatant extracted from Origami 2 cells harvested after 6 h from the protein induction; Lane 4: pellet extracted from Origami 2 cells harvested after 6 h from the protein induction.

Healthcare) (Figure 2.3). After that, the protein fractions of interest were collected and purified by size exclusion chromatography (SEC) followed by SDS-PAGE to monitor the purity of the protein. The resulting size exclusion chromatogram contained a single sharp peak SDS-PAGE was run for

the fractions related to the single peak in SEC showed a single band close to the estimated molecular mass of OnubPBP3. SEC and SDS-PAGE analysis revealed the successful production of a homogeneous and very pure recombinant OnubPBP3 (Figure 2.4 A, B).

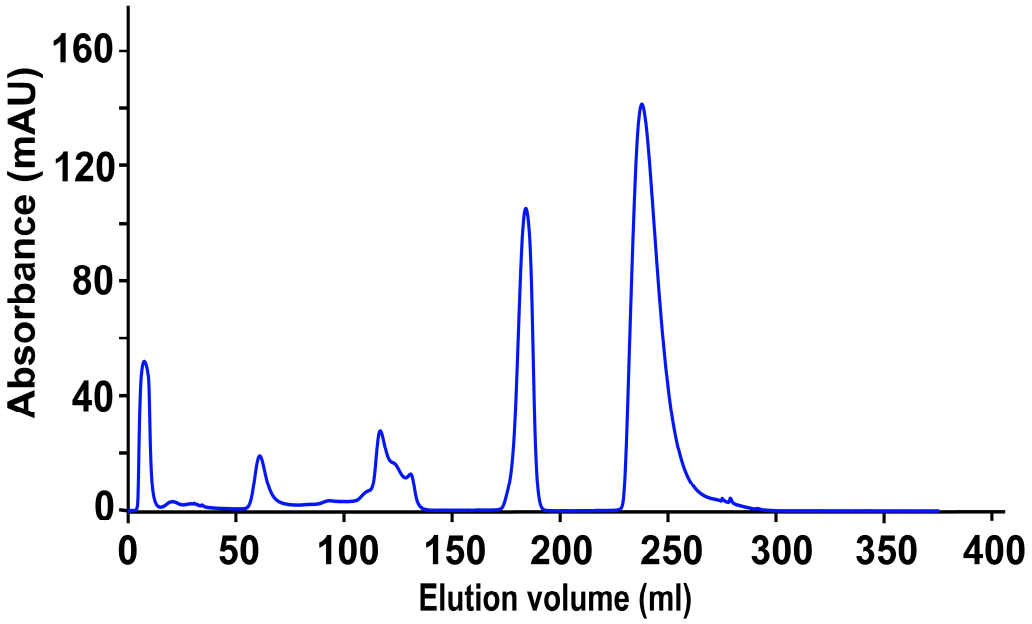

Figure 2.3: Anion exchange DEAE chromatogram of OnubPBP3; the monomer OnubPBP3 was eluted at volume range of 170-190 ml.

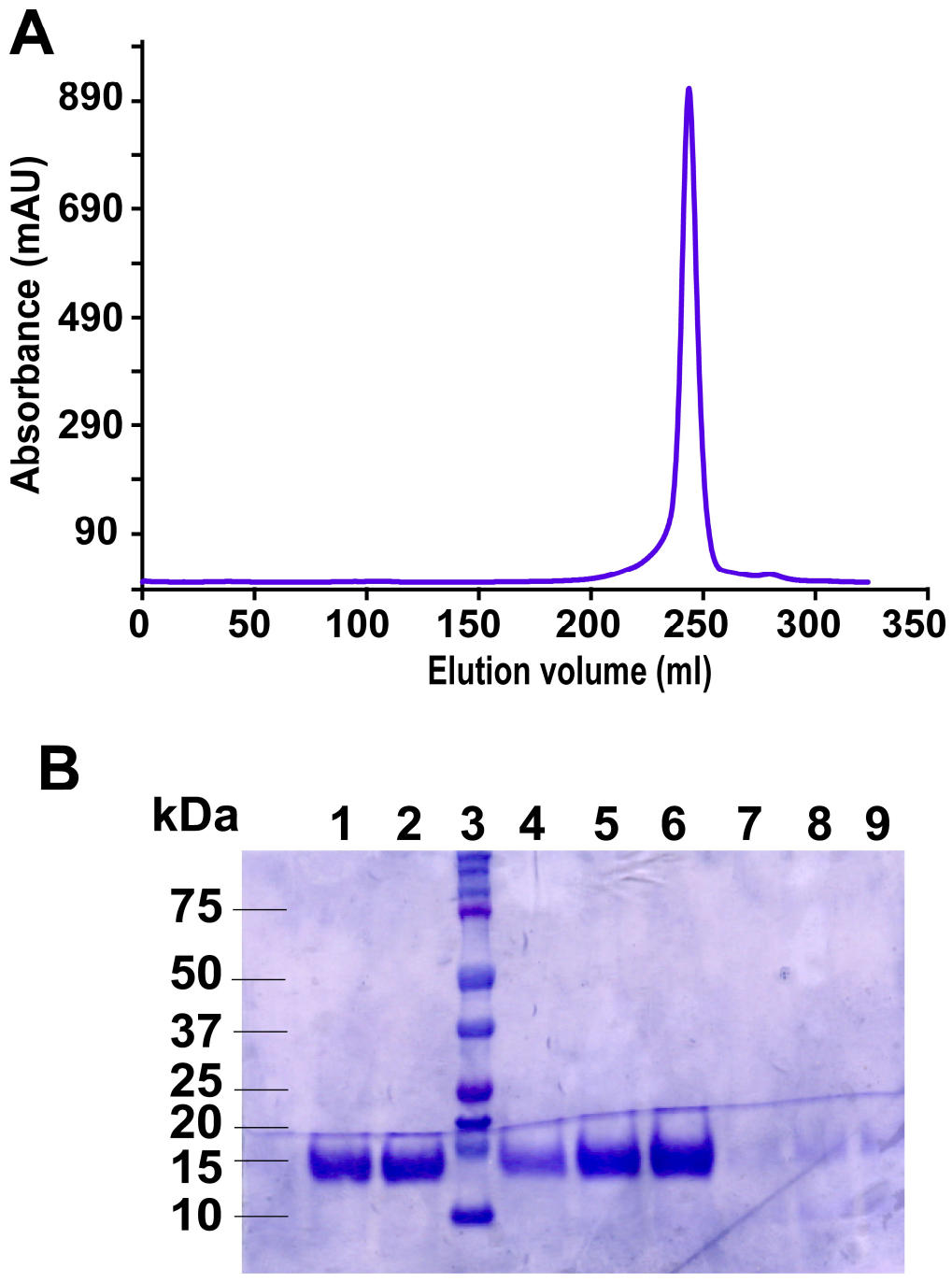

Figure 2.4: Size exclusion chromatography and Coomassie-stained SDS-PAGE of refolded and purified OnubPBP3. (**A**) Size exclusion chromatogram of OnubPBP3 (**B**) SDS-PAGE of OnubPBP3; Lane 1,2,4,5, and 6: pure protein from different fractions after size exclusion chromatography; lanes 3: protein molecular weight marker; Lanes 7,8, and 9: fractions after size exclusion chromatography which do not have protein.

2.3.2 Pheromone binding affinity by fluorescence

N-phenyl-1-naphthylamine (1-NPN) was employed as an external fluorescent probe, to study the hydrophobic cavity/pocket of the recombinant OnubPBP3. In an aqueous solution and without protein, 1-NPN showed a very low fluorescence intensity of 140 by adding 25.6 µM 1-NPN. However, the presence of undelipidated OnubPBP3 immensely amplified the intensity of this probe to 740 (Figure 2.5). The λ_{max} of the 1-NPN showed a blue shifting where the peak shifted to a shorter wavelength by binding to the hydrophobic cavity of the protein. For the dissociation constant $K_{1\text{-NPN}}$ determination, the normalized fluorescence intensity was plotted against the concentration of the free 1-NPN. The $K_{1\text{-NPN}}$ of the OnubPBP3:1-NPN complex was calculated equal to be 3.804 ± 0.830 µM (Figure 2.6).

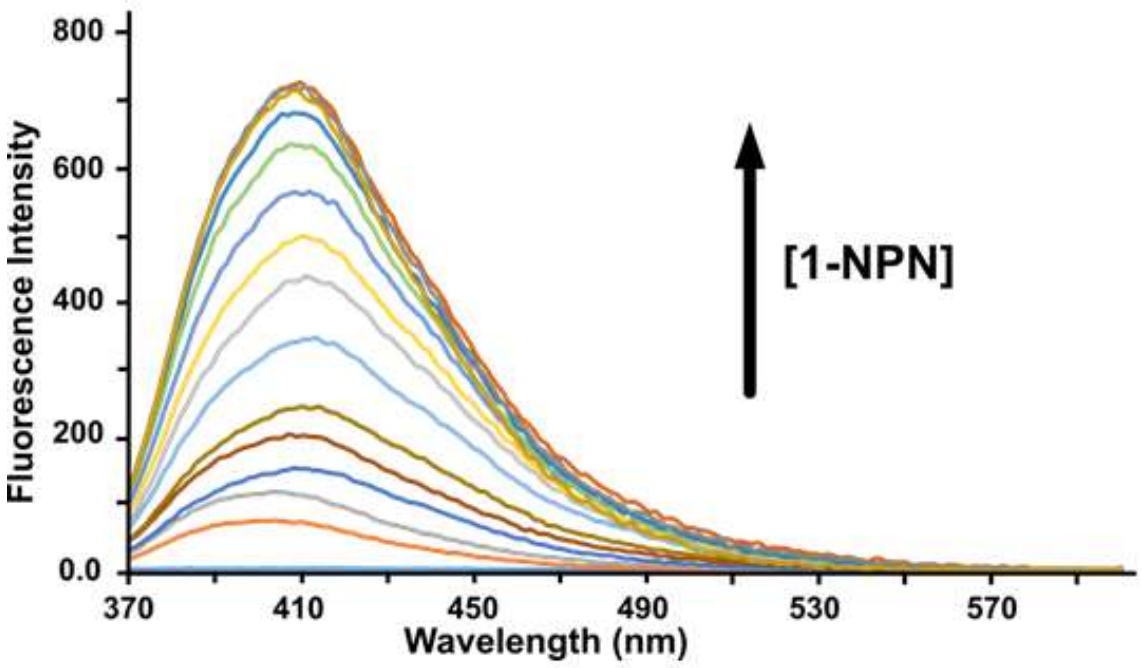

Figure 2.5: Fluorescence intensity increase in the binding assay of 1-NPN with OnubPBP3. OnubPBP3 (1 µM) in 15 mM phosphate buffer, pH 6.5, was titrated with increasing amounts of 1-NPN to a final concentration of 25.6 µM, the fluorescence intensity increased by adding 1-NPN.

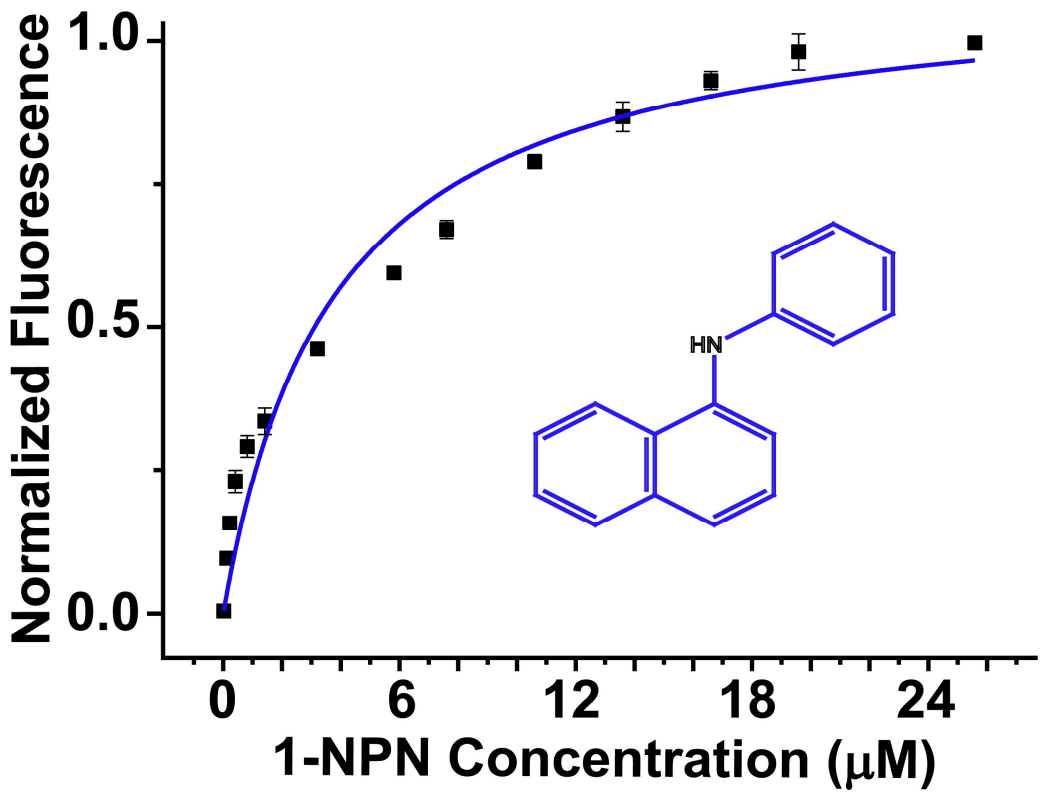

Figure 2.6: Binding assay of 1-NPN with OnubPBP3. Normalized fluorescence intensity as a function of 1-NPN concentration used to calculate $K_{1\text{-NPN}}$. The chemical structure of 1-NPN is shown in the plot. The assay were carried out in triplicate sets.

A fluorescent method based on competitive displacement of 1-NPN from a protein:NPN complex by a ligand has been frequently utilized to determine the dissociation constant (K_d) of the ligand.[89-91] In this experiment, as the aliquots of a ligand are added, the fluorescence intensity decreases until all 1-NPN is replaced by the ligand in the hydrophobic cavity. The binding affinity of the delipidated OnubPBP3 with the two sex pheromones of *O.nubilalis* was studied by carrying out the fluorescence competitive displacement assay using 1-NPN. The fluorescence intensity of 1-NPN decreased successively with the increase in the concentration of the *E*- or *Z*- pheromone (Figure 2.7). The reduction in the fluorescence intensity did plateau after the saturation point was reached

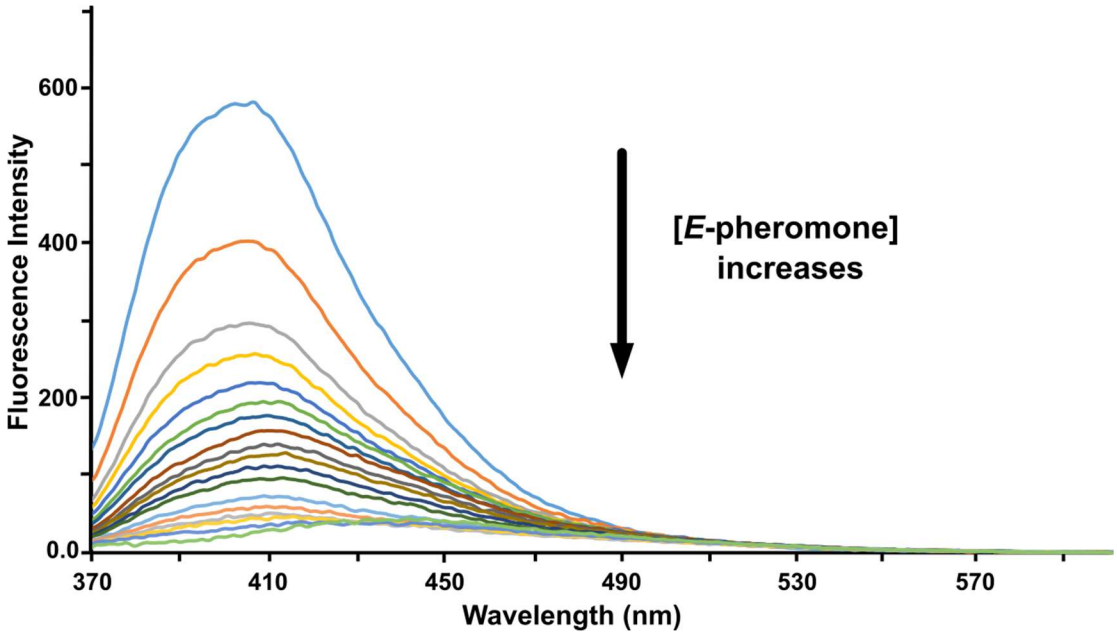

Figure 2.7: Fluorescence intensity decrease in the competitive displacement fluorescence binding assay [2 μM OnubPBP3 in 15 mM phosphate buffer, pH 6.5, in complex with 1-NPN (2 μM)]. 1-NPN was displaced by *E*-pheromone gradually leading to intensity reduction.

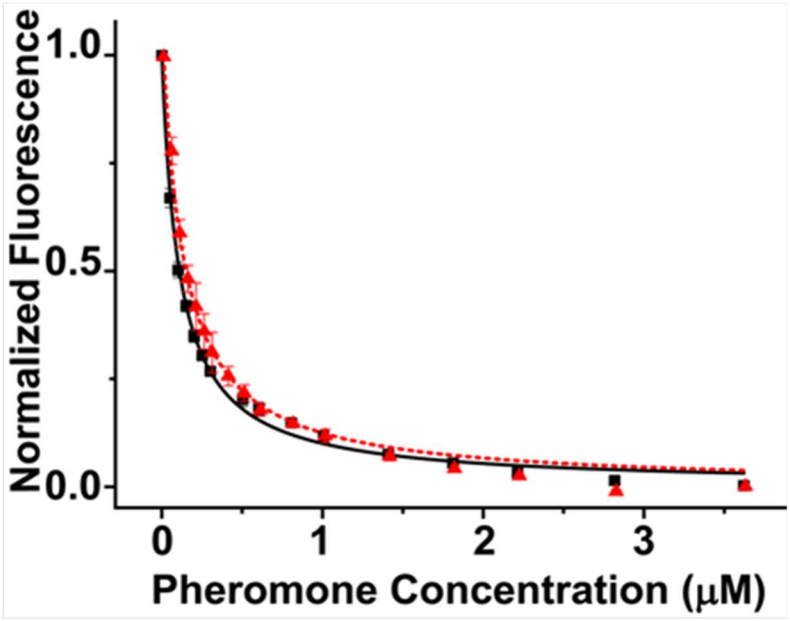

Figure 2.8: Competitive displacement fluorescence binding assay [2 μM OnubPBP3 in 15 mM phosphate buffer, pH 6.5, in complex with 1-NPN (2 μM)]. Normalized maximum fluorescence emission of 1-NPN was recorded after increasing concentrations (0–3.6 μM) of *E*-11-14:OAc (solid black) and *Z*-11-14:OAc (dashed red) were added. Both plots were employed to determine IC_{50} of each pheromone. The assays were carried out in triplicate sets.

when 1-NPN molecules in the hydrophobic pockets of the protein were replaced by pheromone molecules. The normalized fluorescence intensity was plotted against the concentration of the free pheromone to calculate IC_{50}. The IC_{50} of both pheromones, *E*- and *Z*-isomers, was determined to be 108.1 ± 3.13 nM, and 139.4 ± 5.14 nM, respectively (Figure 2.8). Subsequently, K_d values were calculated to be 71 ± 5.92 nM for *E*-pheromone and 91 ± 7.62 nM for *Z*-pheromone. The nanomolar affinity to each isomer of the pheromone, *E*- and *Z*-11-tetradecenyl acetate, indicates the recombinant OnubPBP3 is folded properly and present in the pheromone binding conformation at pH 6.5.

2.3.3 Effect of Delipidation on the Conformation of OnubPBP3

The two-dimensional [^{1}H, ^{15}N] HSQC spectrum provides the fingerprint of a protein. The HSQC spectrum of a protein is very sensitive and its pattern changes noticeably with any environmental changes, such as pH, temperature, solvent, salt concentration, and ligand binding. The changes in the values of the chemical shifts of the protein HSQC spectrum are an indication of a conformational change. In addition, HSQC is an experiment that can be trusted to check the suitability of a protein sample for structure determination. The HSQC spectrum of the refolded undelipidated OnubPBP3 at pH 6.5 exhibited good quality with well-dispersed peaks indicating a well-folded protein (Figure 2.9). Thus, the refolded protein is suitable for structural and functional studies by solution-state NMR. To check the impact of the delipidation on the recombinant OnubPBP3, an HSQC was collected for the delipidated protein at pH 6.5. The HSQC of the delipidated protein showed peak shifting and diminished peak dispersion (Figure 2.10). By overlaying NMR spectra of delipidated and undelipidated proteins, two discernible sets of peaks were found. Indeed, the HSQC spectra of both delipidated and undelipidated proteins were present in different conformations at pH 6.5 (Figure 2.11 A). The extended portion from the overlay HSQC spectra illustrates this conclusion (Figure 2.11 B).

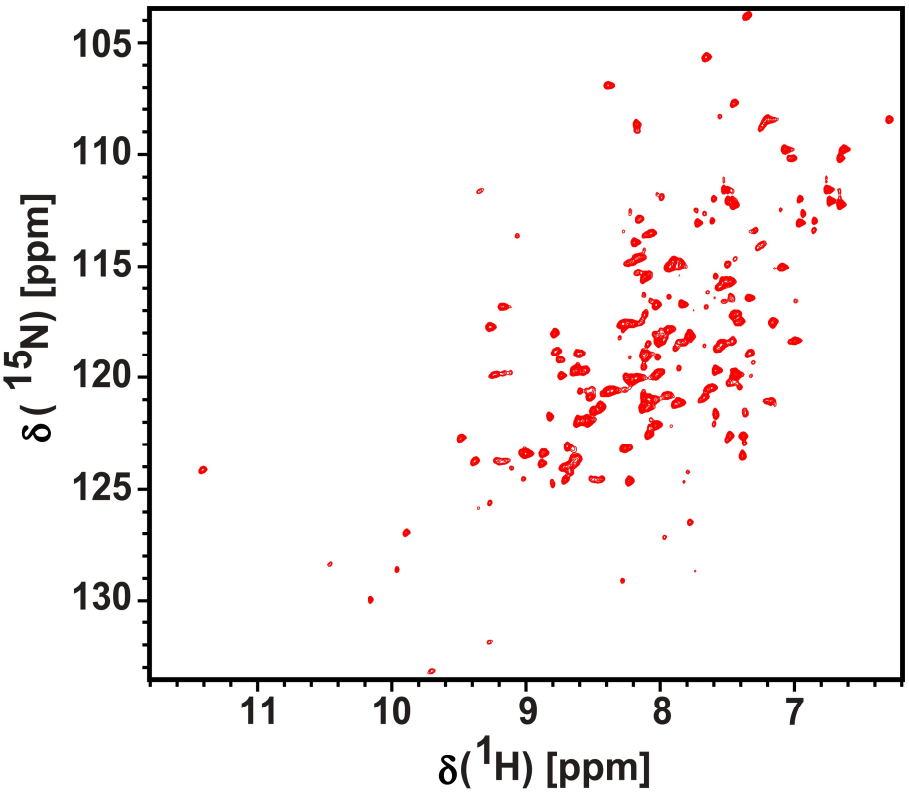

Figure 2.9: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of undelipidated OnubPBP3 at pH 6.5.

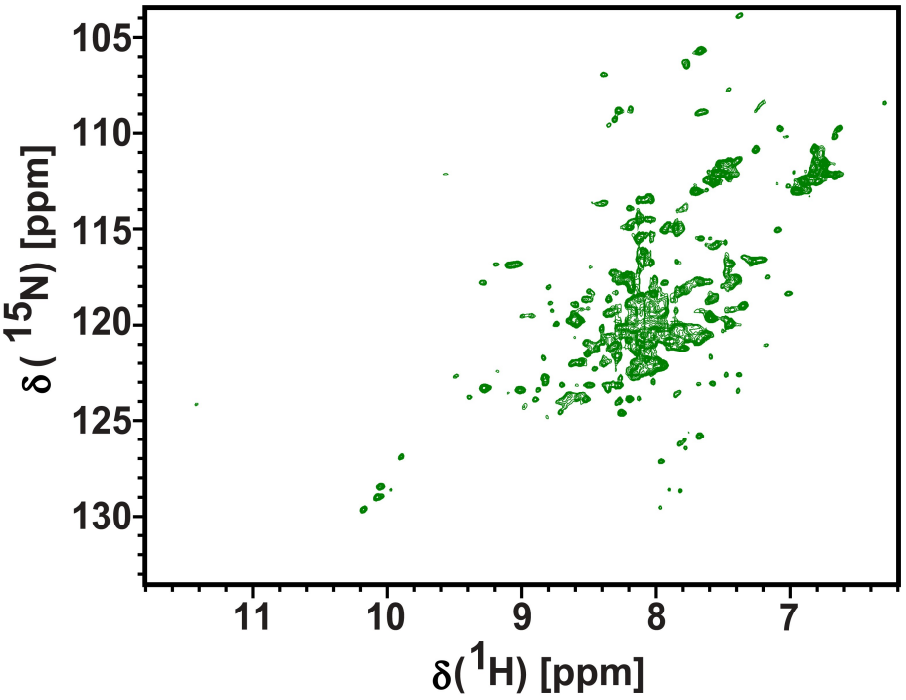

Figure 2.10: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of delipidated OnubPBP3 at pH 6.5.

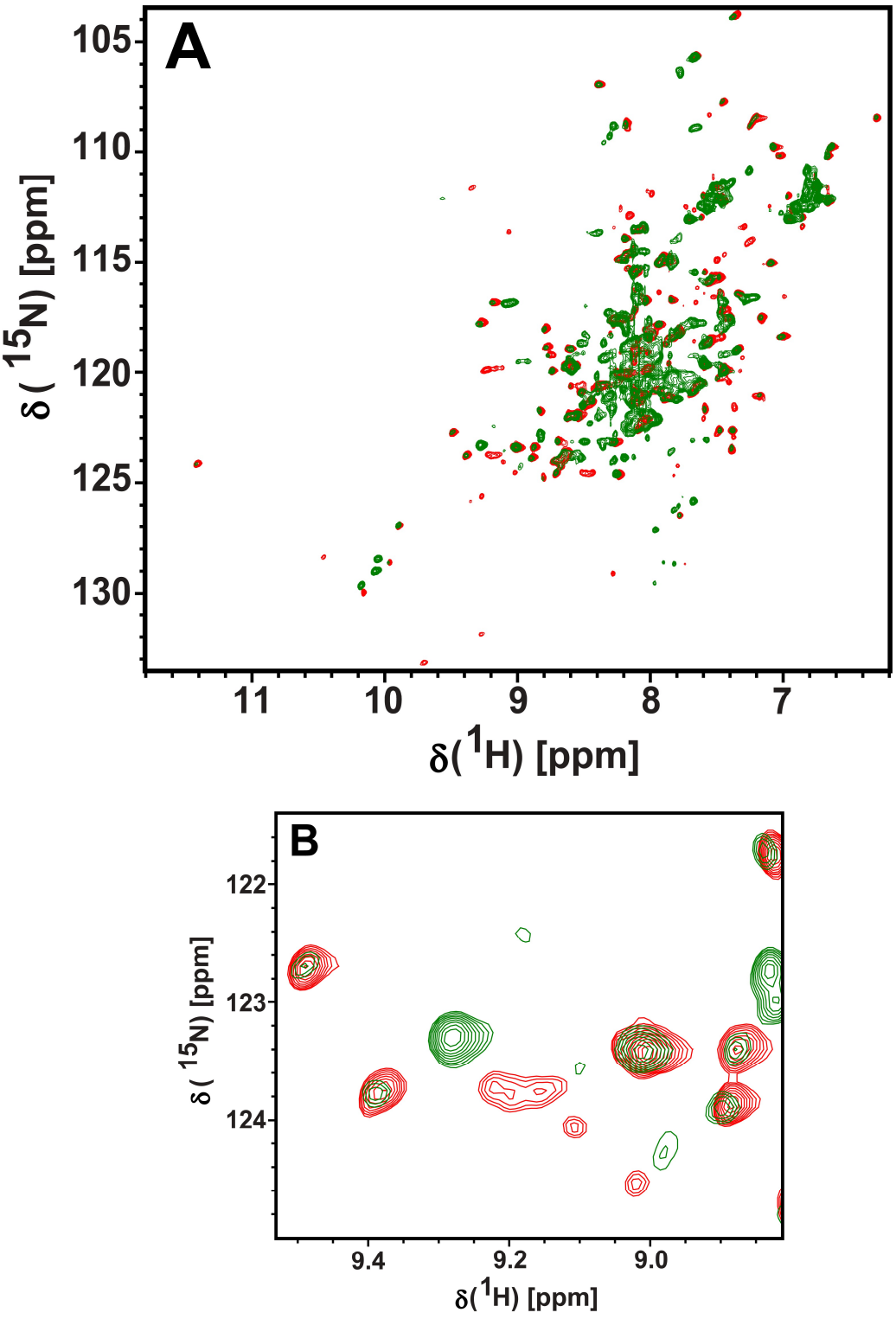

Figure 2.11: Overlay two-dimensional [^{1}H, ^{15}N] HSQC spectra of delipedated and undelipidated OnubPBP3. (**A**) Overly full spectra of delipidated (green) and undelipidated OnubPBP3 (red) (**B**) Expanded view of the boxed region from the overlay spectra of delipidated and undelipidated OnubPBP3.

2.3.4 Effect of Pheromones on the Conformation of Delipidated OnubPBP3

To study the effect of pheromones (*E*- and *Z*-11-tetradecenyl acetate) on the conformation of OnubPBP3, NMR ligand binding studies were performed on the delipidated protein at pH 6.5 with the two *O. nubilalis* sex pheromones. The effects of both pheromones on the conformation of delipidated OnubPBP3 were more or less the same. By overlaying NMR spectra collected at different pheromone concentrations, two sets of peaks were found for free and bound states (Figure 2.12 A, 2.13 A). The protein:pheromone ratio was changed through six steps 1:0, 1:0.2, 1:0.6, 1:1, 1:3, and 1:5 to see the influence of the pheromone concentration on the conformation. At an approximate protein:pheromone ratio of 1:1, the free conformation of the delipidated OnubPBP3 was mostly converted into the bound conformation. The peaks related to the delipidated conformation (free state) of the protein became progressively weak and finally disappeared, while the peaks of the bound conformation appeared and strengthened gradually with each addition of the pheromone. The extended portion from the overlay HSQC spectra at each concentration of both pheromones shows some of the peaks that were affected by pheromone binding to the protein (Figure 2.12 B, 2.13 B). Based on these observations, the interactions of OnubPBP3 with the pheromones are in slow exchange on the NMR time scale suggesting that both pheromones have very high affinities toward the protein.

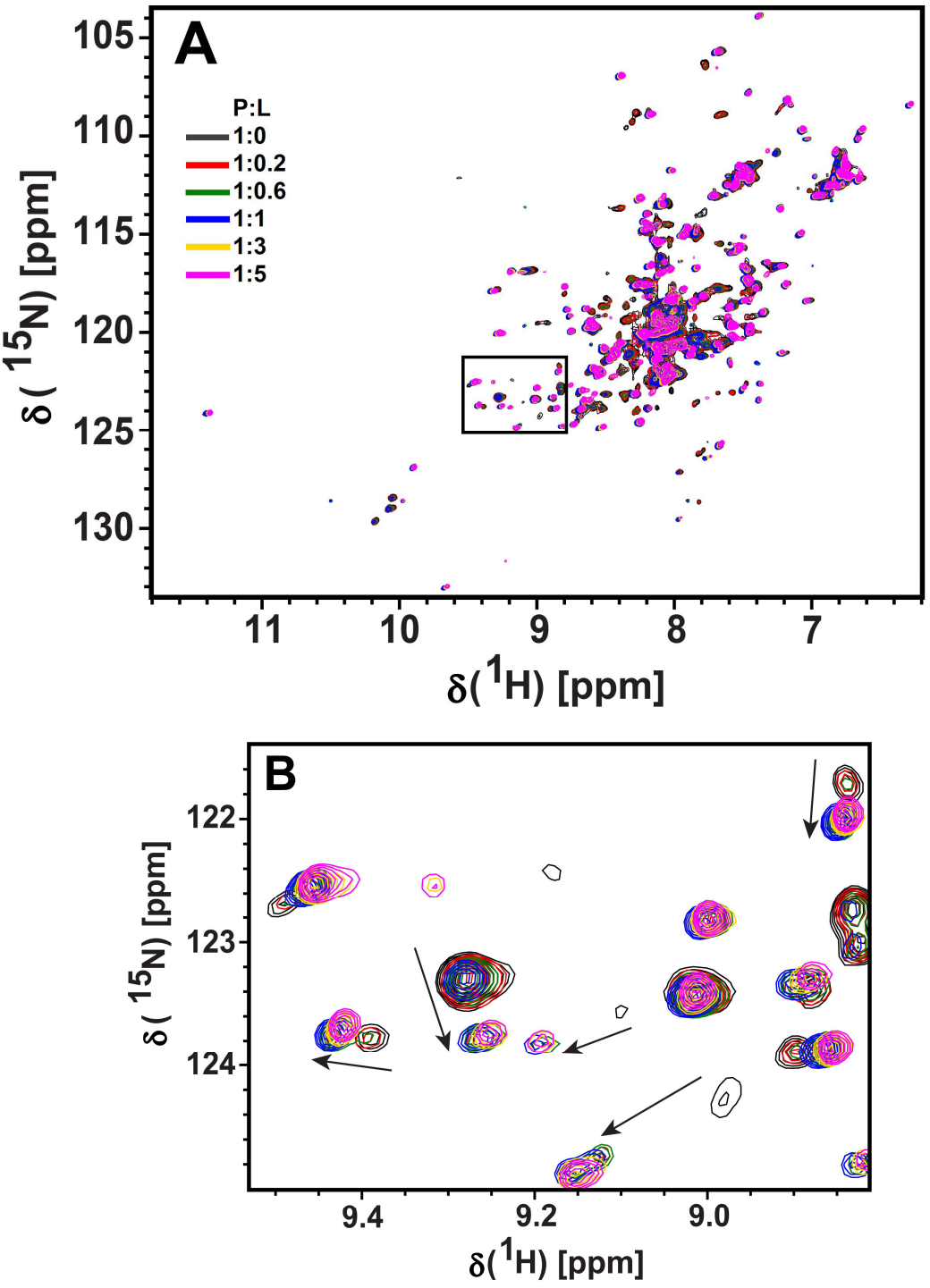

Figure 2.12: Superposition of two-dimensional [^{1}H, ^{15}N] HSQC spectra collected for OnubPBP3 at different *E*-pheromone concentrations. (**A**) superposition of the full HSQC spectra (**B**) The expansion shows the area marked by the solid box in full spectra. The peaks that show slow exchange during the pheromone titration were pointed by arrows.

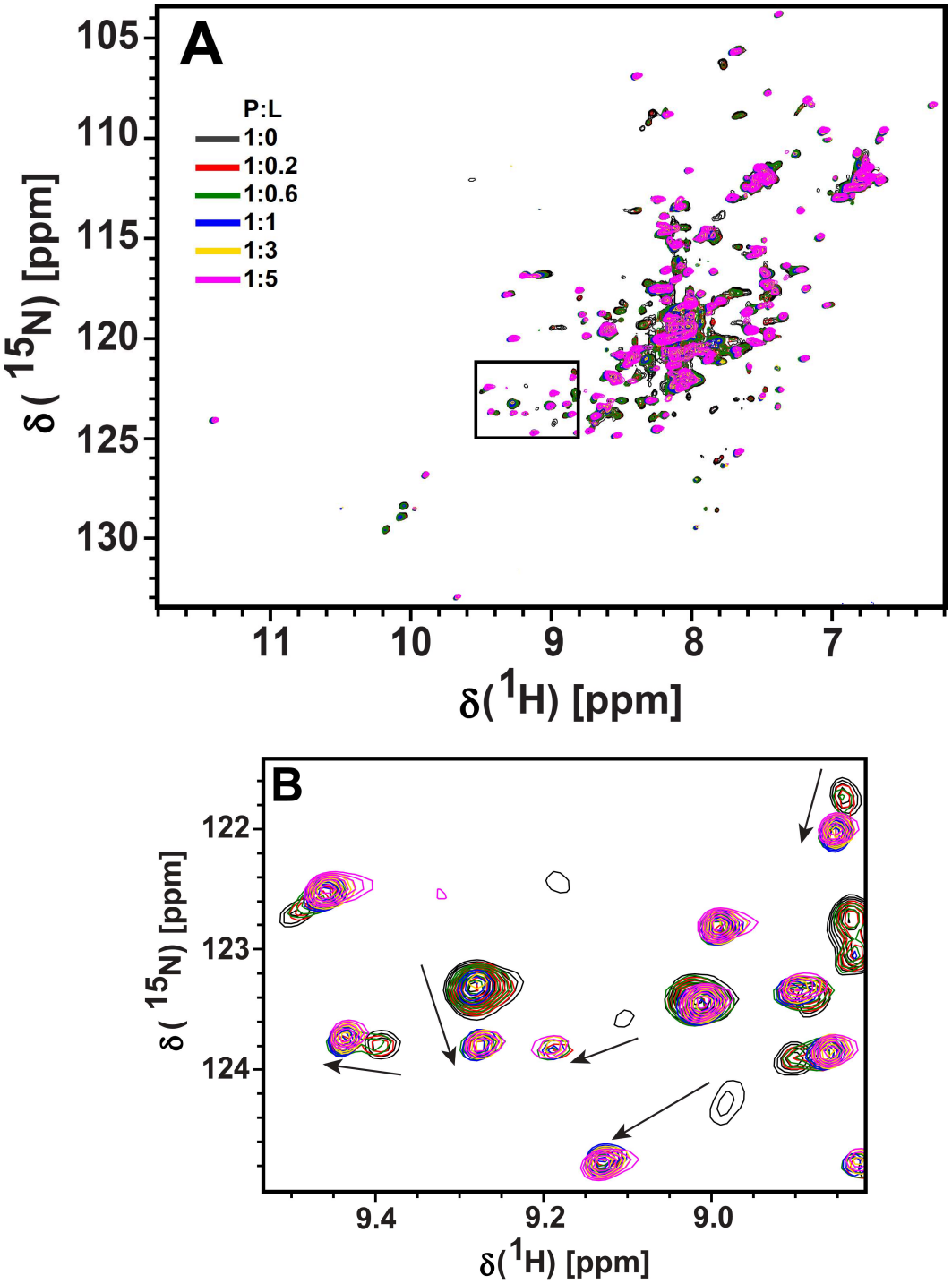

Figure 2.13: Superposition of two-dimensional [^{1}H, ^{15}N] HSQC spectra collected for OnubPBP3 at different *E*-pheromone concentrations. (**A**) superposition of the full HSQC spectra (**B**) The expansion shows the area marked by the solid box in full spectra. The peaks that show slow exchange during the pheromone titration were pointed by arrows.

To conclude the present study, we have demonstrated the overexpression, refolding, and purification of recombinant OnubPBP3 for the first time. The produced recombinant protein is highly active toward the *O. nubilalis* pheromones. The delipidation process applied in this work successfully removed the endogenous *E. coli* lipid that occupied the hydrophobic pocket of the recombinant OnubPBP3. The [^{1}H,^{15}N] HSQC spectrum showed nicely dispersed peaks suggesting that the produced protein is amenable to further NMR investigations. The NMR studies showed that the free and bound OnubPBP3 are present in different conformations at pH 6.5.

CHAPTER III

NMR ASSIGNMENTS OF ONUBPBP3

3.1 Introduction

The biological functions of proteins directly depend on the arrangement of their atoms in the three-dimensional structure. The ability of proteins to bind to their ligands for specific functions depends on the orientation and location of some amino acid residues, that form a binding site, in the tertiary structure of the protein. Obtaining the three-dimensional structure of a protein provides a greater level of understanding of how the protein works and how it can be inhibited. For example, having the three-dimensional structure of a protein helps to predict molecules that bind to the protein or design site-directed mutations[95] with the intent of understanding the roles of some residues in the protein function. To have a deep understanding of pheromone recognition, specificity, and the mechanism of pheromone binding by OnubPBP3, we have initiated a detailed structural investigation of this protein by solution-state NMR spectroscopy.

The typical process of protein structure determination via NMR spectroscopy usually requires a few essential steps. Highly specialized techniques are needed in each step. The required steps include the following:

1. Sample preparation

   Overexpression of a double-labeled protein with two isotopes, ^{13}C and ^{15}N, is needed to prepare a sample for NMR structure calculation.

2. Data acquisition

   This step includes collecting a series of 2D and 3D heteronuclear NMR spectra.

3. Resonance assignments

   Three different resonance assignments are employed in this step including backbone, side chain, and NOE assignments.

4. Extraction of structural information from the spectra

   Structural restraints must be generated from different spectra to be able to calculate the protein structure. The restraints used mostly in structure determination are interproton distances derived from the NOE information and torsion angles derived from the backbone and side-chain chemical shifts.

5. Structure calculation by computer programs

   The three-dimensional structure calculation of the protein is initiated by employing the NOE and dihedral angle restraints. Both types of restraints are essential input information for computer programs like CYANA or XPLOR-NIH to calculate the protein structure from NMR data.

3.1.1 Backbone assignments and secondary structure prediction

Triple resonance NMR experiments (e.g. 3D HNCACB, 3D CBCA(CO)NH, 3D HNCA, 3D HN(CO)CA, 3D HNCO, and 3D HN(CA)CO) are used commonly to accomplish sequential backbone assignments of a protein. The basics of these experiments and the information that they can provide were discussed previously in Chapter 1. Tables 3.1 shows the correlations provided by these six experiments. Sequential NMR backbone assignments are a crucial initial step in protein structure determination by which the resonances of the backbone nuclei (^{1}H, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$) of the protein chain are assigned. Resonance assignments must be sequence-specific; in other words, each resonance must be assigned to a specific spin in an amino acid residue in the protein primary structure.

Table 3.1: The most commonly used triple-resonance NMR experiments for protein backbone assignment. The correlations provided by each experiment are listed in the table.

Experiment	Correlation
3D CBCA(CO)NH	Cβ(i-1), Cα(i-1), H^N(i), ^{H}N(i)
3D HNCACB	Cβ(i-1), Cα(i-1), Cβ(i), Cα(i), H^N(i), ^{H}N(i)
3D HNCA	H^N(i), ^{H}N(i), Cα(i) and Cα(i-1)
3D HN(CO)CA	H^N(i), ^{H}N(i), Cα(i-1)
3D HNCO	CO(i-1), H^N(i), ^{H}N(i)
3D HN(CA)CO	CO(i), H^N(i), ^{H}N(i)

The local structure of a protein and its NMR chemical shifts are highly correlated with each other. The dependence of the chemical shifts on the local structure provides the opportunity to obtain secondary structure elements and torsion angles (phi φ and psi ψ) (Figure 3.1) from chemical shifts. Indeed, particular chemical shifts belonging to residues they are part of an α-helix or β-sheet are different from the expected random coil. This deviation of chemical shifts depends on the type of secondary structure; therefore, it is a helpful value to predict the secondary structure. The deviation is called secondary chemical shift which is defined as Δδ (Eq. 3.1)

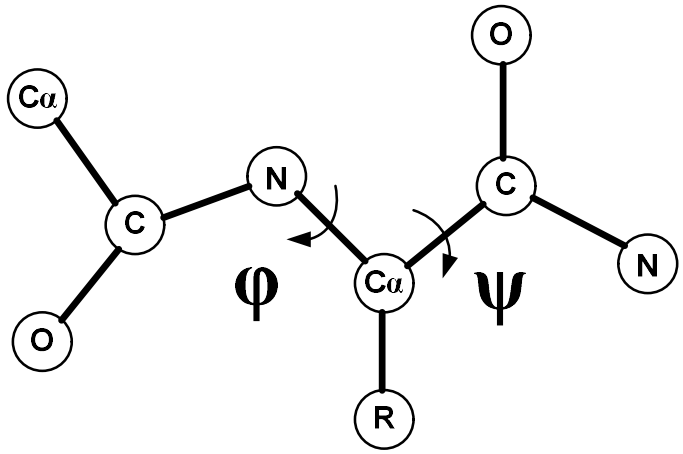

Figure 3.1: Schematic representation of the backbone torsion angles (φ and ψ) of proteins.

$$\Delta\delta = \delta_{observed} - \delta_{random\ coil} \quad (3.1)$$

The secondary chemical shift of α-helical Cα tends to be positive, while the value of Cα in a β-sheet tends to be negative. In comparison to Cα, the secondary chemical shift of Cβ in an α-helix or β-sheet is completely opposite. The chemical shifts of Hα, H^N, N^H, and C' can be helpful for the prediction of the secondary structure. However, Cα and Cβ chemical shifts are the most important values for the prediction. Showing Δδ Cα – Δδ Cβ values is a common practice in reporting secondary chemical shifts. The value of the difference is positive for an α-helix and negative for a β-sheet.[96]

The secondary structure can also be predicted by calculating the backbone torsion angles (φ and ψ) from the chemical shifts. The range of the φ and ψ torsion angles available for a polypeptide chain explains the flexibility of the backbone and the ability of the polypeptide to adopt a particular fold. A plot of the φ against ψ torsion angles of amino acid residues in a protein is known as the Ramachandran plot (Figure 3.2). The plot is used as a tool to view the distribution of torsion angles contained in a protein structure. The plot shows that both the torsion angles are locolized to discrete regions. Each region in the plot represents one of the major secondary structure elements, α-helices or β-sheets.

The TALOS+ is one of the programs that uses the chemical shifts of a protein to calculate backbone torsion angles. The torsion angles obtained with the help of artificial neural networks (ANNs) is then used to locate all secondary structures present in the protein. ANNs are utilized to compare chemical shift patterns of investigated proteins to a huge database of previously assigned proteins with high-resolution three-dimensional structures.[97]

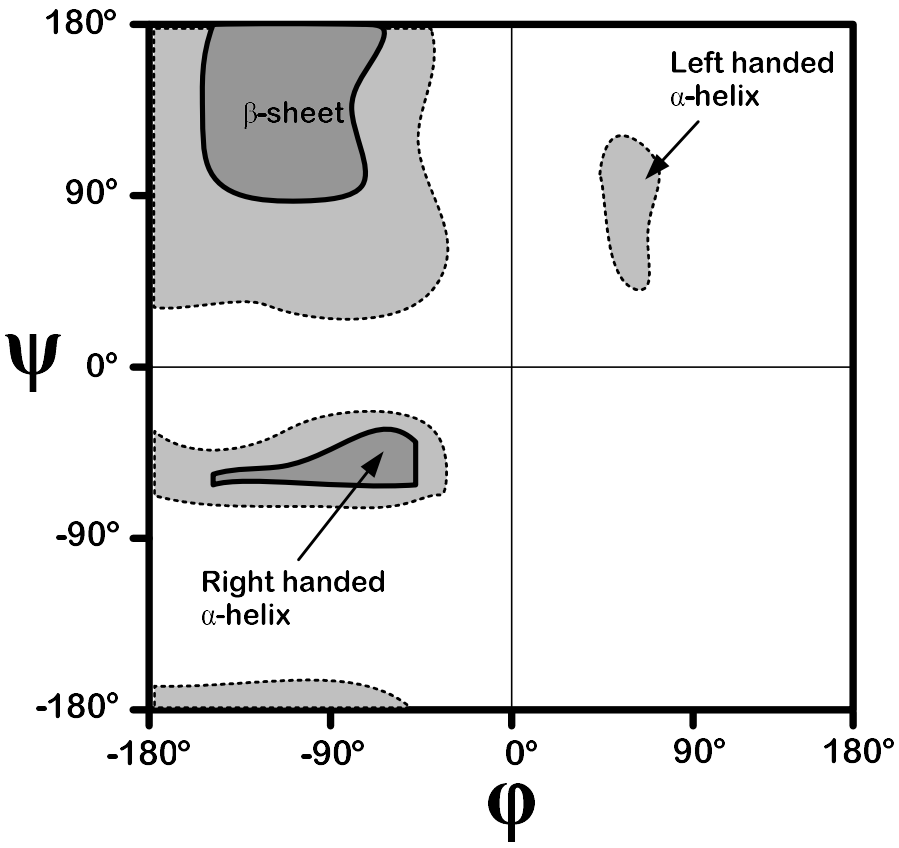

Figure 3.2: The Ramachandran plot shows sterically allowed φ and ψ angles of the secondary structure elements.

3.1.2 Side chain assignments

The determination of a three-dimensional protein structure by NMR spectroscopy depends on interproton distance restraints derived from NOE measurements. NOESY experiments, such as 3D ^{15}N-edited NOESY-HSQC and 3D ^{13}C-edited NOESY-HSQC are the source for NOE distance restraints. To assign 3D ^{15}N-NOESY-HSQC and 3D ^{13}C-NOESY-HSQC spectra, side chain assignments of proton atoms/spins are necessary. In addition to their use for three-dimensional structure determination, the side chain assignments are important for verifying the sequential backbone assignments by assigning a spin system to a residue type. The side chain assignments are straightforward in comparison to the sequential backbone assignments. There are different NMR experiments used to accomplish the side chain assignments, such as 3D H(CCO)NH, 3D ^{15}N-

TOCSY-HSQC, 3D HCCH-TOCSY, and 3D CC(CO)NH. Each of these experiments has something unique to offer. 3D H(CCO)NH and 3D CC(CO)NH provide the hydrogen and carbon side-chain chemical shifts, respectively, for the residue preceding each amide group (i-1). 3D ^{15}N-TOCSY-HSQC is utilized to obtain the hydrogen side-chain chemical shifts for the residue (i) with an NH group (Table 3.2). By using 3D HCCH-TOCSY, all side-chain hydrogen resonances are visible through each carbon frequency in the same residue.

Table 3.2: The most commonly used triple-resonance NMR experiments for protein side-chain assignment. The correlations provided by each experiment are listed in the table.

Experiment	Correlation
3D ^{15}N-HSQC-TOCSY	$H^{C}(i)$, $H^{N}(i)$, $^{H}N(i)$
3D H(CCCO)NH	$H^{C}(i-1)$, $H^{N}(i)$, $^{H}N(i)$
3D (H)CC(CO)NH	$H^{N}(i)$, $^{H}N(i)$, C(i-1)

3.1.3 NOE assignments

The structure determination of proteins by solution-state NMR spectroscopy relies mainly on interproton distance restraints, which are the distances between each pair of protons. To obtain the distance restraints of a protein, the NOE cross-peaks in nuclear Overhauser effect spectroscopy (NOESY) spectra must be assigned. The NOE assignment is the most critical stage to obtain the right three-dimensional structure. NOEs are assigned based on the chemical shifts earlier determined by sequential backbone and side-chain assignments. Depending on the distance between interacting hydrogen atoms, NOEs can be classified into short, medium, and long-range. The former two types of NOE cross-peaks are far more common in NOESY spectra than the latter. However, long-range NOEs are the most important ones for constraining the overall folding of proteins. Therefore, a large number of NOE cross-peaks have to be assigned to obtain accurate and

precise structural results. Practically, NOESY spectra are complicated due to peak overlap, spectral artifacts, and noise.[98] The extraction of a complete structural restraint set from NOESY spectra is an iterative process. At the beginning of the assignment, some NOESY cross-peaks are assigned unambiguously to produce preliminary structures or models from limited numbers of distance restraints. The preliminary structures are employed to decrease the ambiguity of the initial cross-peak assignments.

3.2 Methods and materials

3.2.1 Circular dichroism spectroscopy

The circular dichroism (CD) experiment was carried out to measure the melting temperature of the protein using a Jasco J-810 automatic recording spectropolarimeter. A quartz cell cuvette with a path length of 0.05 cm was used for this work. The far-UV CD data of the unlabeled OnubPBP3 were collected with a protein concentration of 30 μM in 15 mM phosphate buffer at pH 6.5. CD spectra of the 15 mM phosphate buffer at pH 6.5 were collected as a control experiment. To produce one final spectrum at each temperature, four repeat spectra were collected. The data were collected over the wavelength range from 185 to 250 nm and over the temperature range from 40 °C to 108 °C in 2 °C increments, with a ramp rate of 5 °C per min.

3.2.2 Protein sample preparation

The [^{1}H,^{13}C,^{15}N]-labeled recombinant OnubPBP3 protein was expressed in *E. coli* Origami 2 cells using pET21a vector. Saturated overnight Lysogeny broth (LB) bacterial culture containing ampicillin and tetracycline was diluted (1:50 v/v) in M9 minimal medium. The medium contained 4 g/L [^{13}C] glucose, 1.2 g/L [^{15}N] ammonium chloride, 1 mg/L thiamine, 10 μg/ml tetracycline, 100 μg/ml ampicillin, 100 μM trace elements, 2 mM $MgSO_4$, and 50 μM $CaCl_2$. After that, the

cells were allowed to grow at 37 °C until the OD_{600} reached 0.5. Ampicillin and tetracycline were added to the culture to grow the Origami 2 cells containing the OnubPBP3 gene. The fresh bacterial culture was shaken at 37 °C for growing until the optical density at 600 nm (OD_{600}) became 0.50, where the OD was monitored by UV/Visible spectrophotometer (Ultrospec 2100 pro). Expression was induced by adding 1 mM isopropyl β- D-1-thiogalactopyranoside (IPTG), and then the culture was incubated at 30 °C with shaking for 16 h. Bacterial cells were harvested by centrifugation (Sorvall LYNX 4000 centrifuge) at 9000 rpm and 4 °C for 30 min.

Extracting the inclusion bodies (IBs) from the cells were carried out by solubilizing the IBs, refolding the denatured ^{13}C,^{15}N-labeled protein, and purifying the refolded protein following the same protocols as described previously in Chapter 2. To prepare the NMR sample, the ^{13}C,^{15}N-labeled protein was exchanged four times with a buffer containing 50 mM phosphate buffer at pH 6.5, H_2O, 1 mM EDTA, and 0.01% (w/v) NaN_3 by the Millipore ultrafiltration concentrator (capacity 15 ml, molecular weight cutoff (MWCO) 3000). D_2O was added to the sample after the concentration to contain 5% D_2O.

3.2.3 NMR experiments

NMR data collection was accomplished at 35 °C on either a Bruker Avance II 800 MHz spectrometer equipped with a 5 mm triple resonance cryoprobe and Z-axis pulsed-field gradient at the National High Field Magnetic Laboratory, Tallahassee, Florida, or a Bruker Neo 600 MHz spectrometer at Oklahoma Statewide Shared Nuclear Magnetic Resonance Facility, Stillwater, Oklahoma. The data were collected by Dr. Bharat Chaudhary and Omar Al-Danoon. Backbone and side-chain assignments were carried out using the following 2D and 3D NMR experiments: 2D [^1H,^{15}N]-HSQC, 2D [^1H,^{13}C]-HSQC, 3D CACB(CO)NH, 3D HNCACB, 3D HNCA, 3D HN(CO)CA, 3D HNCO, 3D HN(CA)CO, 3D CC(CO)NH, 3D H(CCCO)NH, 3D HCCH-TOCSY, 3D ^{15}N-edited HSQC TOCSY, and 3D ^{15}N/^{13}C-edited NOESY (with mixing times of 120 ms). 2D

[^{1}H,^{15}N]-HSQC, 3D CACB(CO)NH, 3D HNCACB, and 3D ^{15}N-edited NOESY were carried out on a Bruker Neo 600 MHz spectrometer, while all other experiments were performed on the Bruker Avance II 800 MHz spectrometer. NMRPipe was employed to process and visualize all collected spectra and NMRFAM-SPARKY was used for analysis. The pulse program and the parameters used in these experiments are listed in Table 3.3.

Table 3.3: Shows all pulse programs and parameters of the used experiments.

Experiment Name	Pulse program	Frequency offset	Spectral width (Hz)	Number of scans	Complex point
[^{1}H,^{15}N]-HSQC	hsqcetgpsi2	$^{1}H_N$= 4.677, ^{15}N= 117.0	$^{1}H_N$= 9615.38, ^{15}N= 3035.82	32	$^{1}H_N$= 2048, ^{15}N= 512
3D HNCA	hncagp3d	$^{1}H_N$= 4.7, ^{15}N= 117, ^{13}C= 54	$^{1}H_N$=11160.71, ^{15}N= 2919.45 ^{13}C= 6451.91	8	$^{1}H_N$= 2048, ^{15}N= 80, ^{13}C= 228
3D HNCACB	hncacbgp3d	$^{1}H_N$= 4.678, ^{15}N= 116.984, ^{13}C= 47.641	$^{1}H_N$= 9615.38, ^{15}N= 2216.31 ^{13}C= 10548.523	16	$^{1}H_N$= 2048, ^{15}N= 40, ^{13}C= 150
3D CACB(CO)NH	cbcaconhgp3d	$^{1}H_N$= 4.7, ^{15}N= 116.983, ^{13}C= 45	$^{1}H_N$= 9615.385, ^{15}N= 2216.312 ^{13}C= 10548.523	16	$^{1}H_N$= 2048, ^{15}N= 40, ^{13}C= 128
3D HNCO	hncogp3d	$^{1}H_N$= 4.7, ^{15}N= 117, ^{13}C= 175	$^{1}H_N$= 9615.385, ^{15}N= 2216.312 ^{13}C= 1808.318	16	$^{1}H_N$= 2048, ^{15}N= 40, ^{13}C= 128
3D HN(CA)CO	hncacogp3d	$^{1}H_N$= 4.7, ^{15}N= 117, ^{13}C= 175	$^{1}H_N$= 9615.385, ^{15}N= 2216.312 ^{13}C= 1808.318	16	$^{1}H_N$= 2048, ^{15}N= 40, ^{13}C= 156
3D CC(CO)NH	hccconhgpwg3d3	$^{1}H_N$= 4.754, ^{15}N= 121, ^{13}C= 39	$^{1}H_N$= 12820.513, ^{15}N= 2919.471 ^{13}C= 15695.616	16	$^{1}H_N$= 2048, ^{15}N= 40, ^{13}C= 144

3D H(CCCO)NH	hccconhgp3d1	$^{1}H_N$= 4.7, ^{15}N= 121, ^{1}H= 4.7	$^{1}H_N$= 12820.513, ^{15}N= 2919.459 ^{1}H= 6401.870	16	$^{1}H_N$= 2048, ^{15}N= 80, ^{1}H= 200
3D ^{15}N-edited HSQC TOCSY	dipsihsqcf3gpsi3d	$^{1}H_N$= 4.7, ^{15}N= 117, ^{1}H= 4.705	$^{1}H_N$= 12820.513, ^{15}N= 2919.839 ^{1}H= 12803.74	8	$^{1}H_N$= 2048, ^{15}N= 100, ^{1}H= 300
3D ^{15}N-edited NOESY (mixing times of 120 ms)	noesyhsqcf3gpsi3d	$^{1}H_N$= 4.7, ^{15}N= 117, ^{13}C= 175	$^{1}H_N$= 9615.385, ^{15}N= 2216.312 ^{1}H= 1808.318	16	$^{1}H_N$= 2048, ^{15}N= 100, ^{1}H= 300
3D ^{13}C-edited NOESY (mixing times of 120 ms)	noesyhsqcetgp3d	$^{1}H_N$= 4.7, ^{15}N= 42, ^{1}H= 4.7	$^{1}H_N$= 12820.513, ^{15}N= 12878.492 ^{1}H= 12803.74	8	$^{1}H_N$= 2048, ^{15}N= 126, ^{1}H= 200

3.3 Results and Discussion

All NMR experiments were conducted at 35 °C providing spectra with peaks of good resolution. It has been reported that the resolution of peaks depends on temperature and varying the temperature may increase the peak resolution.[99] Protein stability at high temperatures was substantiated by a Circular Dichroism (CD) melting point measurement. The melting point of the recombinant OnubPBP3 at pH 6.5 was determined to be 92.19 °C (Figure 3.3).

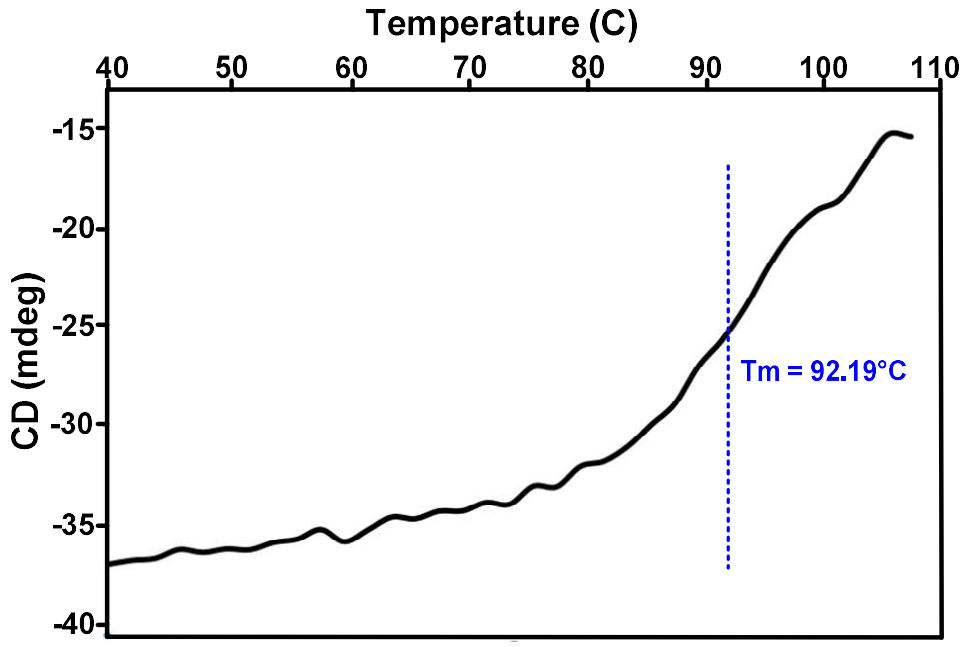

Figure 3.3: CD melting temperature curve of 30 µM OnubPBP3 in 15 mM phosphate buffer (pH 6.5) obtained at 209 nm.

Sequential NMR backbone assignments correspond to the chemical shifts from backbone atoms (^{1}H, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$). The backbone assignments were established using 3D HNCACB and 3D HNCA spectra. The 3D HNCACB spectrum provided intraresidual (residues i) and sequential cross-peaks of Cβ and Cα (residue i-1) (Figure 3.4), whereas the 3D HNCA gave information only about Cα (Figure 3.5). The assignments of Cα, Cβ were confirmed by using 3D HN(CO)CA and 3D CBCACONH spectra since they provide only sequential cross-peaks from residues i-1. 3D HNCO and 3D HN(CA)CO are complementary pair experiments. 3D HNCO was used to assign

the chemical shift of the carbonyl carbons of residue i, while 3D HN(CA)CO was employed for both residues i and i-1. The sequential cross-peaks provided by the assignments of 3D HNCO and 3D HN(CA)CO (Figure 3.6) also confirmed the sequential assignments obtained from the other experiments. The Biological Magnetic Resonance Data Bank, BMRB (http://www.bmrb.wisc.edu/) provides an updated table of statistics calculated from the full BMRB database for all chemical shifts from atoms in the 20 common amino acids. The table was utilized as a reference to confirm the type of the assigned residue or eliminate some ambiguities. Amino acid residues, such as Ala, Gly, Ser, and Thr have unique chemical shifts and are recommended as a starting point.[24] The BMRB statistics table shows that the dispersion of Cα chemical shifts is small, thus it is usually hard to align residues on the sequence only by relying on the chemical shifts of Cα. Because the dispersion of the Cβ chemical shifts is higher, both Cα and Cβ chemical shifts make it easier to connect neighboring residues.

The sequential assignments were started from a Gly which not only has a distinct Cα chemical shift but also does not have a Cβ, in contrast to the other residues. Sequentially residues were connected until proline or another residue of an unambiguous type was encountered. Another stretch of amino acids containing a residue with characteristic chemical shifts was employed to continue the sequential assignments. This process was successfully finished by assigning 97% of the backbone. The assignment of backbone resonances (^{1}H, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$) of OnubPBP3 was completed for all residues in the 2D [^{1}H,^{15}N]-HSQC except for Ser1, Gln2, Lue132, and Leu137 (Figure 3.7).

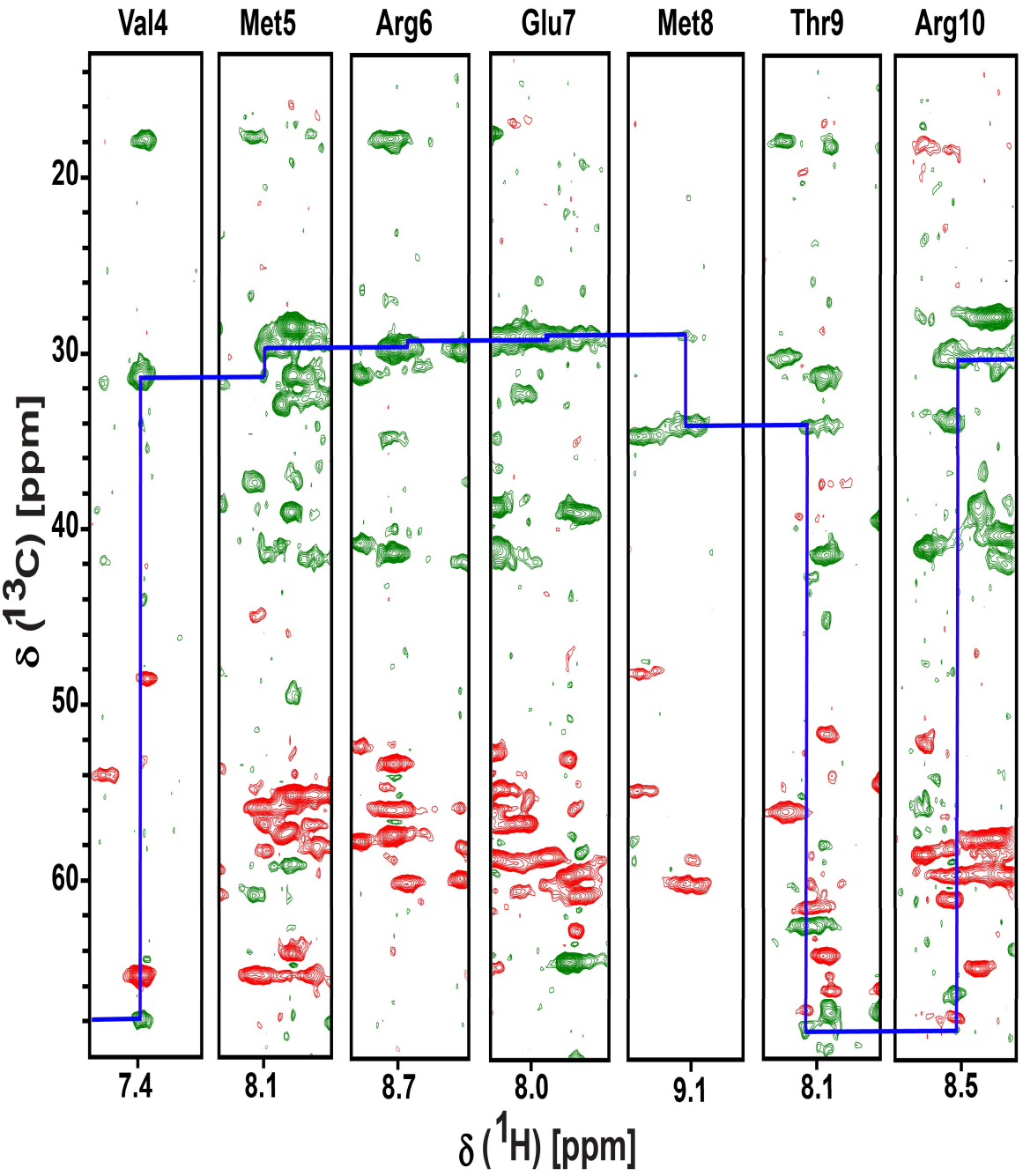

Figure 3.4: Sequential assignment walk for OnubPBP3 by 3D HNCACB spectrum. Strips from the 3D HNCACB experiment are shown for residues Val4-Arg10. Green (Cβ) and red (Cα) represent negative and positive signal intensity, respectively. The corresponding Cβ resonances are connected by blue lines.

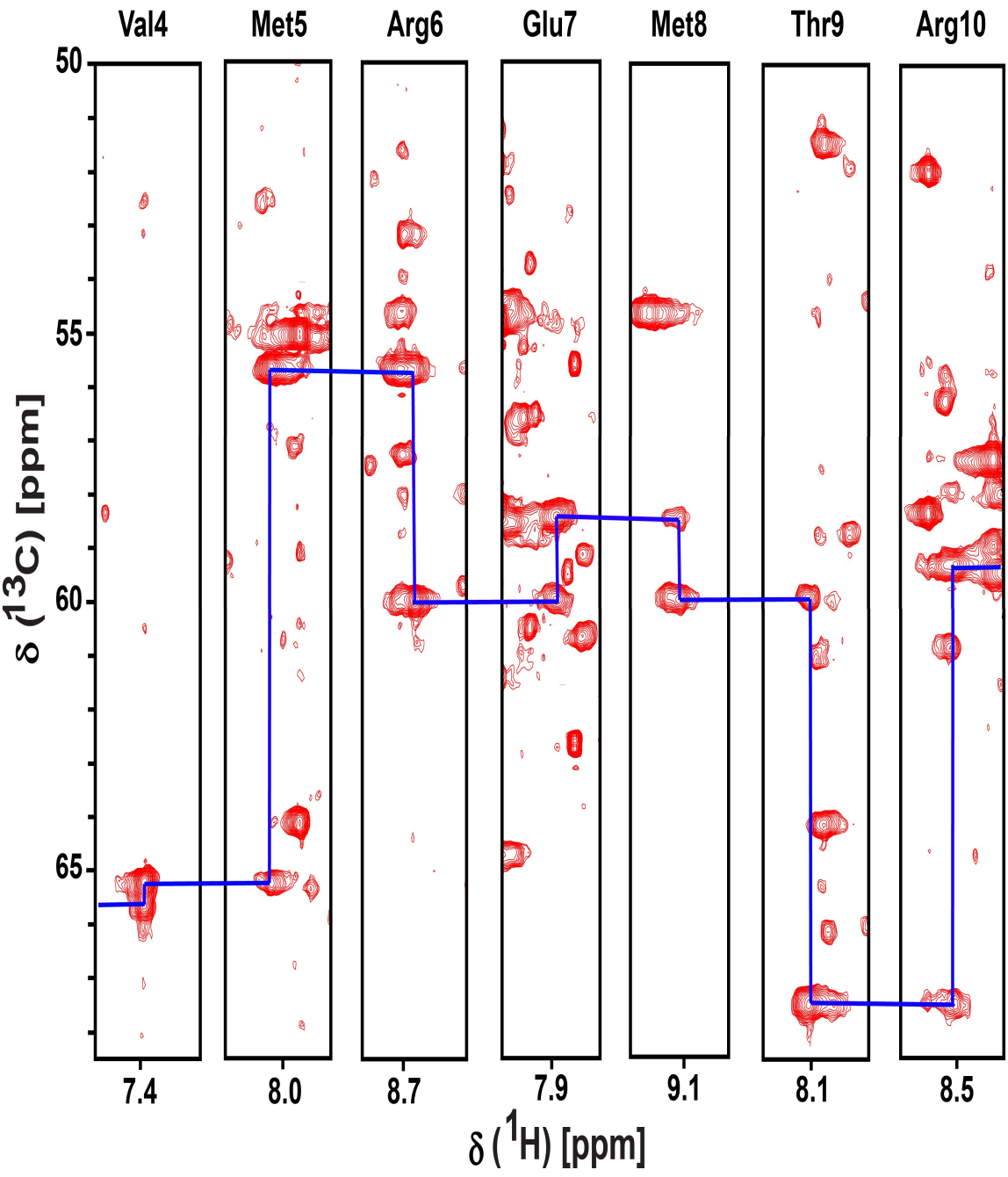

Figure 3.5: Sequential assignment walk for OnubPBP3 by 3D HNCA spectrum. Strips from the 3D HNCA experiment are shown for residues Val4-Arg10. The corresponding Cβ resonances are connected by blue lines.

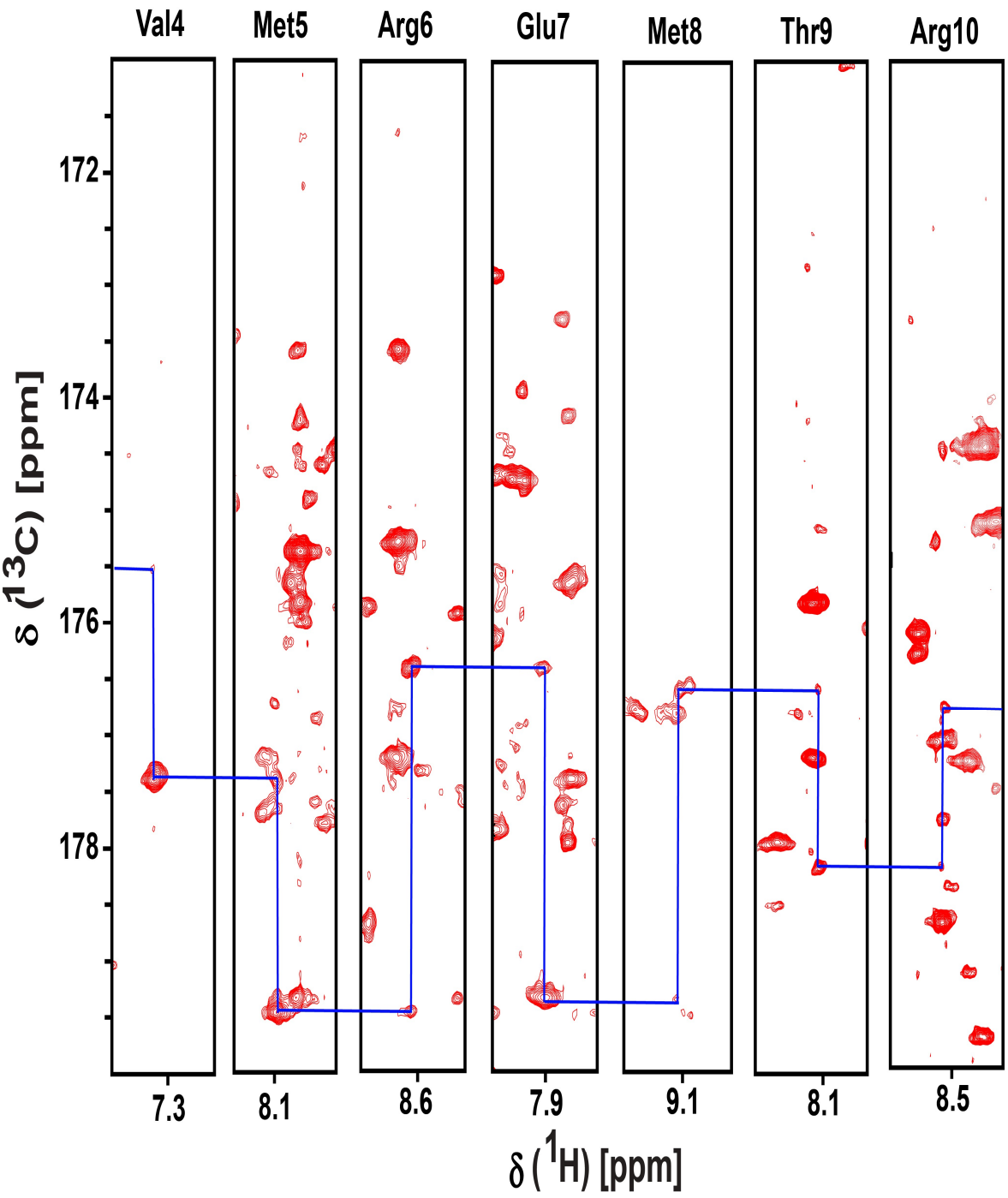

Figure 3.6: Sequential assignment walk for OnubPBP3 by 3D HN(CA)CO spectrum. Strips from the 3D HN(CA)CO experiment are shown for residues Val4-Arg10. The corresponding C' resonances are connected by blue lines.

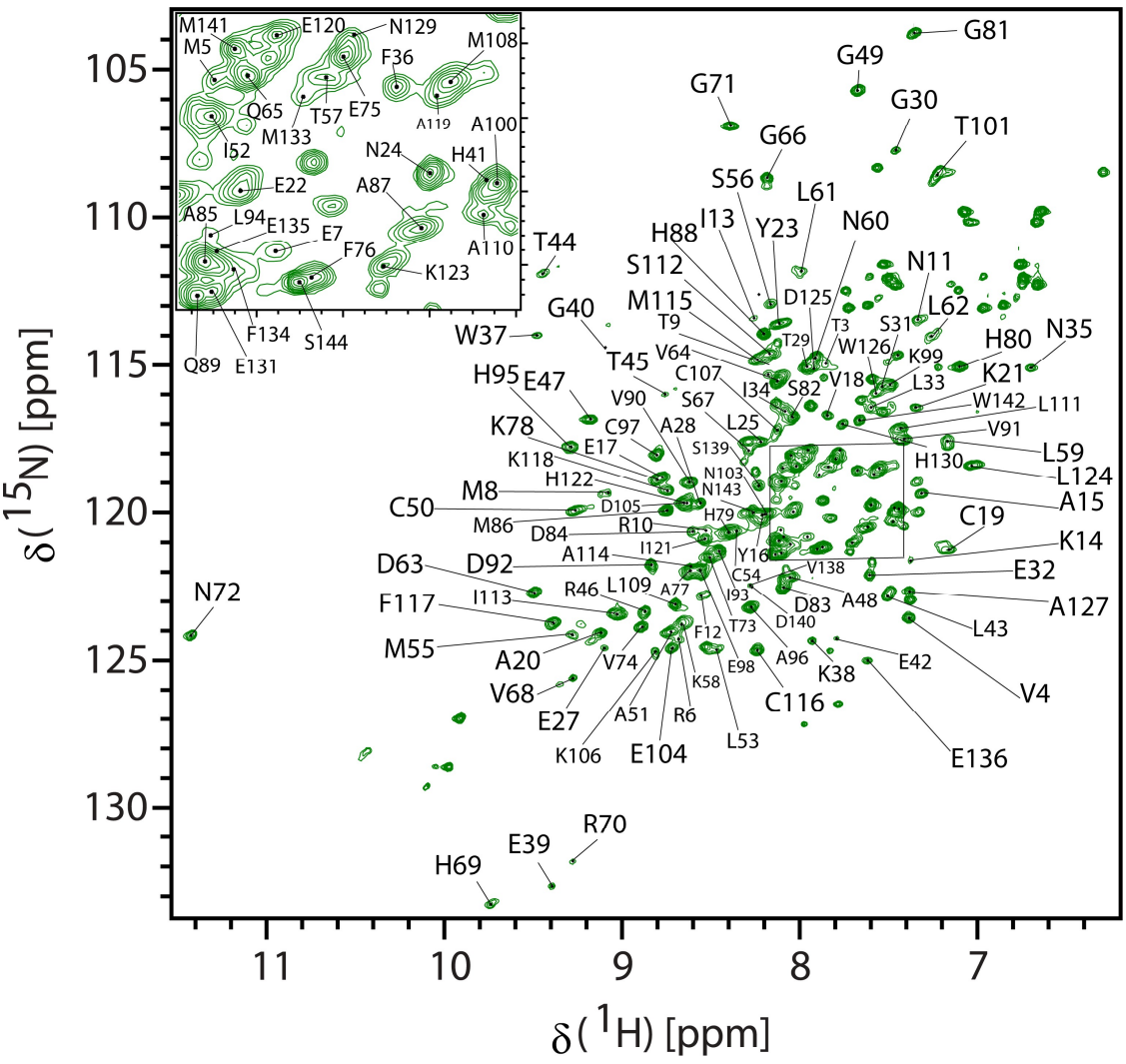

Figure 3.7: 2D [^{1}H, ^{15}N]-HSQC spectrum of uniformly ^{15}N,^{13}C-enriched OnubPBP3 at pH 6.5 and temperature 35 °C. The protein sample contains 50 mM phosphate buffer at pH 6.5, 95% H2O, 5% D_2O, 1 mM EDTA, and 0.01% (w/v) NaN3. The spectrum was recorded at a ^{1}H field-strength of 800 MHz. The structure of OnubPBP3 consists of 144 residues. Backbone amide cross peaks were labeled with the letter of residues type and the sequence number. The residues S1, Q2, L132, and L137 were not found in the spectra. Resonances of the side-chain glutamine and asparagine amide groups were not labeled. An extended region of the HSQC spectrum is shown in the sequare inset.

It has been reported that disulfide bonds in lepidopteran PBPs are extremely important for protein stability and the formation of the hydrophobic binding pocket which plays a pivotal role in PBP function. To detect the presence of the disulfide bonds in the proteins by NMR spectroscopy, the Cβ chemical shifts of cysteine residues are used as a probe. If the chemical shift of Cβ is greater than 35 ppm, the cysteine is oxidized. However, if the chemical shift is less than 32 ppm, the cysteine is reduced. Sometimes the Cβ chemical shifts can be within the range of 32-35 ppm. In this case, if the protein has an even number of cysteine and the Cβ chemical shifts of other cysteines are higher than 35 ppm, the cysteine of the in-between value is oxidized. The oxidation of two cysteine residues is an excellent sign for disulfide bond formation.[100] The Cβ chemical shifts of the six cysteine residues in the recombinant OnubPBP3 were higher than 35 ppm except one with a value close to 35 ppm indicating the protein contains three disulfide bonds (Table 3.4).

Table 3.4: The chemical shift of the six cysteine residues in OnubPBP3.

Cysteine residue	Cys19	Cys50	Cys54	Cys97	Cys107	Cys116
Cβ chemical shifts	41.1	34.7	40.2	39.2	44.9	39.2

The secondary structures of OnubPBP3 were predicted independently using chemical shifts with two different programs: Secondary Structure Propensity (SSP)[101] and TALOS+.[97] By using SSP, the deviations of the Cα and Cβ chemical shifts (Δδ) from mean random coil values were calculated. The difference between ΔδCα and ΔδCβ was determined and plotted against the OnubPBP3 sequence (Figure 3.8). The values of (ΔδCα – ΔδCβ) showed that OnubPBP3 contains seven α-helices. The C-terminus of the protein is random coil or unstructured, while the N-terminus is the first α-helix in the protein. These structural characteristics are similar to the structures of the well-studied lepidopteran PBPs (ApolPBP[46] and BmorPBP[51]) at pH 6.5. The helicity of OnubPBP3 was predicted to be 61.1% by using SSP. However, SSP showed 2.3% of the protein was β-strands and the rest was unstructured.

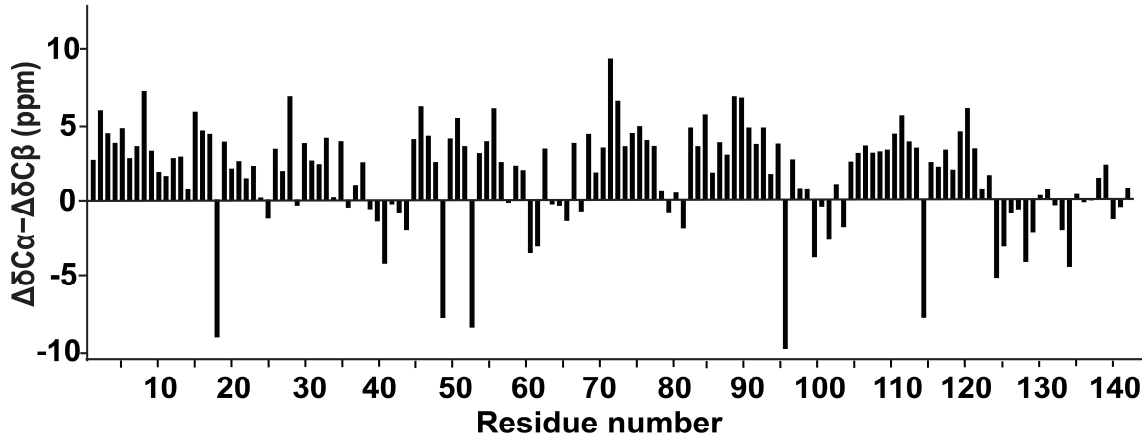

Figure 3.8: Secondary chemical shifts, ΔδCα − ΔδCβ, are plotted against the linear amino acid sequence. ΔδCα, ΔδCβ, and ΔδCα − ΔδCβ are calculated using the Secondary Structure Propensity (SSP) algorithm.

Furthermore, TALOS+ predicted the secondary structure of OnubPBP3 (Figure 3.9) by determining the φ and ψ torsion angles. The TALOS+ secondary structure is consistent with the one calculated by SSP.

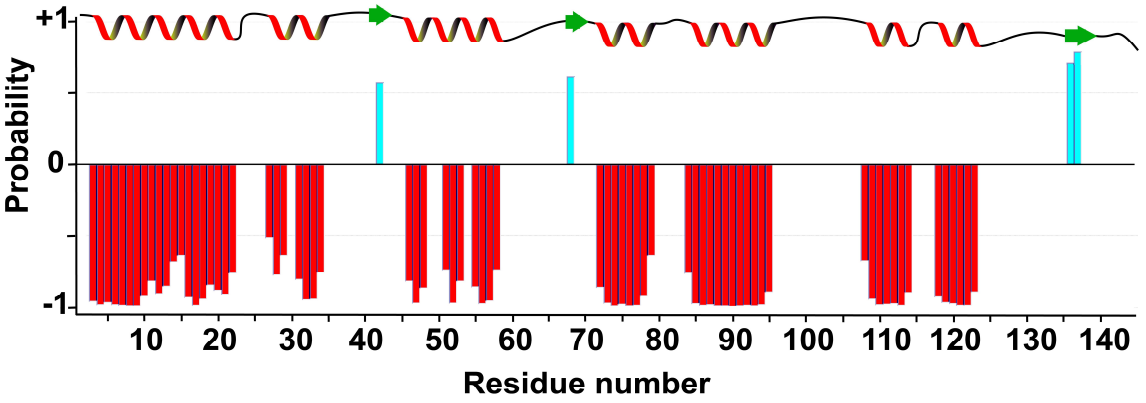

Figure 3.9: Secondary structure prediction of OnubPBP3 accomplished by TALOS+. The chemical shifts of ^{1}H, ^{15}N, $^{13}C\alpha$, $^{13}C\beta$, and $^{13}C'$ isotopes was as inputs. The probability of secondary structures is plotted against the linear amino acid sequence; the secondary structure prediction is shown as cyan bars for β-strands and red bars for α-helices.

In order to determine the three-dimensional structure of the protein, most of the side-chain proton and carbon resonances must be assigned, which always follows the backbone assignments. Although the completion of the backbone assignment facilitates the side-chain assignment, it remains a difficult task, especially when it is done with one experiment. Hence, two three-dimensional NMR experiments, 3D H(CCCO)NH and 3D ^{15}N-edited HSQC-TOCSY, were used specifically to assign the proton resonances of the protein side-chain. This assignment starts with choosing a backbone amide peak from [^{1}H,^{15}N]-HSQC to find the resonances of the side-chain protons of the same residue in the 3D ^{15}N-edited HSQC-TOCSY spectrum (Figure 3.10A) or the protons of the preceding residue in the 3D H(CCCO)NH spectrum (Figure 3.10B). Thus, it is a good practice to have both spectra side-by-side for assigning side-chain protons to reconfirm the assignments. The proton side-chain assignments of OnubPBP3 were accomplished and reconfirmed by these two spectra. Since the side-chain assignments reveal the residue type, it is important to compare this assignment with the results of the backbone assignments. For example, all glycine residues of OnubPBP3 gave two peaks in the 3D ^{15}N-edited HSQC-TOCSY and 3D H(CCCO)NH spectra for two alpha hydrogens, confirming our backbone assignment of the protein. The side-chain assignment was successfully done by assigning 89.2% of the protons. 3D ^{13}C-edited HSQC-NOESY was helpful in assigning some of the unresolved resonances in the 3D H(CCCO)NH and 3D ^{15}N-edited HSQC-TOCSY spectra. During the backbone assignment, only Cα and Cβ were assigned (Figure 3.11A). However, the assignment of the side-chain carbons is essential for the NOE assignments by 3D ^{13}C-edited HSQC-NOESY. Therefore, the 3D CC(CO)NH experiment was accomplished to obtain the resonances of all aliphatic side-chain carbons of residue i-1 (Figure 3.11B).

The final assignment that must be achieved before the structure calculation is the NOE assignment. It is the most critical stage among all resonance assignments and the main reason to go through the backbone and side-chain assignments. This assignment provides the distance between protons,

known as distance constraints. To accomplish this assignment, two NMR experiments, the 3D ^{13}C-edited HSQC-NOESY and the 3D ^{13}C-edited HSQC-NOESY, were performed on the doubled labeled OnubPBP3. In both experiments, the NOE is utilitized to exchange the magnetization between all hydrogens. In the 3D ^{15}N-edited NOESY-HSQC experiment, after NOESY the magnetization is transferred to neighboring ^{15}N nuclei and back to ^{1}H for detection, while the magnetization is transferred to neighboring ^{13}C nuclei and back to ^{1}H for detection in the 3D ^{13}C-edited NOESY-HSQC experiment. Consequently, each strip of the 3D ^{15}N-edited NOESY-HSQC spectrum in the NH plane contains NOEs from the NH group to all other hydrogens nearby. However, each strip of the 3D ^{13}C-edited NOESY-HSQC spectrum in the CH plane includes NOEs from the CH group to adjacent hydrogens (Figure 3.12). The number of carbons is much higher than nitrogens in proteins. Thus, the 3D ^{13}C-edited NOESY-HSQC spectrum was basically more informative than the 3D ^{15}N-edited NOESY-HSQC spectrum. To determine the global fold of the protein, a sufficient number of long-range NOEs are required. Unfortunately, long-range NOEs are typically weak and can easily be eclipsed by the intra-residue, short- and medium-range NOEs. The computational calculation of the protein structure, which is an iterative process, assisted in correcting NOESY cross peak assignments, and resolved ambiguities.

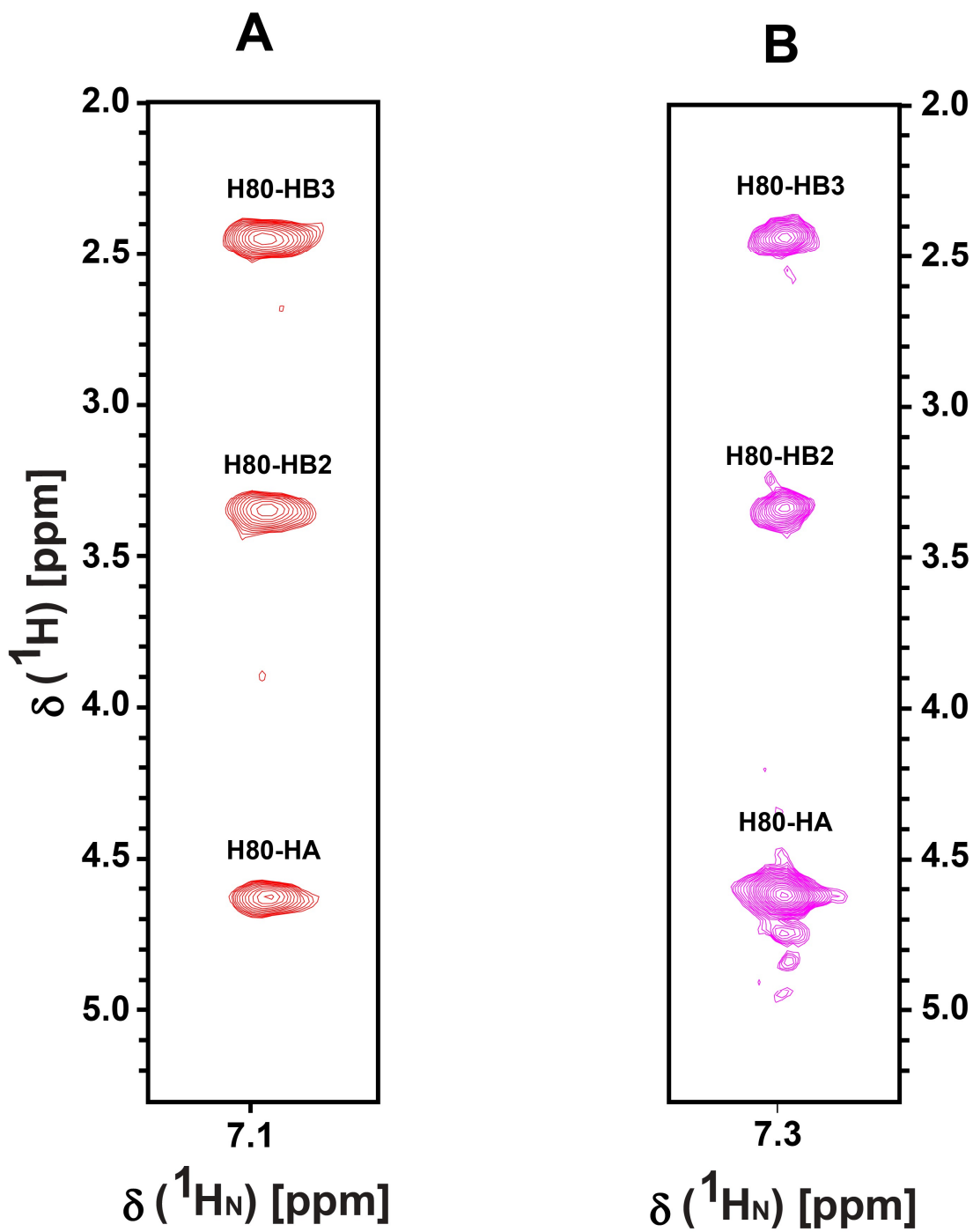

Figure 3.10: Proton side-chain assignment for OnubPBP3. (**A**) A strip from the 3D ^{15}N-edited HSQC-TOCSY experiment is shown for His80 as the i residue. (**B**) A strip from 3D H(CCCO)NH experiment is shown for His80 as the i-1 residue. Both spectra confirm each other by exhibiting the same resonances.

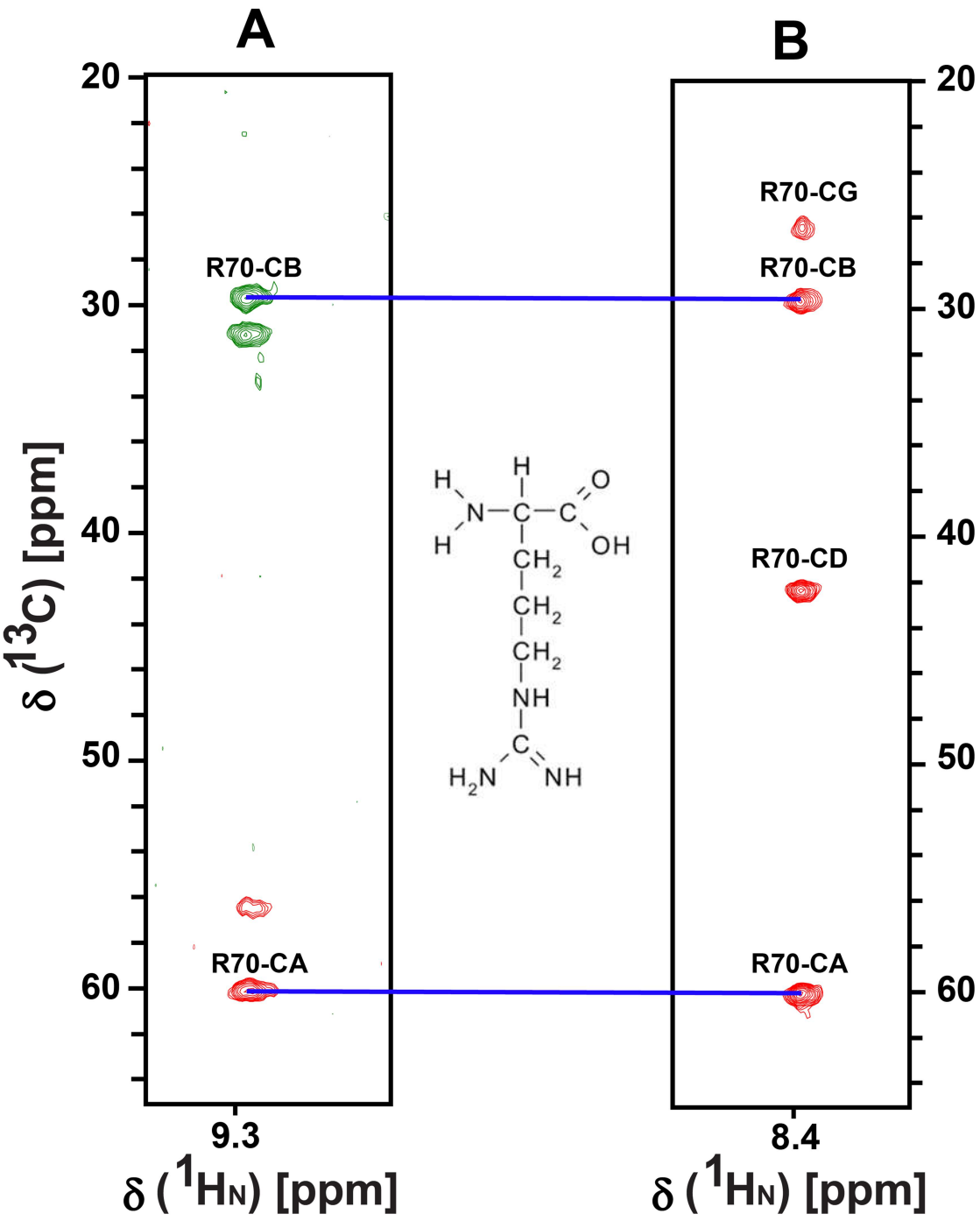

Figure 3.11: Carbon side-chain assignment for OnubPBP3. (**A**) A strip from the 3D HNCACB experiment is shown for Arg70 as the i residue; the only resonances that can be seen in this spectrum are of Cα and Cβ. (**B**) A strip from 3D CC(CO)NH experiment is shown for Arg70 as the i-1 residue; the spectrum provides all the resonances of the carbon side-chain in addition to the Cα. The chemical structure of the arginine amino acid is placed between the strips.

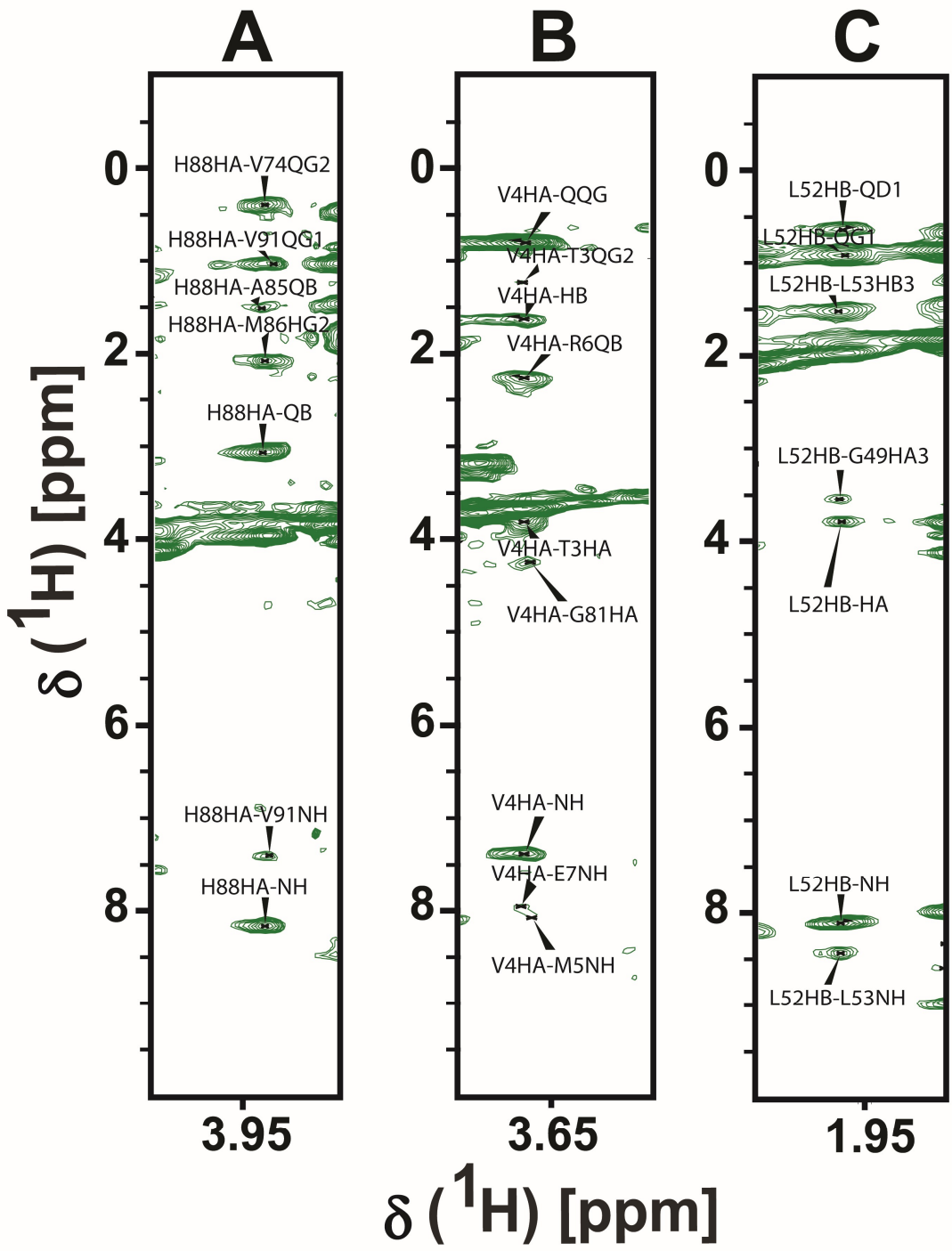

Figure 3.12: NOE assignment of OnubPBP3 by ^{13}C-edited HSQC-NOESY spectrum. (**A**) A strip from the spectrum is shown for H88 HA proton. (**B**) A strip from the spectrum is shown for V4 HA proton. (**C**) A strip from the spectrum is shown for L52 HB proton.

CHAPTER IV

STRUCTURE DETERMINATION OF ONUBPBP3 BY SOLUTION NMR

4.1 Introduction

4.1.1 Structure determination

NMR spectroscopy and X-ray crystallography are techniques that can be applied to the investigation of tertiary structures of proteins at the atomic level. However, NMR spectroscopy is the only approach that is utilized to determine detailed three-dimensional structures of proteins in the solution state. Nowadays, many of NMR experiments with proteins are carried out routinely to determine their three-dimensional structures.[46, 53, 55] Since Wüthrich and coworkers reported the first protein structure in 1985,[102] thousands of protein structures have been solved by NMR and deposited in the Protein Data Bank.[103] While NMR spectroscopy is unable to provide an image of a protein with atomic resolution instantly, it is capable of producing a wealth of indirect structural information, which helps in the determination of the protein tertiary structure after performing extensive calculations. The NMR structural informations are always provided as input to a computer algorithm as a distance and angular restraints for the calculation of protein tertiary structures. The distance restraints are the distances between each hydrogen pairs located within the distance range of 0-5 angstroms in a NOESY spectrum. As explained in Chapter 3, the NOE cross peaks assigned in a NOESY spectrum provide the distance restraint between pairs of protons within

a protein. The angular restraints are the backbone torsion angles (φ and ψ) of the protein which are obtained using the TALOS+ program as explained in Chapter 3.

Various computer programs such as CYANA[104] or XPLOR-NIH[28] are available to calculate the three-dimensional structure of a protein using information provided by NMR experiments. The programs require all the necessary restraints to be prepared in a particular format as input files to calculate an ensemble of tertiary structures. The computational calculation of a protein structure is an iterative process which helps to remove structural violations, correct NOESY cross peak assignments, and resolve ambiguities. The final result of this process is an ensemble of three-dimensional structures with a satisfactory root mean square deviation value (RMSD). Reported values of RMSD have been calculated differently from one protein to another. It is a usual practice to calculate the RMSD for the model ensemble by excluding the unstructured termini. However, sometimes the RMSD is calculated based only on the structured segments of the protein.

Three dimensional structures of a few PBPs have been determined by solution-state NMR. The NMR structure of BmorPBP has been determined under various pH conditions pH 6.5 and 4.5 and the C-terminal truncated protein (1-128) at pH 6.5.[50-51, 105] The PBP structure of the giant silk moth Antheraea polyphemus (ApolPBP) has been determined by NMR spectroscopy at pH 6.3, 5.2, 4.5.[46, 54-55] The PBP structure of navel orange worm, Amyelois transitella (AtraPBP1) has been also determined at pH 4.5 by NMR spectroscopy.[52] Recently, the NMR structure of the Gypsy Moth Lymantria dispar PBP1 (LdisPBP1) at pH 4.5 was calculated.[53] The RMSD of these proteins were calculated based on specific segments. Table 4.1 contains the RMSD values of the protein mentioned above and the proteins segments that are included in the calculation.

Table 4.1: RMSD values of NMR structure ensembles, segments included in the calculation, and sample conditions of different pheromone binding proteins

Protein name	Conditions	RMSD (Å)	Segments included in RMSD calculation
BmorPBP	**Ligand-Free, pH 6.5**	**0.95 ± 0.16**	**1-128**
BmorPBP	**Truncated protein (1-128), pH 6.5**	**0.47 ± 0.05**	**1-128**
BmorPBP	**pH 4.5**	**0.46 ± 0.07**	**8-142**
ApolPBP	**pH 6.3**	**0.85**	**1-125**
ApolPBP	**pH 5.2**	**0.56**	**10-125**
ApolPBP	**pH 4.5**	**0.46 ± 0.06**	**10-142**
AtraPBP1	**pH 4.5**	**0.68 ± 0.09**	**Structured segments**
LdisPBP1	**pH 4.5**	**0.28 ± 0.07**	**13-143**

4.1.2 Molecular Dynamic Simulation

Proteins are essential components of living cells, responsible for many indispensable biological processes including enzyme catalysis, molecular recognition, signal transduction, and protein localization; in addition, structural proteins maintain cell shape. These functions are directed by

protein motions and physical interactions formed between proteins and associated molecules such as ligands, peptides, proteins and nucleic acids.

The development in instrumentation (X-ray crystallography, nuclear magnetic resonance (NMR) and cryo-electron microscopy) has provided a new boost in structural biology studies, resulting an exponential increase in the number of three-dimensional structures of proteins that have been determined and deposited to Protein Data Bank (PDB).[106] Protein structures encourage the study of protein dynamics to understand the functions of a protein at a molecular level. Protein dynamins can be studied either experimentally or computationally by molecular dynamics (MD) simulation. The first publication of a molecular dynamic simulation of a biomacromolecule was in 1977 by Martin Karplus and coworkers.[107] They studied the dynamics of a small 58-residue (6.5 kD) water-soluble protein called bovine pancreatic trypsin inhibitor. The protein dynamics were simulated in vacuum and only for 9.2 ps. This field has developed along with computer evolution to a level that can now be employed to study the dynamics of large proteins both soluble[108] and membrane-bound[109] in water as well as protein complexes including ligand-protein,[110] protein-protein,[111] and protein-nucleic acid complexes.[112]

In the current chapter, the first three-dimensional structure of the ligand-binding conformation of OnubPBP3 at pH 6.5 determined by high-resolution solution-state NMR is presented. Pheromone-docking to the calculated NMR structure of OnubPBP3 and molecular dynamics (MD) simulations were performed to gain insight into the protein structure (free and bound) and dynamics, as well as its interaction with the pheromones.

4.2 Methods

4.2.1 Structure calculation

Structure calculation was performed by CYANA 3.98.[104] It is a program used to calculate the protein structure based on conformational restraints, interproton distance and torsion angle restraints, obtained from NMR experiments. The interproton distance restraints were obtained from the assignment of 3D ^{15}N NOESY-HSQC and 3D ^{13}C NOESY-HSQC spectra while the torsion angle restraints were determined by TALOS+.[97] Disulfide bond restraints were applied. The CYANA calculation included seven cycles of iterative NOE assignment and structure calculation. During the iterative NOE assignments 350 NOEs were corrected, 54 were removed. The computer program initially generated 100 structures or models. Out of this number, 20 structures of lowest potential energy were selected to represent the ensemble conformation of the OnubPBP3. These structures were energy-minimized in a water shell with the YASARA Energy Minimization Server.[113] The structures were validated based on Ramachandran statistics determined by PROCHECK.[114] The root-mean-square deviations from the averaged coordinates of the 20 structures ensemble were calculated using PSVS 1.5.[115] Visualization of the OnubPBP3 structure was performed with the program MOLMOL 1.0.7[116] and PyMOL. The CASTp 3.0 server[117] was used to measure the volume of the hydrophobic cavity of OnubPBP3.

4.2.2 Molecular docking and Molecular dynamics simulation

To create a ligand-bound state of OnubPBP3, AutoDock Vina in the UCSF Chimera 1.13.1rc[118] program was used to dock *E*- and *Z*-pheromone (*E*11-tetradecenyl acetate and *Z*11-tetradecenyl acetate) individually into the calculated NMR structure of OnubPBP3. The lowest energy conformations of the protein-pheromone complexes were selected as the initial structure for the molecular dynamics (MD) simulation study. The GROMACS 2018.1 software package[119] was used to perform 400 ns MD simulations for the free protein and its complexes with either the *E*- or *Z*-

pheromone. The Gromos54a7 force field[120] was chosen to generate a protein topology file by using pdb2gmx command. The topology and force-field parameters for both pheromones were generated using the Automated Topology Builder (ATB) server.[121] MD simulations were carried out in a periodic cubic box where the protein alone or the protein-ligand complex was placed at least 1.0 nm from the box edges. The system was solvated by adding SPC water and then neutralized by adding 9 Na^+ ions. The energies of the systems were minimized through the steepest descent approach. The modified Berendsen thermostat with a time constant of 1.0 ps was used to maintain the temperature of systems at 300 K. To start with a reasonable structure, regarding geometry and solvent molecule orientation, the solvent and ions around the protein must be equilibrated. Equilibration often involves two phases: NVT and NPT. NVT equilibration was performed to bring the systems to the temperature of interest for the simulation and to establish the appropriate orientation of the solvent molecules around the protein. NPT equilibration was conducted to bring the system to the proper density. The computing for this project was performed at the OSU High-Performance Computing Center at Oklahoma State University (OSU). The essential steps involved in the performance of the MD simulations are summarized in Figure 4.1. The LIGPLOT[122] program was used to show the pheromone interactions with the protein. VMD software[123] was utilized to show 3D structures of free and bound OnubPBP3. The secondary structure database (DSSP) program[124-125] enabled us to follow the evolution in the secondary structure of free and bound OnubPBP3 over the entire simulation time. The GROMACS package was utilized to calculate root mean square deviation (RMSD) values for the protein as well as the ligands, the root mean square fluctuation (RMSF) values for each residue, and the ΔRMSF. Furthermore, GROMACS was employed to determine the number of hydrogen bonds created between ligands and the protein.

The RMSD is used to measure global protein flexibility and stability. The RMSD of particular atoms in a molecule, protein or ligand, with respect to a reference structure, was calculated based on the following equation:

$$\mathbf{RMSD}\ (\mathbf{t}) = \sqrt{\frac{1}{N}\sum_{i}^{N}\left\|\boldsymbol{X_i(t)} - \boldsymbol{X_i^{ref}}\right\|^2} \qquad \textbf{(4.1)}$$

where X_i (t) is the position of an atom i at time t, X_i^{ref} is the initial position of atom i, and N is the number of atoms.

The RMSF is a measure of the flexibility of a residue. It determines the deviation between the position of atom i and a reference position in a molecule such as a protein. The RMSF was calculated using the equation below.

$$\mathbf{RMSF}\ (\mathbf{i}) = \sqrt{\frac{1}{T}\sum_{t_j=1}^{T}\left\|\boldsymbol{X_i(t_j)} - \boldsymbol{X_i^{ref}}\right\|^2} \qquad \textbf{(4.2)}$$

where T is the time and X_i^{ref} is the reference position of atom i.

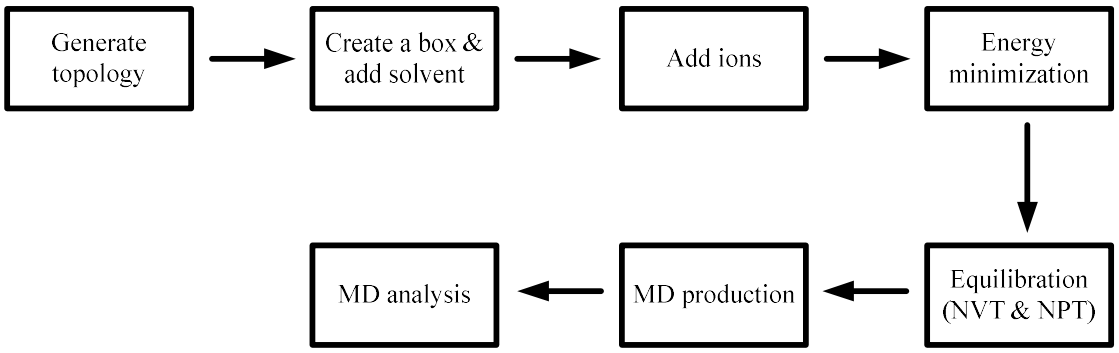

Figure 4.1: Block diagram representing the major steps involved in the performance of the molecular dynamics simulation

$\Delta RMSF_{C\text{-}F}$ and $\Delta RMSF_{C\text{-}Z}$ values were simply calculated from the resultant RMSF of free and bound OnubPBP3. The $\Delta RMSF_{C\text{-}F}$ values were obtained by subtracting the RMSF of the free protein from the bound protein. The $\Delta RMSF_{C\text{-}Z}$ values were calculated by subtracting the RMSF of bound protein to the Z-pheromone from the RMSF of the bound protein.

The binding energy of both pheromones with OnubPBP3 was determined by using the Molecular Mechanics-Poisson-Boltzmann Surface Area (MM-PBSA) method[126] to evaluate the relative stabilities of OnubPBP3 complexes. It is one of the methods used extensively to estimate the binding free energy. The g_mmpbsa tool utilizes some of files (tpr and xtc files) generated by GROMACS to calculate essentially three constituents of the binding energy, (1) molecular mechanical energy, (2) polar solvation energy, and (3) apolar solvation energy. The binding energys of OnubPBP3 to the pheromones were calculated based on 200 snapshots taken from the last 20 ns of the simulations.

Based on the MM-PBSA method, the binding free energy (ΔG_{blind}) can be represented by the following equations:

$$\mathbf{\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand})} \quad \textbf{(4.3)}$$

Where $G_{complex}$ represents the free energy of the protein-ligand complex, while $G_{protein}$ and G_{ligand} represent the free energies of the unbound protein and free ligand, respectively.

Moreover, the free energy for each individual component $G_{complex}$, $G_{protein}$ and G_{ligand} was estimated by:

$$\mathbf{G_x = E_{MM} + G_{solvation}} \quad \textbf{(4.4)}$$

where x is the ligand, protein, or protein-ligand complex, E_{MM} is the molecular mechanics energy in the vacuum state, and $G_{solvation}$ is the solvation free energy.

The vacuum molecular mechanics potential energy was calculated based on the molecular mechanics (MM) force-field parameters as in Eq. 4.5.

$$\mathbf{E_{MM} = E_{bonded} + E_{non\text{-}bonded} = E_{bonded} + (E_{vdw} + E_{elec})} \quad \textbf{(4.5)}$$

where E_{bonded} represents the bonded interactions which consist of bond, angle, dihedral and improper interactions, $E_{non\text{-}bonded}$ represents the nonbonded interactions which include both electrostatic (E_{elec}) and van der Waals (E_{vdW}) interactions.

The solvation free energy ($G_{solvation}$) is the amount of energy needed to transfer a solute from vacuum into a solvent. The solvation free energy was calculated based on the following equation:

$$\mathbf{G_{solvation} = G_{polar} + G_{non-polar}} \quad \textbf{(4.6)}$$

The two terms, G_{polar} and $G_{non\text{-}polar}$ represent the electrostatic and non-electrostatic contributions to the solvation free energy, respectively.

G_{polar} was computed based on the Poisson-Boltzmann (PB) equation,[127] and $G_{non\text{-}polar}$ was determined by using the solvent accessible surface area (SASA) in the following equation:

$$\mathbf{G_{non-polar} = \gamma\ SASA + b} \quad \textbf{(4.7)}$$

where γ is a coefficient related to the solvent surface tension and b is a fitting parameter.

4.3 Results and discussion

4.3.1 Structure calculation

The solution structure of OnubPBP3 at pH 6.5 is determined for the first time. The ensemble of 20 structures with the lowest energy of OnubPBP3 was generated with a backbone RMSD of 0.5 Å for the residues 3–124. This result is very good compared to the RMSD values calculated for the proteins listed in Table 3.1. A view of a superimposed 20-member structural ensemble of OnubPBP3 is represented in Figure 4.2. The determined OnubPBP3 structure was validated by Ramachandran plot. This method has been used for decades and it is a key quality metric based on values of phi and psi angles for protein structure validation.[128] The Ramachandran plot of OnubPBP3 shows 117 residues (88%) in the most favored regions and 16 residues (12%) in additional allowed regions indicating a good quality structure (Figure 4.3). Restraints and structural statistics for the 20-structure ensemble of OnubPBP3 are listed in Table 4.2. The mobility of the OnubPBP3 backbone was investigated by analyzing sequential NOE resonances (Figure 4.4). Some parts of the backbone showed increased mobility as detectable from a weak or missing NOE resonance leading to an increase in the local RMSD. The flexible segments of OnubPBP3 include part of the loop between α2 and α3 (residues 36-40), part of the loop between α3 and α4 (residues 61-69), part of the loop between α4 and α5 (residues 82-83), part of the loop between α5 and α6a (residues 101-108), and part of the unstructured C-terminus (residues 127-144).

Table 4.2: Restraints and structural statistics for OnubPBP3

Property	**Value**
All distance restraints	1961 (100.0%)
Intraresidue distance restraints, |i-j|=0	432 (22.0%)
Sequential distance restraints, |i-j|=1	590 (30.1%)
Short-range distance restraints, |i-j|<=1	1022 (52.1%)
Medium-range distance restraints, 1<|i-j|<5	548 (27.9%)
Long-range distance restraints, |i-j|>=5	391 (19.9%)
Dihedral angle restraints	238
Ramachandran plot statistics	**(%)**
Residues in most favored regions	88.0
Residues in additional allowed regions	12.0
Residues in generously allowed regions	0.0
Residues in disallowed regions	0.0
RMSD	**(Å)**
Backbone of the residues 3–124	0.5
All heavy atoms of the residues 3–124	1.0

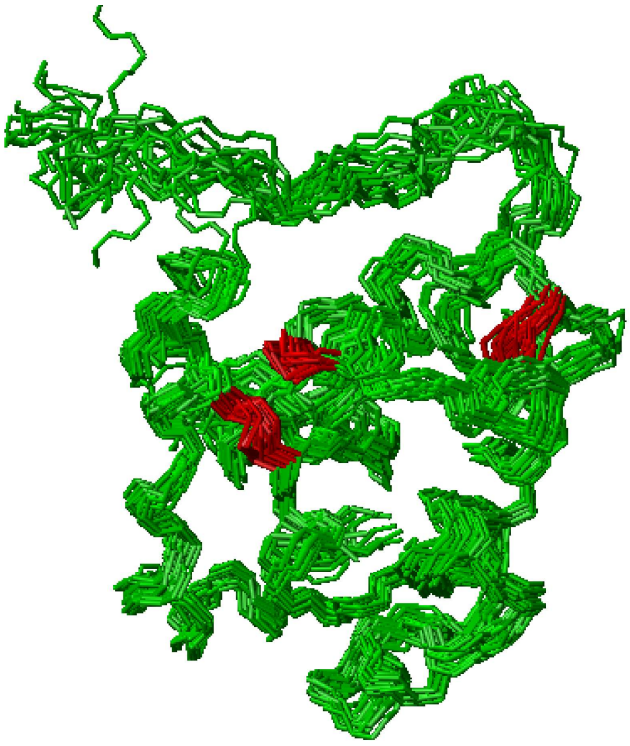

Figure 4.2: A view of superimposed 20-member structural ensemble of OnubPBP3. The protein backbone and disulfide bonds are presented in green and red, respectively.

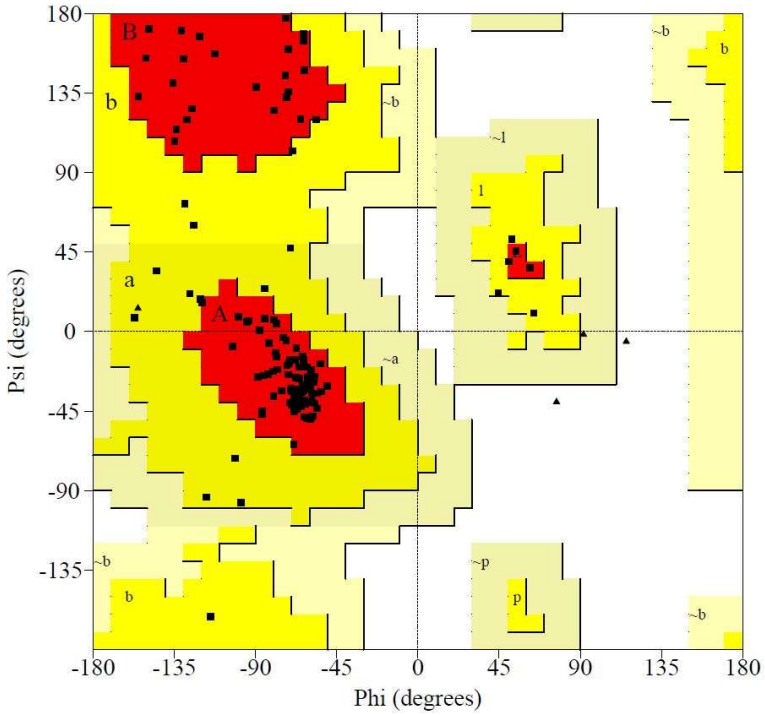

Figure 4.3: The Ramachandran plot of OnubPBP3.

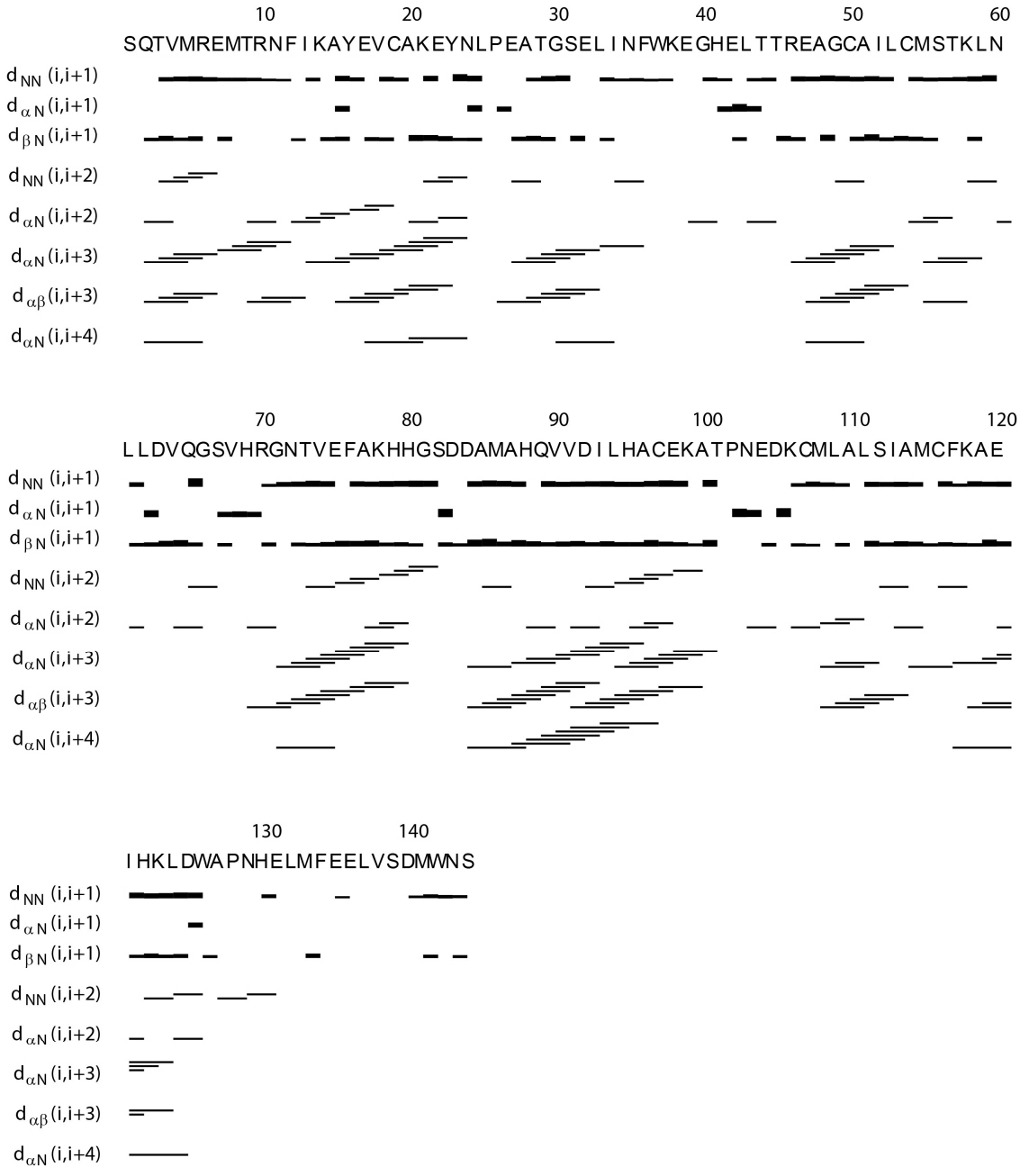

Figure 4.4: Summary of sequential NOE contacts of OnubPBP3

The structure of OnubPBP3 consists of eight α-helices with residues 3–9 (α1a), 15–23 (α1b), 27–35 (α2), 48–57 (α3), 72–79 (α4), 84–98 (α5), 114–116 (α6a), and 120–123 (α6b) (Figure 4.5). The most prominent characteristic of an α-helix is intrehelical hydrogen bonds formed between the backbone CO groups of (i) residue and the NH groups of the (i+4) residue. However, this is not the case with the first and last four residues of α-helices. It was determined that there is a preference for certain residues at both termini that can fulfill the hydrogen bonding requirements forming an N-cap or C-cap. N-cap is defined as the first residue of the helix that forms a hydrogen bond through the CO group with the NH group of residue i+4. On the other hand, the C-cap is described as the NH group of the last residue that forms a hydrogen bond to CO group of residue i-4. Both N- and C-capping are found to stabilize the protein.[129-130] The structure of OnubPBP3 contains two N-caps, one in α4 and the other in α5. The N-cap of α4 includes a hydrogen bond between Phe76 as donor

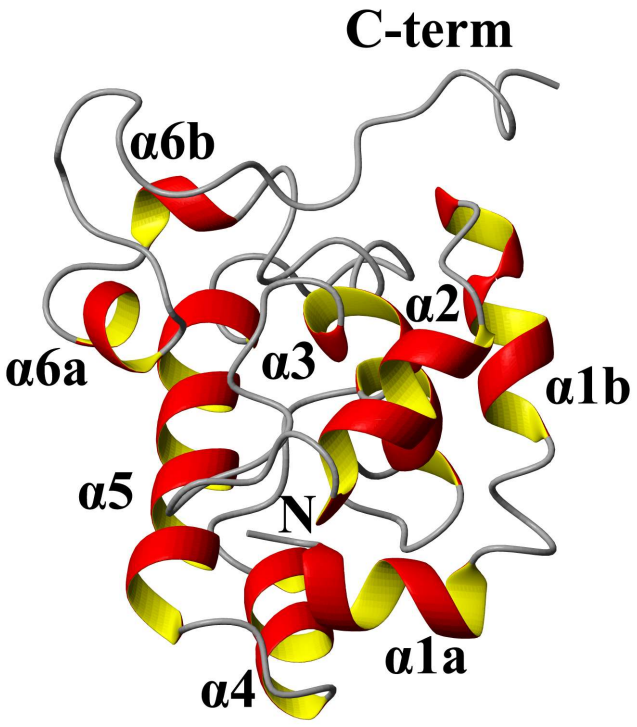

Figure 4.5: Ribbon structure of one of the OnubPBP3 conformers. Helices and both termini are designated.

and Asn72 as an acceptor. The N-cap of α5 is formed as a result of hydrogen bonding between His88 as a donor and Asp84 as an acceptor. A C-cap is also present in the OnubPBP3 structure precisely α1a, α2a, α4, and α5 where hydrogen bonds are formed between the following donor-acceptor pairs: Thr9-Met5, Tyr23-Ala20, and His79-Glu75, respectively. The six cysteine residues in OnubPBP3, as discussed in the Chapter 3, are oxidized by forming disulfide bonds with each other as follows: Cys19-Cys54, Cys50-Cys107, and Cys97-Cys116. These disulfide bonds help the protein to form a hydrophobic cavity by connecting the helices with each other. The bond between Cys19 and Cys54 connects the α1b to the α3 helix. The disulfide bridge of Cys50-Cys107 connects the α3 helix to a loop present between the α5 and α6a helix. The bond between Cys97 and Cys116 links helices α5 and α6a. A large number of hydrophobic interactions also exist between the side chains of the different helices. The eight α-helices are organized into a globular structure of residues 1–124 and an unstructured tail of residues 129–142 that extends into the solvent. Although the antiparallel helices α4 and α5 are not connected by a disulfide bond, they are packed nicely as a result of forming a salt bridge between Lys78 (α4) and Asp84 (α5), hydrophobic interactions, and a π-cation interaction between His 95(α5) and Arg70 (α4) (Figure 4.6). At acidic pH, it is expected that His95 becomes positively charged and repels the Arg70. Consequently, the packed antiparallel helices α4 and α5 may be disrupted. Here, we suggest that His95 and Arg70 may work together as a biological gate to maintain the pheromone in the pocket at pH 6.5 and facilitate the release of the pheromone at acidic pH.

Since the function of OnubPBP3 is to bind and transfer hydrophobic ligands (pheromones) in an aqueous environment to olfactory receptors, OnubPBP3 must have a hydrophobic pocket. Further analysis has revealed that the structure of OnubPBP3 contains a large cavity with volume of 273.979 $Å^3$ (Figure 4.7). This volume is similar to the cavity volume of the well-investigated lepidopteran PBPs at pH > 6, such as ApolPBP (282 $Å^3$)[46] and BmorPBP (272 $Å^3$).[51] The cavity of

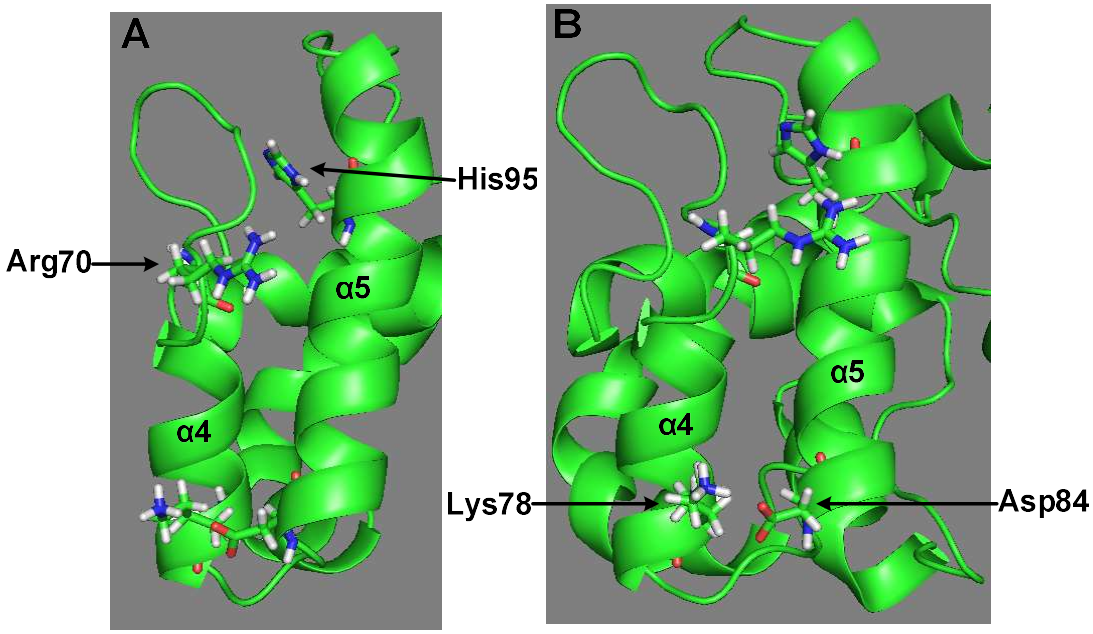

Figure 4.6: Close-up views of the helices 4 and 5 of onubPBP3 in the NMR structure. (**A**) Shows Arg70 close to His95 forming a π-cation interaction. (**B**) Shows Lys78 close to Asp84 forming a salt bridge.

OnubPBP3 is lined with side-chains of 29 residues: Gln2, Met5, and Thr9 from the helix α1a; Ile13 from the helix α1a; Glu42, Leu43, Thr44, and Arg46 from the loop between α2 and α3; Ala48, Gly49, Ile52, Leu53, Met55, and Ser56 from α3; Leu59, Leu61, Leu62, Gly66, and Val68 from the loop between α3 and α4; Thr73 from the helix α4; Val90, Ile93, Leu94, and Glu98 from the helix α5; Lys106, Leu109, and Ala110 from the loop between α5 and α6a; Phe117 from the loop between α6a and α6b; and Leu124 from the unstructured C-terminus. Most of these residues are hydrophobic which make the protein cavity suitable to bind the hydrophobic pheromones.

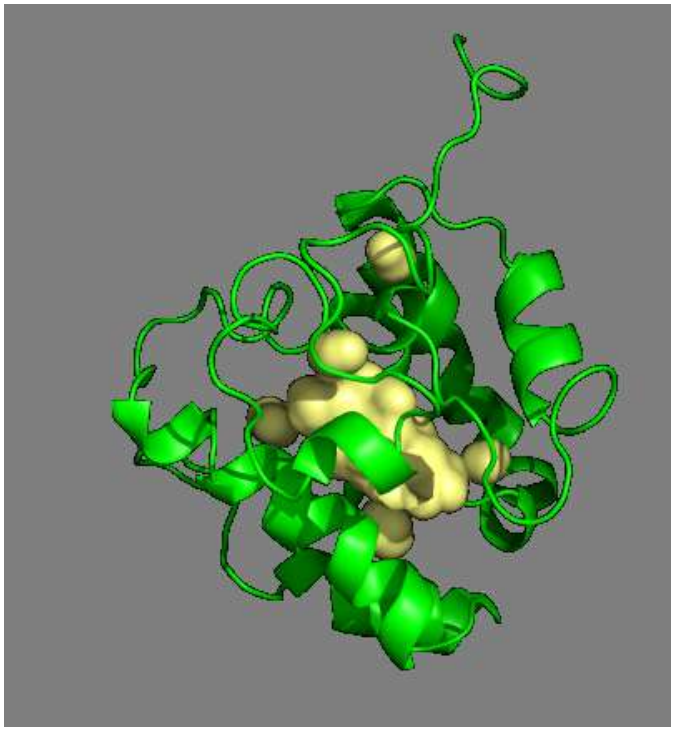

Figure 4.7: Binding cavity of OnubPBP3 viewed from the side of the helix α6a.

4.3.2 Molecular docking and Molecular Dynamic simulation

Molecular docking of the pheromone (*E*- or *Z*-isomer) to OnubPBP3 was conducted to predict the position and orientation of the pheromones when they are bound to the protein. By using AutoDock Vina, the optimal binding between the pheromone (*E*- or *Z*-isomer) and the protein was determined. Analysis of the docking results revealed all the amino acid residues that interact with the pheromones. The *E*-pheromone formed hydrophobic interactions with Gln2, Glu42, Leu43, Ala48, Ile52, Leu53, Leu61, Gln89, Val90, Ile93, Leu94, Ala110, Phe117, and Leu124 (Figure 4.7A). Similarly, the *Z*-pheromone showed hydrophobic interactions with Leu43, Ala48, Ile52, Leu53, Leu61, Gln89, Val90, Ile93, Leu94, Ala110, Phe117, and Leu124 (Figure 4.7B). However, it also showed hydrophobic interactions with Thr44 and Lys123 instead of Gln2 and Glu42 as in the

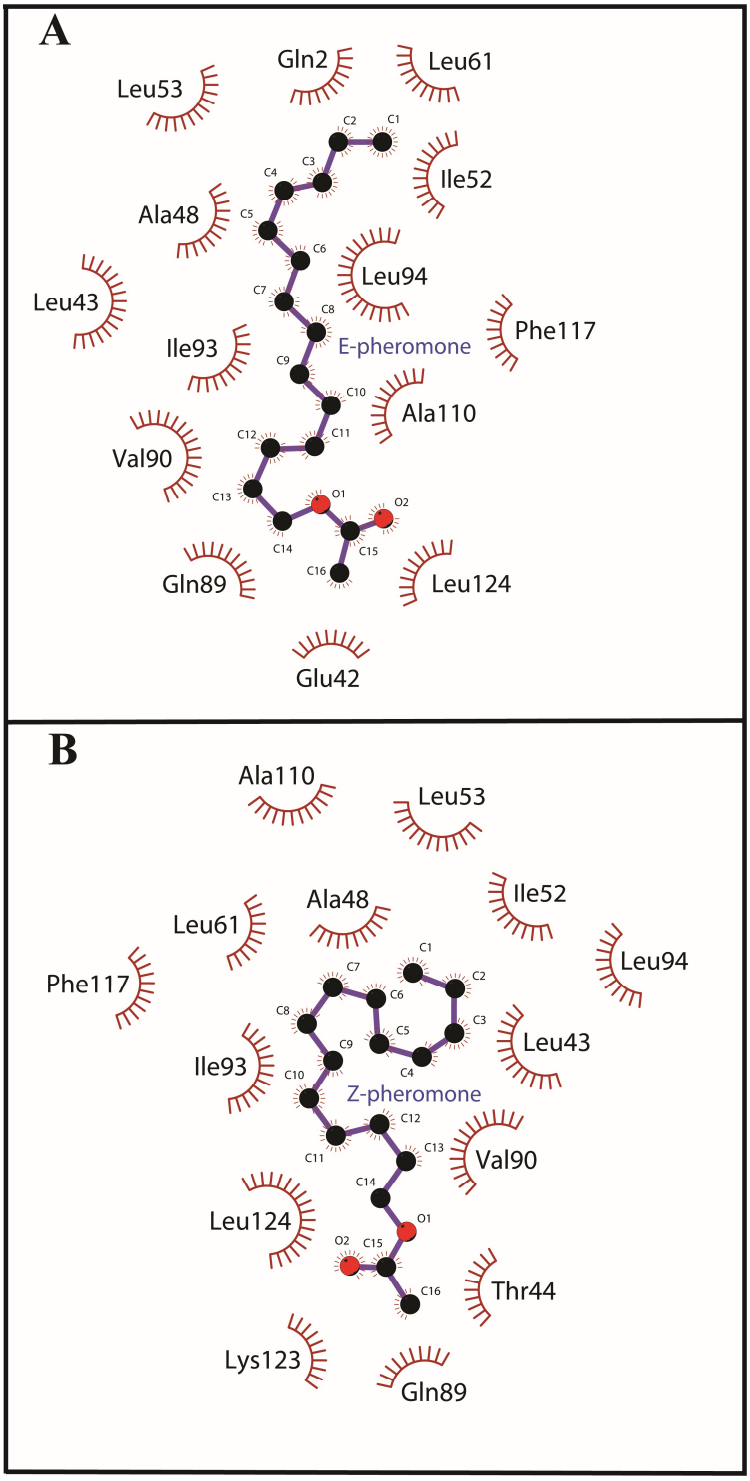

Figure 4.8: LIGPLOT diagram illustrating the interactions between pheromones and OnubPBP3 in each complex generated by molecular docking. Interactions of OnubPBP3 with (**A**) *E*-11-14:OAc and (**B**) *Z*-11-14:OAc where the hydrophobic interaction is displayed as arcs with spokes radiating towards the pheromone molecule. The atoms are color coded (carbon, black; oxygen, red; nitrogen, blue).

complex of the *E*-pheromone. Although docking programs are generally fast, one of the main methodological issues is obtaining accurate outcomes. Therefore, it is necessary to use another computational technique after molecular docking to obtain more reliable results. Molecular dynamics (MD) simulations have been used widely to optimize the provided outcomes by molecular docking.[131]

MD simulations were performed to get an insight into the binding of OnubPBP3 to the pheromones and to compare their binding interactions and protein dynamics. MD simulation of 400 ns was performed on each of the free OnubPBP3 and each of the docked OnubPBP3-pheromone complex structures. The root-mean-square deviation (RMSD) is used to measure global protein flexibility and stability. RMSDs of backbone atoms of free and bound OnubPBP3 were calculated. Each system (i.e. free protein and each complex) reached equilibrium at a different time in the simulation (Figure 4.8). The RMSD of OnubPBP3:*E*-pheromone complex converged at 80 ns with an average value of 0.54 nm. However, OnubPBP3:*Z*-pheromone complex reached equilibrium at 120 ns with an average RMSD value of 0.52 nm. Compared with the *E*-pheromone, the protein with the *Z*-pheromone showed some fluctuation at 80-340 ns indicating that OnubPBP3 is more stable with the *E*-pheromone. RMSD of free OnubPBP3 with an average value of 0.67 nm is higher than RMSD values of both complexes by more than 0.1 nm. This suggests that the free OnubPBP3 undergoes more changes in the global structure before reaching an equilibrium compared to the complexes. The free OnubPBP3 reached the maximum stability at 40 ns but had some fluctuation (relative to the OnubPBP3:*E*-pheromone complex) throughout the 400 ns simulation. The ligand-RMSD of both pheromones were also calculated to observe the stability of the ligand inside the hydrophobic pocket of OnubPBP3 (Figure 4.9). Interestingly, the average RMSD of *E*- and *Z*-pheromones were found to be 0.25 nm and 0.26 nm, respectively, and both of them were very stable along the entire simulation suggesting that the pheromones are stable inside the hydrophobic pocket.

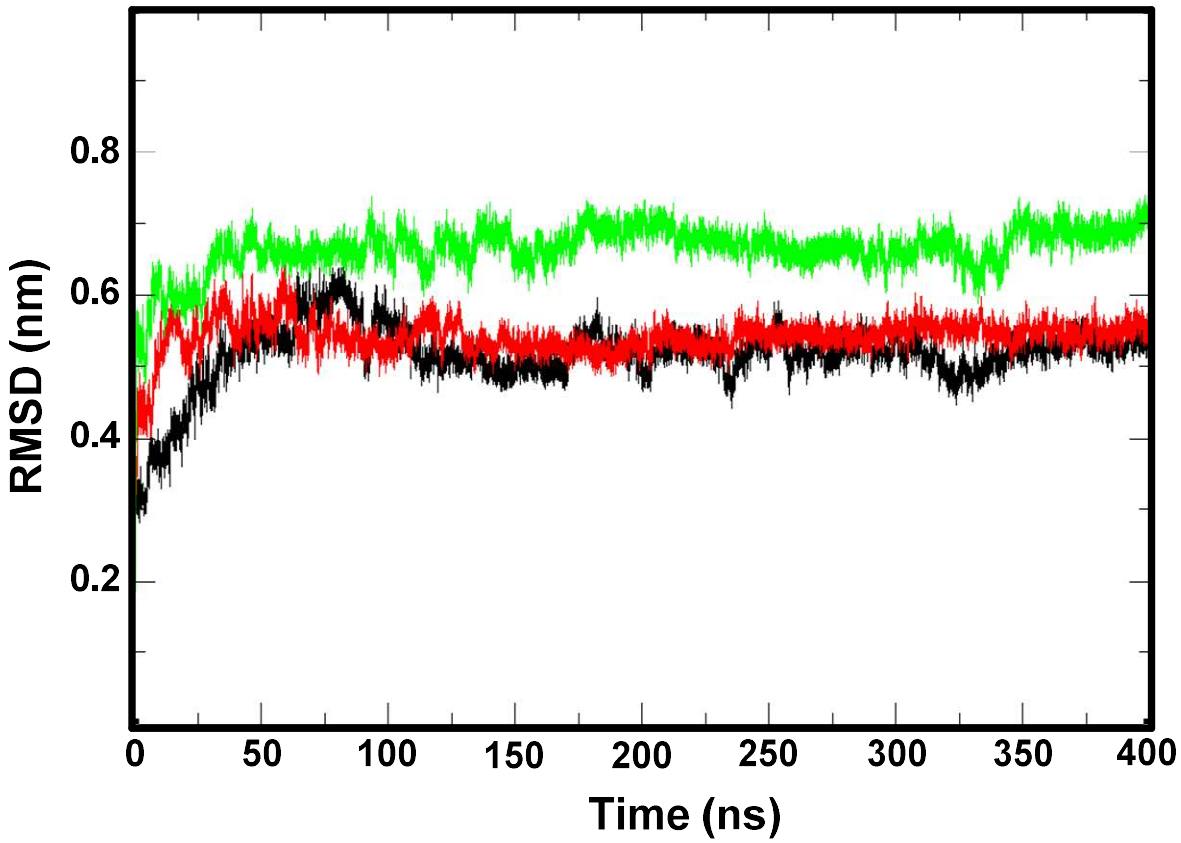

Figure 4.9: RMSD plots of OnubPBP3 in all three 400 ns MD simulations, where green represents free OnubPBP3, red color represents OnubPBP3 bound to the *E*-pheromone, and black represents OnubPBP3 bound to the *Z*-pheromone.

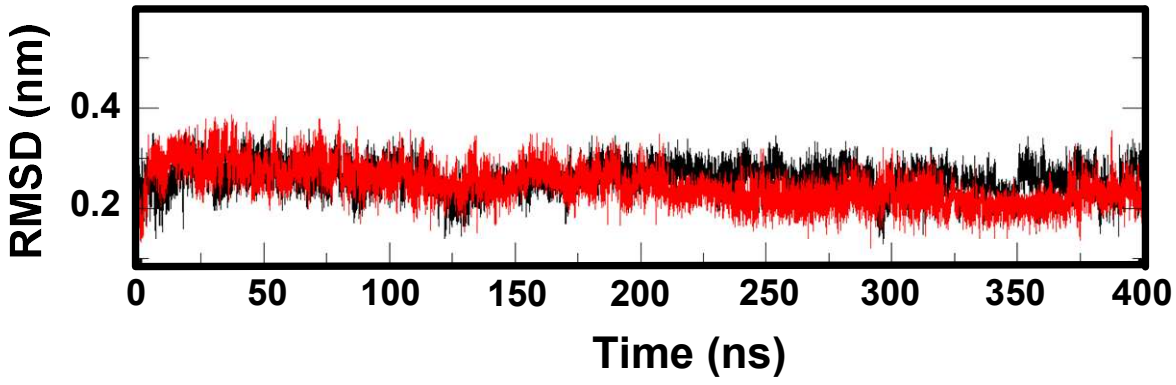

Figure 4.10: Ligand-RMSD plots of OnubPBP3 complexes in two 400 ns MD simulations, where red represents bound the *E*-pheromone to OnubPBP3 and black represents the bound *Z*-pheromone to OnubPBP3.

Although the docking results of both OnubPBP3 complexes did not show a hydrogen bond between the protein and the pheromones, the MD simulations did show the presence of the protein-pheromone hydrogen bonding. The *E*-pheromone formed hydrogen bonds with three residues (His41, Thr44, and Arg46), each one time during the MD simulation. The distance between the donor and the acceptor of these hydrogen bonds was around 0.3 nm. Arg46 formed, through the guanidinium group more than one hydrogen bond with the *E*-pheromone and they were the most stable compared to the ones created by the other residues (Figure 4.10). The *Z*-pheromone formed hydrogen bonds with five residues (His41, Thr44, Thr45, Arg46, and Ser139), each one time during the MD simulation. Similar to the OnubPBP3:*E*-pheromone complex, the distance between the donor and the acceptor of these hydrogen bonds was around 0.3 nm. Remarkably, Arg46 also created more than one hydrogen bond with the *Z*-pheromone and they were the most stable. The strong hydrogen bond must have a distance of $\leq$ 0.25 nm to be classified as mostly covalent. All hydrogen bonds formed between OnubPBP3 and the pheromones were classified as mostly electrostatic.[132] To get an insight into protein dynamics upon ligand binding, RMSF values of free and bound proteins were calculated. The RMSF is the root-mean-square fluctuation of each residue in a protein and is used to determine the residue-specific flexibility. The Cα-RMSFs for the three simulations were calculated over the 400 ns trajectories and averaged over each individual residue in the protein (Figure 4.11). Analysis of the free protein fluctuations revealed the loops, which are present in the regions Phe12-Ala15, Tyr22-Gly30, Trp37-Thr45, Leu59- Asn72, Gly81-Asp83, and Thr101-Met108 joining the α-helices and C-terminus, were the most flexible (Figure 4.11A).

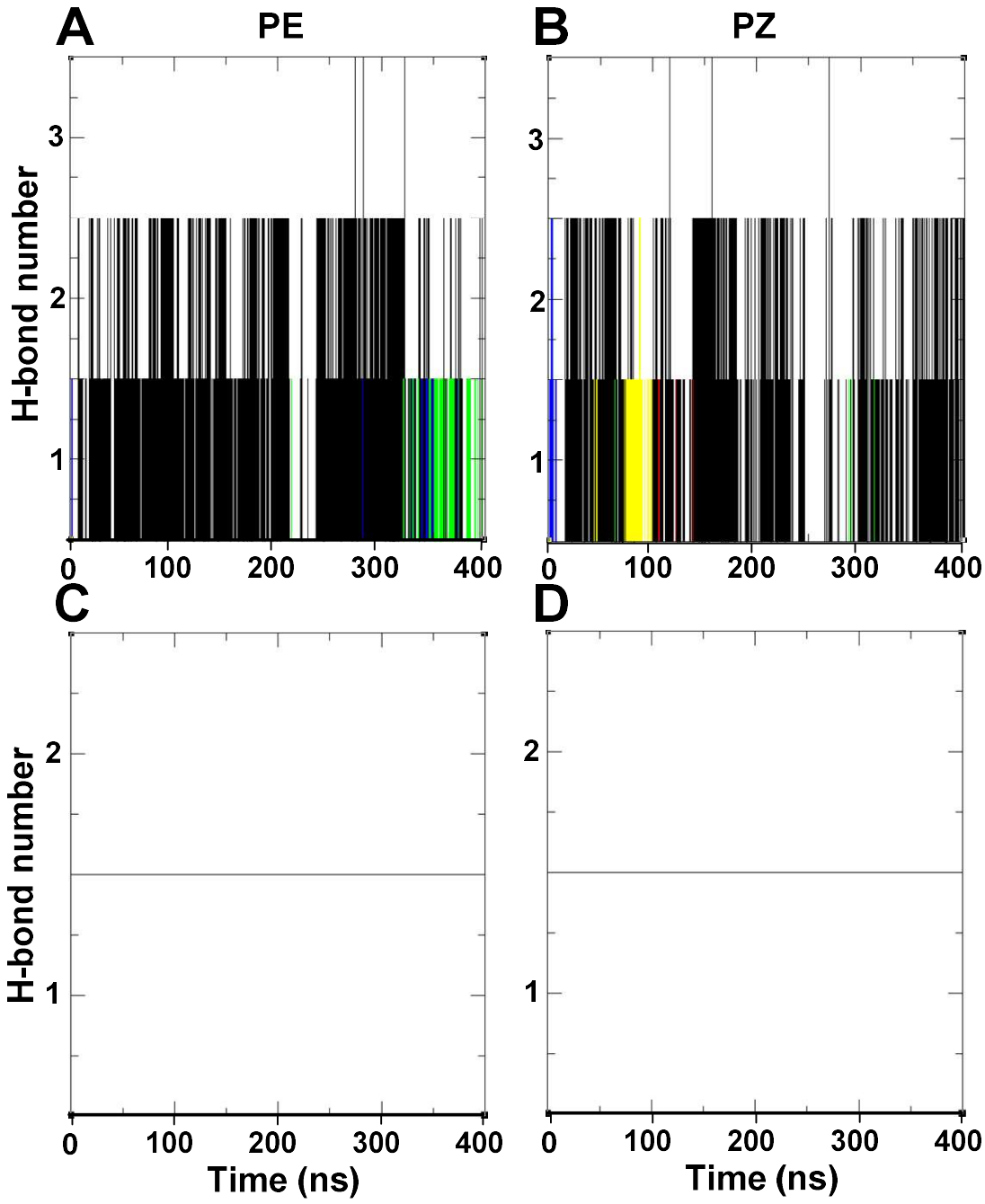

Figure 4.11: Time evolution of the number of intermolecular hydrogen bonds formed between the pheromones and active site residues of OnubPBP3, His41 (blue), Thr44 (green), Thr45 (red), Arg46 (black), and Ser139 (yellow). (**A**) and (**B**) hydrogen bonds between the protein and the *E* and *Z*-pheromones, respectively calculated between the donor and acceptor atoms with a maximum distance of 0.35 nm. (**C**) and (**D**) hydrogen bonds between the protein and the *E* and *Z*-pheromones, respectively calculated between the donor and acceptor atoms with a maximum distance of 0.25 nm.

By interacting the pheromones with the protein, the N-terminus became more flexible. Moreover, the binding to the *E*-pheromone increased the protein flexibility of different regions including loop1, loop2, loop3, Cys54-Asn60, and Ser112-His129. However, *E*-pheromone reduced the fluctuation of Leu33-Asn35, Gln65-Val74 (part of loop4), and Leu132-Ser139. The *Z*-pheromone increased the fluctuation prominently in the Lys21-Cys50 region and reduced the fluctuation of Gln65-Val74 (part of loop4) and the Leu132-Ser139 region. Both pheromones similarly decreased the flexibility of the Gln65-Val74 and Leu132-Ser139 regions and increased the flexibility of the N-terminus. Additionally, taking the difference of the root mean square fluctuation ($\Delta RMSF_{C-F}$) of each residue between the free protein and each complex provides a better picture of flexibility variations after adding pheromone to OnubPBP3. $\Delta RMSF_{C-F}$ showed clearly that the flexibility of OnubPBP3 was changed as a result of binding to the *E*- or *Z*-pheromone and binding to the *Z*-pheromone produced the highest flexibility (Figure 4.11B). Similarly, the difference of the RMSF for each residue between the complexes ($\Delta RMSF_{C-Z}$) showed the flexibility difference between the complexes (Figure 4.11C). $\Delta RMSF_{C-Z}$ exhibited that most of the residues of OnubPBP3:*E*-pheromone complex are less flexible compared to the complex of the protein with the *Z*-pheromone.

Since the effect of *E*- and *Z*-pheromones on the dynamics of OnubPBP3 are different, we monitored the secondary structural changes induced by the pheromones during MD simulations. The secondary structure of free OnubPBP3 exhibited fluctuation in the last α-helix (α6b) during the MD simulation in water making the region flexible (Figure 4.12A). The last α-helix of both complexes hardly existed raising the flexibility of the region. The first α-helix (α1a) of the free protein was stable during the 400 ns simulation, thus the region exhibited low flexibility. However, the first α-helix of the protein with the *E*-pheromone was unstable (Figure 4.12B). This helix was present constantly during the simulation until 140 ns; after that it barely existed in the rest of the simulation

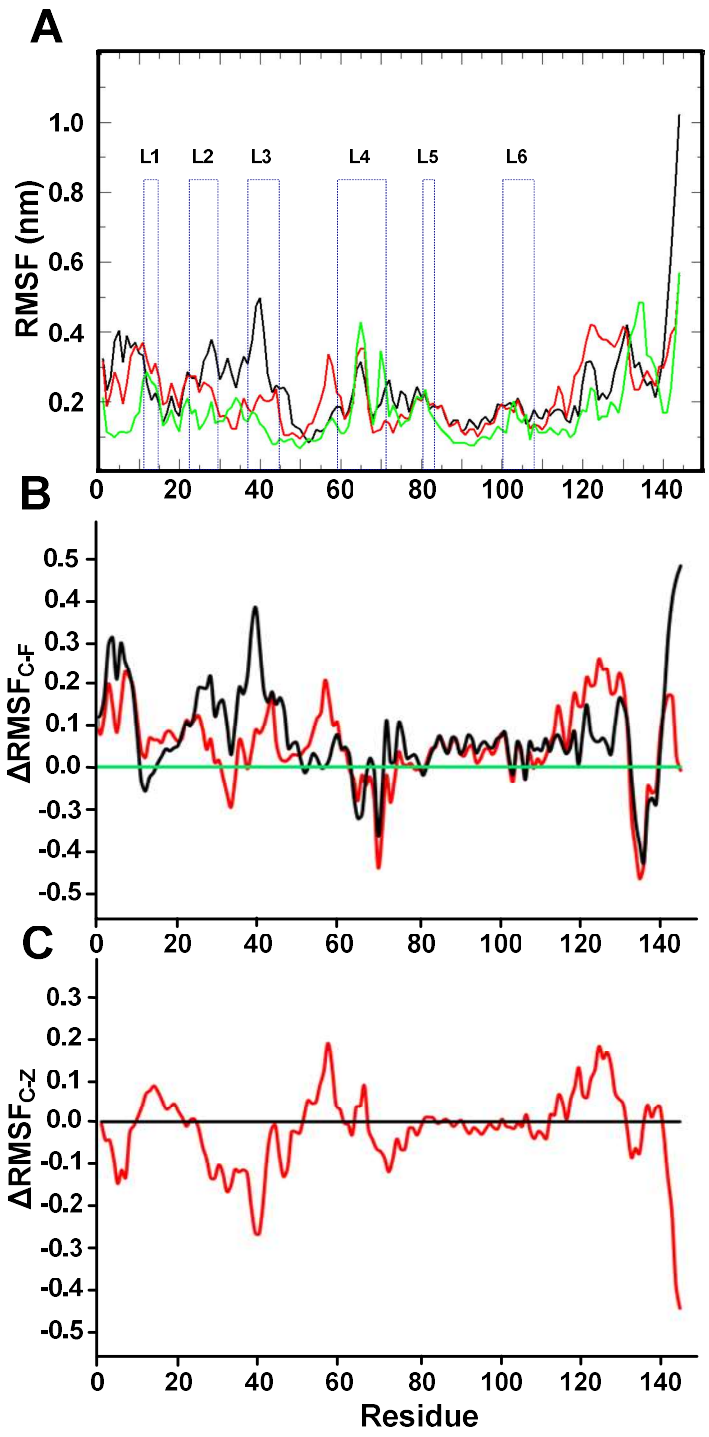

Figure 4.12: RMSF, and ΔRMSF plots of OnubPBP3 in all three 400 ns MD simulations. In all these plots, green represents free OnubPBP3, red color represents bound OnubPBP3 to *E*-pheromone, and black represents bound OnubPBP3 to *Z*-pheromone. (**A**) Cα-RMSF calculated for free and bound OnubPBP3; all loop areas in the plots were pointed via rectangles and labeled as in the NMR structure. (**B**) The differences between RMSF plots of free and bound OnubPBP3. (**C**) The differences between RMSF plots of the protein with Z-ligand and E-ligand.

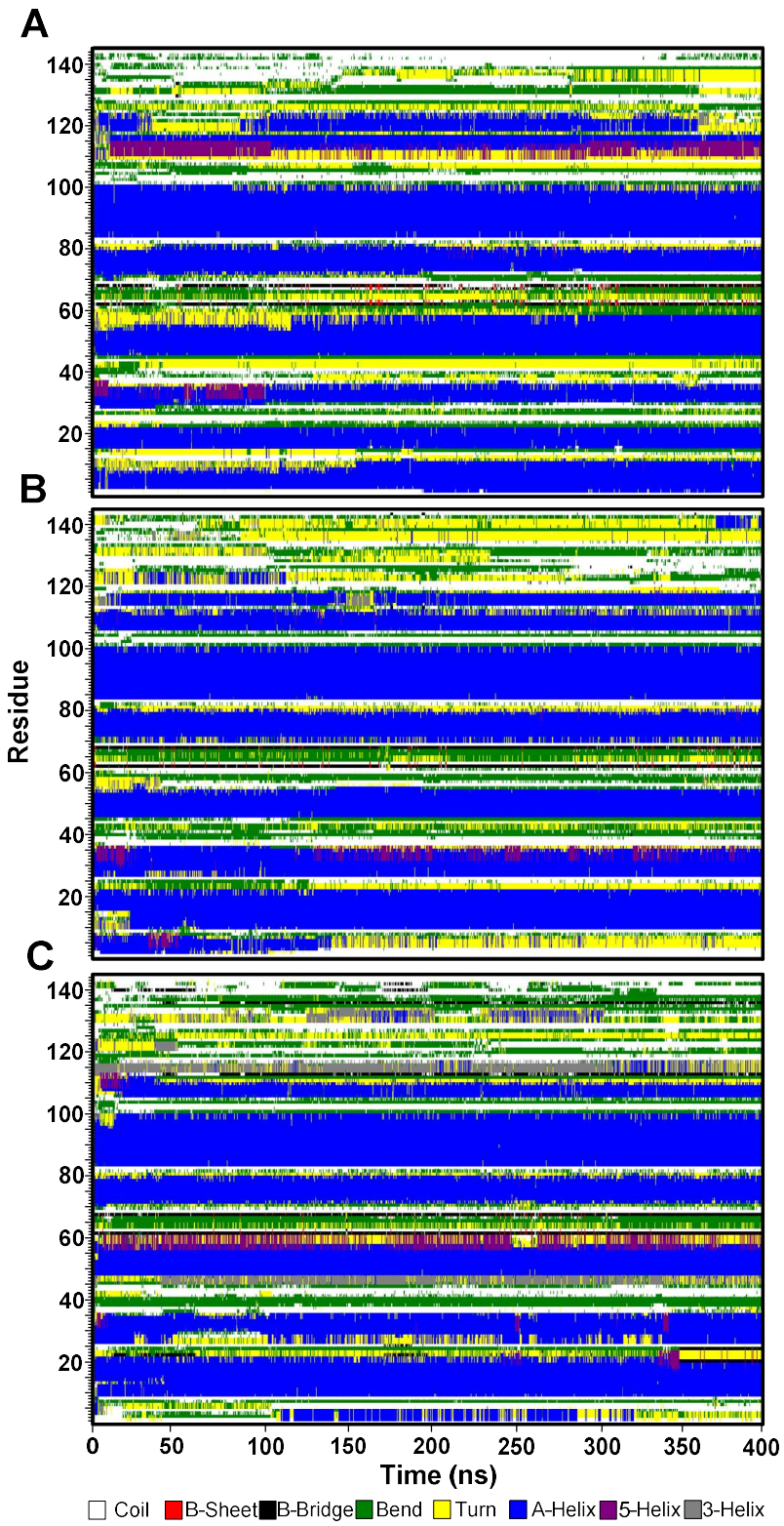

Figure 4.13: Secondary structures of OnubPBP3 in three 400 ns trajectories as a function of time calculated by DSSP, (**A**) the free protein, (**B**) the protein with the *E*-pheromone, (**C**) the protein with the *Z*-pheromone.

Causing a flexibility increase. The same α-helix, but with the *Z*-pheromone, was very unstable (Figure 4.12C); consequently, the RMSF plot showed the highest flexibility at the N-terminus compared to the other two systems. In the free protein and the protein:*Z*-pheromone complex, the Ser56-Lys58 segment was part of α3 during the simulation. This segment lost helicity due to the binding of OnubPBP3 to the *E*-pheromone producing more flexibility. Although some differences in the protein structure were observed in each complex, our data suggest that OnubPBP3 has broad specificity to hydrophobic ligands. A greater degree of flexibility in a protein is known to correlate to higher protein promiscuity.[133-135] It is possible that the flexibility of OnubPBP3 provides its ability to bind to different ligands.

MM-PBSA analysis was conducted on both protein:pheromone complexes to evaluate the affinity of the pheromones to the target protein. In this analysis, the binding energy, equivalent to experimental K_d value, was calculated for both complexes to estimate the stability of each complex. By using the MM-PBSA method, the binding energy was calculated for each complex after calculating four energy terms: van der Waals energy, electrostatics, polar solvation, and SASA energy. The values of all four energies as well as the binding energy for both OnubPBP3 complexes are listed in Table 4.3. The binding energy of OnubPBP3 with the *E*- and *Z*-pheromones were -139.7 (+/- 12.2) kJ/mol and -131.7 (+/-12.0) kJ/mol, respectively. A higher affinity ligand will stabilize a protein more than a ligand with a lower affinity. The K_d values determined by competitive displacement assay (Chapter 2) were 71 nM for the *E*-pheromone and 91 nM for the *Z*-pheromone. Taken together, the binding constants, the binding energies, and ΔRMSF indicate the *E*-pheromone forms a more stable complex with OnubPBP3 than the *Z*-pheromone.

Table 4.3: All energies calculated by MM-PBSA method for both OnubPBP3 complexes.

Energy	*E*-pheromone	*Z*-pheromone
van der Waal (kJ/mol)	-166.034 (+/- 9.125)	-164.995 (+/- 11.158)
Electrostatic (kJ/mol)	-9.334 (+/- 11.339)	-9.467 (+/- 9.252)
Polar solvation (kJ/mol)	54.351 (+/- 17.999)	60.440 (+/- 16.189)
SASA (kJ/mol)	-18.728 (+/- 1.205)	-17.706 (+/- 1.181)
Binding (kJ/mol)	-139.745 (+/- 12.228)	-131.728 (+/- 12.028)

CHAPTER V

pH-INDUCED CONFORMATIONAL CHANGE OF ONUBPBP3

5.1 Introduction

Proteins are composed of building blocks called amino acids which have been classified based on the chemical properties of their side chains into four groups: polar, nonpolar, positively charged, or negatively charged. The charged amino acids are aspartic acid, glutamic acid, arginine, lysine, and histidine which are responsible for the existence of charge on proteins (Figure 5.1). The inter- and intramolecular interactions of charged residue side chains depend on their protonation state.

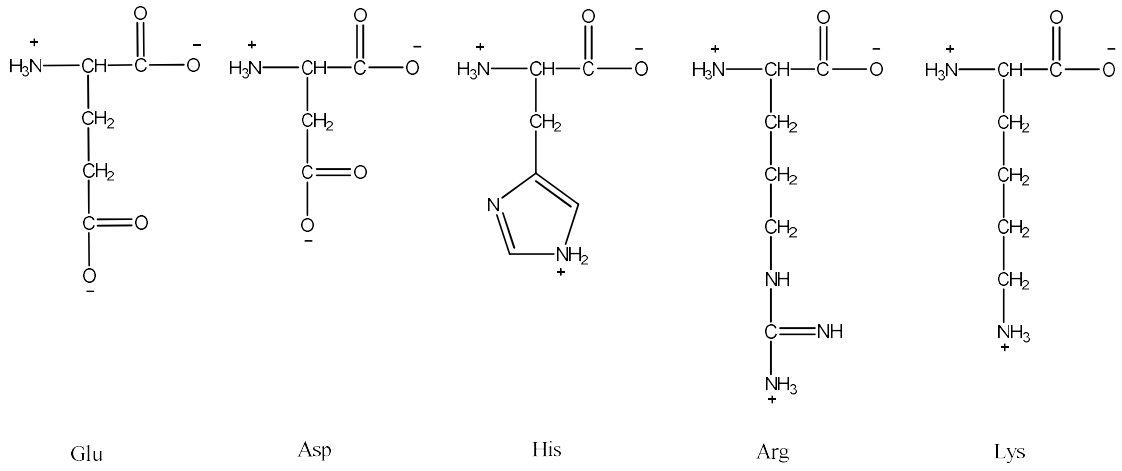

Figure 5.1: Structures of amino acids with charged side groups. Two of them can have negative charge and the others have positive charge depending on the pH of solution.

The charged amino acids have been observed to have essential roles in the binding of proteins to other molecules (micro- and macromolecules) and in enzyme mechanisms.[136-139] Furthermore,

protein structure, stability, and solubility are influenced by the ionization state of these amino acids.[140] The pH value of a solution is an important variable which can change the ionization state of the amino acid side chain. Titration techniques have been utilized to determine the dissociation constant (Ka) of each individual amino acid. The pKa values of the side chain of charged amino acids are shown in Table 5.1.

Table 5.1: The difference in pKa values of side chain amino acids in free amino acids and in a protein.

Amino acid	Aspartic acid	Arginine	Glutamic acid	Histidine	Lysine
Side chain pKa in amino acid	3.9	12.5	4.3	6.0	10.8
Expected side chain pKa in a protein	4.4-4.6	≥12	4.4-4.6	6.5-7.0	10.0-10.2

However, the titration of a protein or a peptide yields different pKa values of amino acid side chains because of the surrounding chemical environment. For instance, the side chain of glutamic acid possesses a carboxylic group with a pKa of 4.3 and its dissociation is represented by the following equation (Eq. 5.1).[141]

$$\mathbf{R-COOH + H_2O \leftrightarrow R-COO^- + H_3O^+}$$

$$\mathbf{Ka = \frac{[R-COO^-][H_3O^+]}{[R-COOH]}} \qquad \mathbf{(5.1)}$$

From Equation (5.1) we can imagine the effects of the environment on the state of the side chain. If the environment stabilizes (COO^-) relative to (COOH), the equilibrium will shift to the right side of the reaction and the pKa value will decrease and vice versa. Table 5.2 shows the effect of the chemical environment on the ionization state of the glutamic acid side chain and its pKa value.

Table 5.2: The ionization state of glutamic acid side chain and its pKa value change relative to the free amino acid for different locations of the amino acid in a protein.

The position of side chain	The effect on side chain	pKa value
On protein surface	The side chain is in touch with water and it is like the free amino acid in solution	Similar value
Positively charged region	The electrostatic interaction will stabilize the charge	Decrease
Negatively charged region	The electrostatic interaction will destabilize the charge	Increase
Nonpolar region	The hydrophobicity will destabilize the charge and makes it less hydrated than the free amino acid in solution	Increase
Buried inside a protein	The side chain will not be in touch with water and it is surrounded with non-aqueous environment	Unknown

The pH of the aqueous phase plays a critical role in the biological processes involving macromolecules and their assemblages. The influence of pH has been observed clearly in protein-ligand, protein-membrane, and protein-protein associations. Chemoreception is an essential process for the regulation of insect behaviors such as mating, feeding, oviposition, and predator avoidance. Pheromone binding proteins (PBPs) are the only components that serve inside the insect antenna, specifically to transport pheromones across the aqueous lymph to olfactory receptors. The structure and function of pheromone binding proteins of different insects have been studied in detail; such as Antheraea polyphemus (ApolPBP1),[61] Amyelois transitella (AtraPBP1),[56] Bombyx mori (BmorPBP),[105] Lymantria dispar (LdisPBP1 and LdisPBP2),[142] Apis mellifera (ASP1),[143] Blattella germanica (BgerOBP26 and BgerOBP40).[144] All of these proteins rely on a pH-dependent conformational switch mechanism to bind and release pheromones. The proteins bind to their pheromones at pH above 6.0 and release them at pH below 5.0. The sequence of OnubPBP3 presented in Chapter 1 contains 43 charged amino acid residues including 4 Arg, 9 Lys, 7 Asp, 15

Glu, and 8 His. Since the pH of the sensillum lymph in moths antenna has to be near neutral for pheromone binding and the pH at the vicinity of dendritic membrane is lower than 6.0,[145] the ionization state of Asp, Glu, and His could be modified by pH change (Table 5.1), it is interesting to investigate the pH effect on the conformation of OnubPBP3. From the pKa of the charged side chain (Table 5.1), His residues are expected to have the major effect on the protein conformation by reducing the pH from 6.5 to 4.5.

Some of the techniques used to study pH-dependent conformational change in proteins are circular dichroism (CD),[82] NMR spectroscopy,[61] and small angle X-ray scattering (SAXS).[146] CD spectroscopy is an absorption spectroscopy technique that measures the difference in absorbance of left- and right-handed circularly polarized light by chiral molecules in the sample. In CD, there are two regions in the spectrum that should be focused on, far UV and near UV. The far UV region is used to study the secondary structure of proteins. Absorption in this region occurs at < 240 nm and is due mainly to the peptide bond. The different types of regular secondary structure (α-helix and β-sheet) present in proteins provide characteristic CD spectra in the far UV. The secondary structure composition of proteins can be estimated from far UV CD spectra by using computer algorithms; such as CDSSTR, CONTINLL and SELCON3.[147] The near UV region is used to produce the CD fingerprint of the tertiary structure of proteins. Absorptions in this region take place at 260-320 nm and arise from the aromatic side chains of tryptophan, tyrosine, and phenylalanine. Each of these amino acids absorbs UV light at a specific range of wavelength: tryptophan 290-305 nm, tyrosine 275- 282 nm, and phenylalanine 255-270 nm. The factors that influence the magnitude and shape of the CD spectrum of a protein at this region are the content of each type of the aromatic amino acid, their mobility, their environment, and their disposition in the protein. Since CD spectroscopy in the near UV provides a valuable fingerprint of the protein tertiary structure, this region can be used to investigate the pH effect on the protein folding. It has been reported that collecting near UV CD spectra is a good method to observe the presence of the molten globule state

in proteins. The structure of a protein in the molten globule state is compact with a significant amount of secondary structures, but exhibits a loss in the tertiary structure, and 10-30 % increase in the radius of gyration.[148-151] Molten globule structures are characterized by very weak near UV CD signals compared to the native protein structure due to the high mobility of aromatic side chains.[151] Moreover, the thermal stability of a protein can be determined using CD to measure the melting temperature of the protein at specific conditions; such as pH, salt concentration, and denaturant concentration.[152]

NMR spectroscopy has been used to study the pH effect on the structure of proteins including different PBPs.[53, 61, 82, 153] Despite the popular use of conventional spectroscopies, such as fluorescence or CD to study the pH effect on proteins, these methods cannot determine which residues are responsible for the switch in the conformation of the protein. On the contrary, 2D heteronuclear NMR experiments, such as the heteronuclear sequential quantum correlation (HSQC) is used to address the contributions of individual residues.[53] Since the 2D [^{1}H,^{15}N]-HSQC experiment provides the fingerprint of a protein under particular conditions, spectra of the protein at different pH are collected. The chemical shifts of each amino acid residue are compared to recognize which part of the protein is affected by the pH change. However, this is not always the case because some proteins at acidic pH undergo melting globular state, denaturation, or aggregation which make it extremely difficult to keep track of the change in chemical shifts of the residues with the pH change.[154]

SAXS has been employed to study the effect of pH on structure of proteins.[146, 155-156] It is an important tool, not only for protein studies but also for structural biology in general.[157] It is a powerful, versatile, fast technique which requires a moderate amount of highly pure, monodisperse macromolecule that remains soluble at high concentration to provide detailed structural information. Unlike the high resolution methods of X-ray crystallography and NMR spectroscopy, SAXS requires non-crystalline and unlabeled samples of the macromolecule.[158] The speed of data

collection and sample characterization is a distinctive advantage of SAXS, especially with modern synchrotron sources where users may collect data within seconds.[159] Therefore, SAXS can be utilized as an approach for the fast screening of samples at different conditions. This technique is used to study the structure and interactions of proteins in solution under a variety of conditions (physiological or other) without the molecular weight limitations that may be encountered in other techniques such as NMR spectroscopy. Computational methods such as *ab initio* and rigid body modelings are utilized to build three-dimensional models with low-resolution based on SAXS data. Furthermore, SAXS provides quantitative structural information about flexibility, maximum size, and compactness of proteins.[160]

In the SAXS technique, X-ray photons are scattered as a result of an elastic collision between the X-rays and the sample. Inelastic collisions are not taken into account in this technique since they contribute to the background signal and degrade signal to noise. The intensities of the scattered X-rays are measured as a function of the scattering angle (2θ). Since the SAXS sample is in a solution, the scattering is isotropic because of the random orientation of the proteins. Thus, the scattering intensity recorded by a two-dimensional detector provides better statistical accuracy of the signal after radial averaging. Isotropic scattering intensity relies on the momentum transfer q ($q = 4\pi \sin(\theta)/\lambda$), where 2θ is the scattering angle, which is the angle between the incident and scattered beam, and λ is the wavelength of the X-ray beam (Figure 5.2). The final scattering profile used for analysis is a one-dimensional curve known as scattering intensity I (q) versus q, where q is given in inverse angstroms or inverse nanometers.[161]

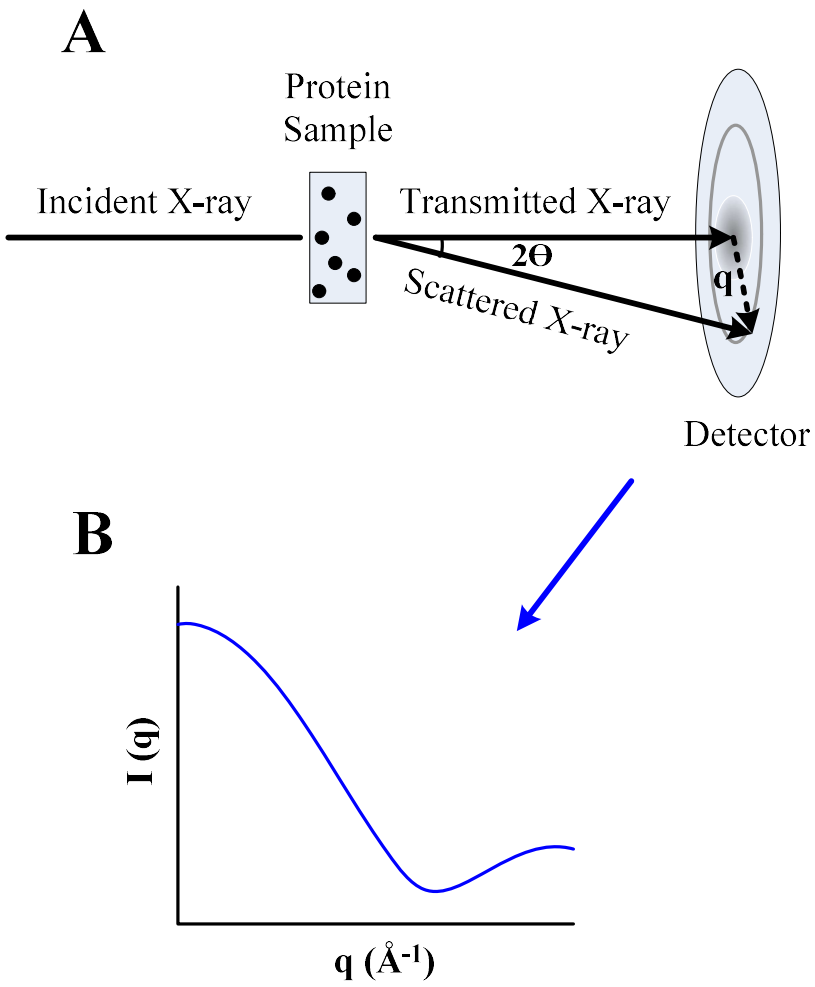

Figure 5.2: Schematic picture of basic SAXS. (**A**) Schematic representation of a SAXS experiment. (**B**) SAXS profile.

5.2 Material and methods

5.2.1 Circular dichroism (CD)

Melting temperature measurement

Circular dichroism (CD) experiments were carried out to measure the melting temperature of the protein at different pH values 6.5 and 4.5 using a Jasco J-810 automatic recording spectropolarimeter and quartz cell cuvette with a path length of 0.05 cm. The far-UV CD data of the unlabeled OnubPBP3 were collected at pH 6.5 and pH 4.5 with a protein concentration of 30 µM

in 15 mM phosphate buffer. CD spectra of the 15 mM phosphate buffer at pH 6.5 and pH 4.5 were collected as a baseline. To produce one final spectrum with averaged random noise at each temperature, four repeat spectra were collected. The data were collected over the wavelength range from 185 to 250 nm and over the temperature range from 40 °C to 108 °C in 2 °C increments, with a ramp rate of 5 °C per min.

Determination of protein secondary structures

To characterize the secondary structure content of OnubPBP3 at four different pH values 8.0, 6.5, 5.5, and 4.5, the far-UV CD spectrum of the protein was used. The experiment was performed by using quartz cell cuvette with a path length of 0.05 cm. The CD data of the unlabeled OnubPBP3 were collected at pH 8.0, 6.5, 5.5, and 4.5 with a protein concentration of 30 μM in 15 mM phosphate buffer. CD spectra of the 15 mM phosphate buffer at the four pH values were collected as a baseline. The data were collected over the wavelength window from 185 to 250 nm at 25 °C. The spectrum at each pH was collected five times and averaged to produce the final spectrum. The secondary structure contents of the protein were estimated by deconvoluting its far-UV CD spectrum by using CDSSTR program incorporated in CDPro software package.

Assessment of protein tertiary structure

Near-UV CD spectrum was recorded to study the tertiary structure of OnubPBP3 at different pH values 8.0, 6.5, 5.5, and 4.5. To carry out this experiment, a quartz cell cuvette with a path length of 10 mm was used. The CD data of the protein were collected at pH 8.0, 6.5, 5.5, and 4.5 with a protein concentration of 250 μM in 15 mM phosphate buffer. CD spectra of the 15 mM phosphate buffer at the four pH values were collected as a baseline. The data were collected over the wavelength window from 250 to 350 nm at 25 °C. The spectrum at each pH was collected five times and averaged to produce the final spectrum.

5.2.2 NMR Spectroscopy

NMR samples consisted of 600 µM uniformly ^{15}N-labeled OnubPBP3 in 50 mM phosphate buffer at pH 6.5, 95% H_2O, 5% D_2O, 1 mM EDTA and 0.01% (w/v) NaN_3 in a Shigemi tube. NMR data were collected on a Bruker Neo 600 MHz spectrometer at 35 °C. The two-dimensional-[^{1}H, ^{15}N] heteronuclear single quantum coherence (HSQC) experiments were performed on the OnubPBP3 sample at three different pH values (6.5, 5.5, 4.5). The HSQC spectra were collected using the Bruker pulse sequence hsqcetgpsi2. The transmitter frequency offset was set to 117 ppm in the ^{15}N dimension and 4.7 ppm in the ^{1}H dimension. The spectral width was 9615.385 Hz and 2216.148 Hz for ^{1}H and ^{15}N dimensions, respectively. The data were collected with 32 scans and 2048 complex points in the ^{1}H dimension and 512 complex points in the ^{15}N dimension. The first spectrum collected for the protein was at pH 6.5. The pH of the sample was reduced to 5.5 and then to 4.5 by adding 1 M HCl to the sample in 1.5 ml tube. After recording the spectrum of the protein at pH 4.5, the pH was changed back to 6.5 by adding 1 M NaOH and an HSQC experiment was collected. The collected data were processed and analyzed by NMRPipe21 and Sparky22 respectively.

5.2.3 Small Angle X-ray Scattering (SAXS)

Small-angle X-ray scattering data were collected at beamline 12-ID-B of the Advanced Photon Source (APS) of Argonne National Laboratory. Four samples of OnubPBP3 were prepared at different pH values 8.0, 6.5, 5.5, and 4.5. The samples at pH 8.0 and 6.5 were prepared by dialyzing 200 µl of 5 mg/ml of OnubPBP3 for each pH with 1 L of 50 mM HEPES buffer containing 5% glycerol at pH 8.0 and 6.5, respectively. On the other hand, the samples with low pH (5.5 and 4.5) were prepared by dialyzing 200 µl of 5 mg/ml of OnubPBP3 for each pH with 1 L of 50 mM acetate buffer containing 5% glycerol at pH 8.0 and 6.5, respectively. The RAW program[162] was used for normalizing, correcting, and reducing the resulting SAXS signals. A sample of the buffer taken

after the dialysis of each protein sample was used for background scattering by subtracting the scattering curve of the corresponding buffer solution from the scattering curve of the protein sample. Moreover, the RAW program was employed to obtain the radius of gyration (Rg), pair distribution function (P(r)), and the maximum dimension (Dmax). GNOM[163] in RAW was used to determine the initial data for reconstructing low-resolution molecular envelopes of OnubPBP3 at various pH values. DAMMIF[164] in ATSAS was utilized to extract the optimized structures. The CRYSOL 3.0 program[165] was employed for evaluating the solution scattering from OnubPBP3 samples by fitting the experimental scattering curves from small SAXS to the calculated NMR OnubPBP3 structure. The Correlation Map (CorMap) test,[166] which is a novel goodness-of-fit test, was used to evaluate the fitness of experimental SAXS curve to the calculated SAXS curve based on NMR structure or by GNOM. CorMap is an independent test of error estimates and uses the probability of similarity (p-value) between the two curves to assess their fitness. If the p-value is higher than 0.01, the observed difference between any two SAXS curves may be considered statistically insubstantial.

On the basis of the scattering intensity curves, the various plots; such as the Guinier plot, Kratky plot, and Pair Distance Distribution Function p(r) were employed to extract the structural information of OnubPBP3 at different pH values.

The Guinier plot (ln I(q) vs q^2) was utilized to acquire the size of the protein expressed by the radius of gyration (Rg) from the low-resolution scattering ($qRg \leq 1.3$):

$$\boldsymbol{I(q) \cong I(0)\exp(-q^2R_g^2/3)} \qquad \textbf{(5.2)}$$

where I(q) is the scattering intensity and I(0) is the intensity at zero angle.

The quality of the experimental SAXS data can be determined from the linearity of the Guinier plot. A deviation from linearity produces either a concave curve in the plot due to aggregations or a convex curve as a result of repulsion between the scatterers.

Rg is a measure of the overall size of a macromolecule and it is defined as the root mean square distance to the center of density in the macromolecule weighted by the scattering length density. Thus, Rg has different values for particles of the same volume but different shape.

To study the folding conformation and flexibility of OnubPBP3, a Kratky plot was used. It is created by plotting $q^2 I(q)$ against q, and uses the high angle region which corresponds to a small distance in real space. The shape of the plot helps to assess the folding state of proteins: a bell-shaped curve with a distinct maximum at low q indicates a globular protein (well-structured), while a plateau curve is characteristic of an unfolded protein (random coil-like conformation/unfolded proteins). The difference can be easily observed by using a normalized Kratky plot $(qRg)^2 \cdot I(q)/I(0)$ vs qRg. In this plot, the well-folded protein will have a peak position at $qR \approx (3)^{1/2} \approx 1.73$ and the peak height should be around 1.1.

The Pair Distance Distribution Function p(r) describes the distribution of atom-to-atom within the protein molecule weighted by the respective electron densities. The point where p(r) is decaying to zero known as the maximum distance (D_{max}), represents the maximum distance within noninteracting protein molecules. An incorrect D_{max} leads to a different Rg value than the one determined by the Guinier approximation. The p(r) function depends on the shape of the protein molecules. Globular proteins produce a sharp, approximately symmetric peak, while proteins with anisotropic shapes result in an asymmetric shape of the peak.

5.3 Results and discussion

5.3.1 Characterization of OnubPBP3 by CD spectroscopy

In order to estimate the thermal stability of the OnubPBP3 at pH 6.5 and pH 4.5, CD spectroscopy was employed. The CD spectra showed that the melting temperature (Tm) of the protein at pH 6.5 was 92.19 °C (Figure 5.3A), indicating a high thermal stability. By reducing the pH to 4.5, the melting temperature of the protein decreased to 83.95 °C (Figure 5.3B) suggesting a reduction in the protein thermal stability. According to the obtained melting curves, OnubPBP3 at either pH would not denature at 35 °C, which is the temperature used to perform NMR experiments. Since the melting temperature of proteins is considered as an indicator of atomic bonding strength,[167] lowering the pH to 4.5 caused OnubPBP3 to lose some of intramolecular interactions leading to the reduction in the melting temperature.

The far-UV CD spectra of OnubPBP3 at pH 8.0, 6.5, 5.5, and 4.5 (Figure 5.4A) showed one positive band at 193 nm and two negative bands around 208-209 and 222-225 nm. This CD spectrum pattern is characteristic of a protein with a high content of α-helices. The CD spectra at pH 8.0 and 6.5 were quite similar; however, the degree of helicity increased as the pH decreased from 6.5 to 5.5 indicating a conformational change occurred in the protein structure. The pH reduction from 5.5 to 4.5 produced a slightly different CD spectrum. Analysis of the secondary structure content of OnubPBP3 at all four pH values by CDPro program suites was performed. The analysis results showed the percentage of helix content increased from 65% at pH 6.5 to 82% at pH 4.5. Moreover, the analysis exhibited the helix content at pH 8.0 (68%) similar to the one at pH 6.5 and the helix content at pH 5.5 (84%) similar to the one at pH 4.5 (Table 5.3).

The near-UV region was used to produce the CD fingerprint of the tertiary structure of OnubPBP3. The near-UV CD spectra of OnubPBP3 at pH 8.0, 6.5, 5.5, and 4.5 (Figure 5.4B) showed a remarkable change in the tertiary structure when the pH was lowered from 6.5 to 5.5. The spectra

at pH 8.0 and 6.5 displayed a similar broad negative band around 270 nm. Disulfide bonds and all aromatic amino acid residues, phenylalanine, tyrosine, and tryptophan, contribute to this band. OnubPBP3 contains in the structure 3 disulfide bonds, 5 phenylalanines, 2 tyrosines, and 3 tryptophans, which supports this practical observation. The spectra of pH 5.5 and 4.5 were similar; however, they showed less intensity in the negative band than the spectra at high pH. This significant change in the spectrum was caused by a change in the local environment of the aromatic side chains obtaining a more flexible orientation as a result of partial unfolding. Far-UV and near-UV CD spectra suggest that OnubPBP3 experienced a molten globule-like state by decreasing the pH to 5.5 and 4.5.

Table 5.3: The secondary structure content of OnubPBP3 at different pH.

pH	Helix	Sheet	Turn	Random
4.5	0.82	0.04	0.050	0.09
5.5	0.84	0.13	0.040	0.08
6.5	0.65	0.10	0.10	0.15
8.0	0.68	0.07	0.10	0.15

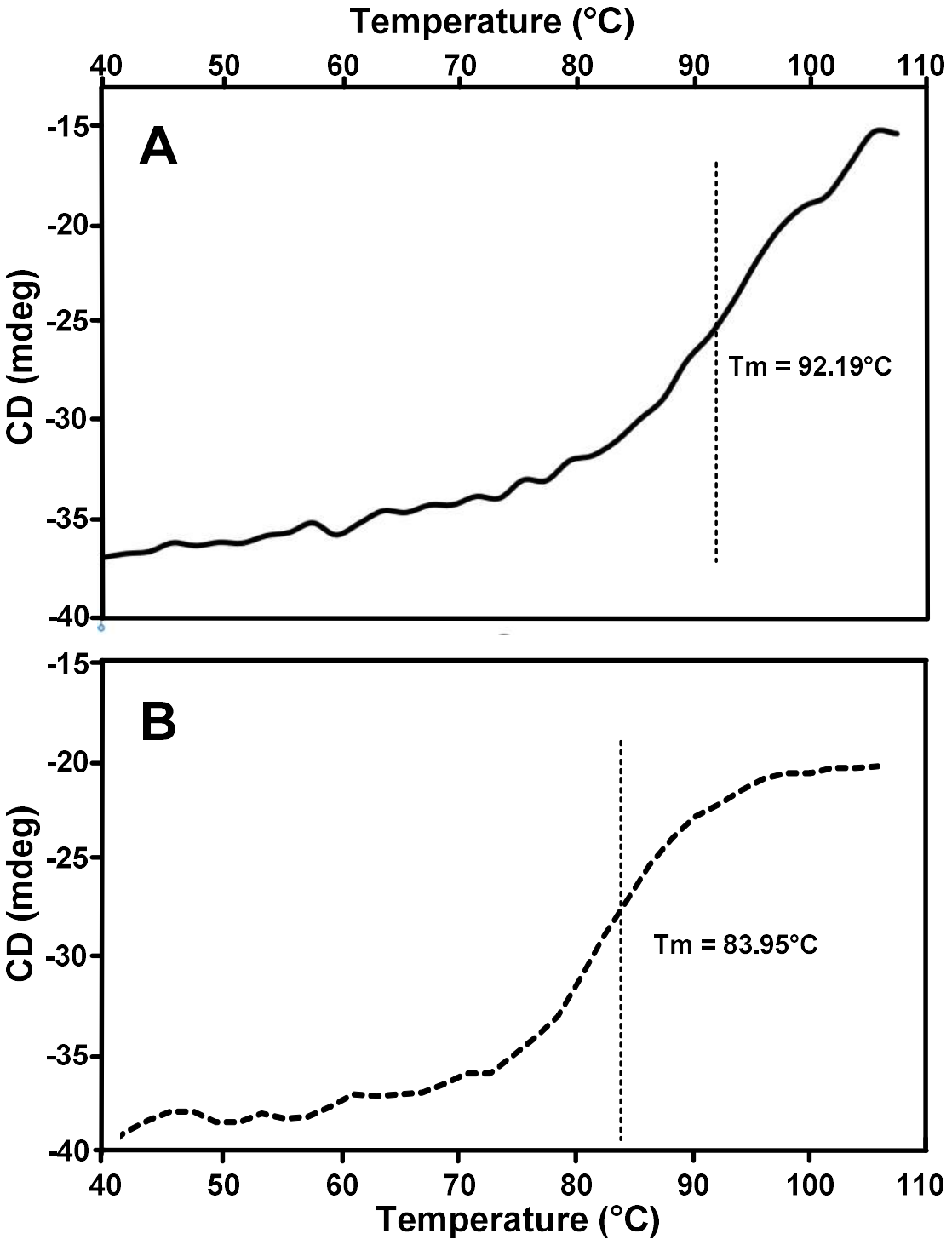

Figure 5.3: CD melting temperature curves of 30 µM OnubPBP3 in 15 mM phosphate buffer obtained at 209 nm. (**A**) Melting temperature curve of the protein at pH 6.5. (**B**) Melting temperature curve of the protein at pH 4.5.

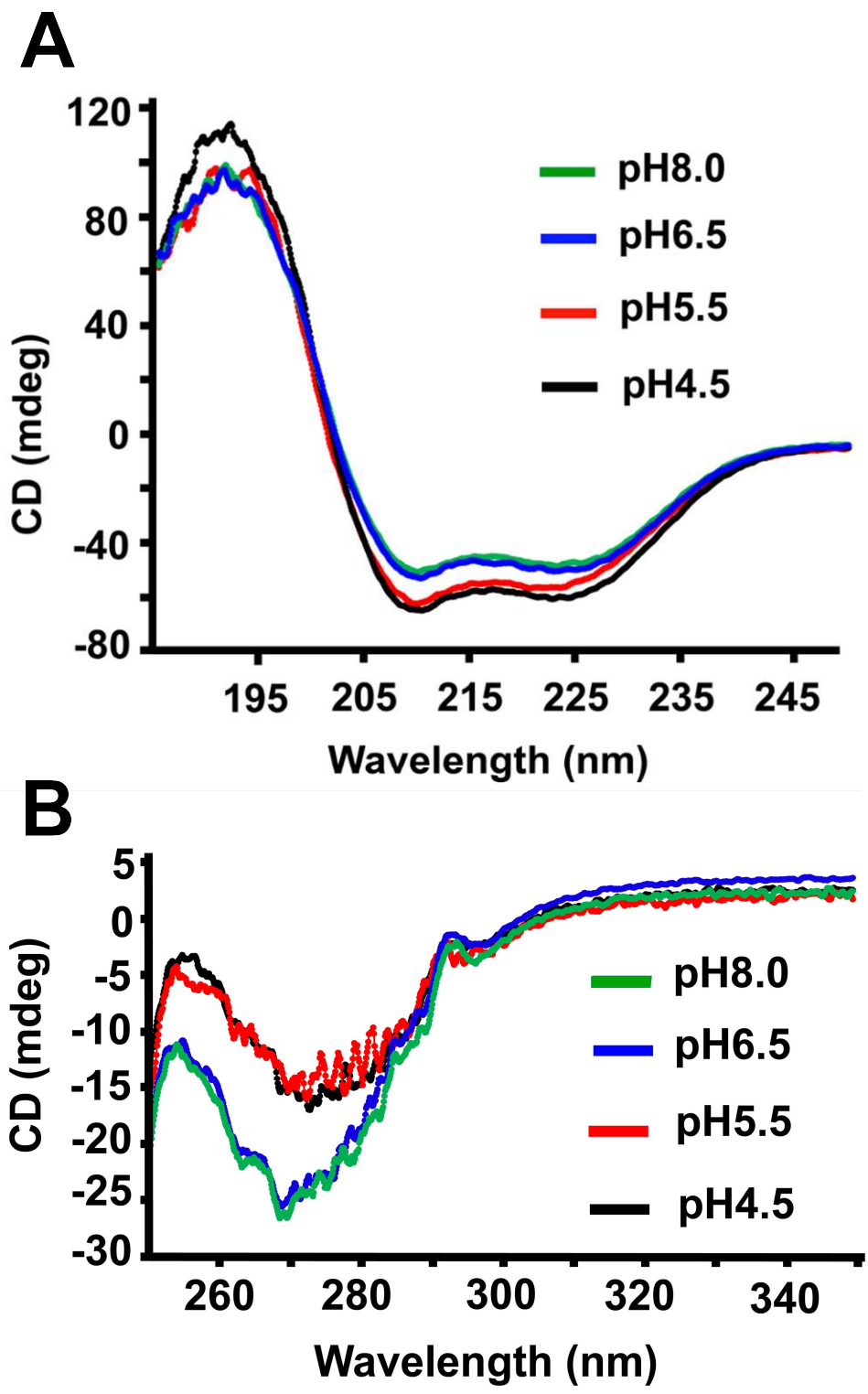

Figure 5.4: CD spectra of OnubPBP3 in 15 mM phosphate buffer at pH 4.5, 5.5, 6.5, and 8.0. (**A**) Far-ultraviolet CD spectra (30 µM protein, 25 °C), (**B**) near-ultraviolet CD spectra (250 µM protein, 25 °C).

5.3.2 NMR spectroscopy

The 2D-[^{1}H,^{15}N] HSQC spectrum was collected for ^{15}N-labeled OnubPBP3 at pH 6.5 (Figure 5.5 A). This HSQC spectrum shows the fingerprint region of OnubPBP3. Any change in this spectrum due to changes in the protein environment such as variation in pH, solvent, temperature, salt concentration, ligand binding, etc. reflects a change in protein conformation. Thus, the conformational change of a protein can be observed conveniently by monitoring the HSQC spectrum under various conditions.

To investigate the effect of pH on the OnubPBP3 conformation, HSQC spectra were collected at pH 6.5, 5.5, and 4.5. The NMR spectra displayed that the peak dispersion diminished with pH reduction from near neutral to the acidic range (Figure 5.5A-D). The spectrum of the protein at pH 5.5 showed that reducing the pH by one unit is sufficient to cause a strong change in the conformation (Figure 5.5B). Lowering the pH to 4.5 showed a dramatic effect, with most of the peaks collapsing to the center of the HSQC (Figure 5.5C). Indeed, many peaks were missing at pH 4.5 which can be noticed from Figure 5.5D. The disappearance of peaks has been observed with proteins in molten globule states due to exchanges between different conformations.[168] The lack of chemical shift dispersion is an indication of protein denaturation including formation of a molten globule.[151] Together with the CD spectroscopy results, this suggests that OnubPBP3 is in a molten globular state at acidic pH (5.5 and 4.5). To test whether OnubPBP3 was irreversibly denatured at pH 4.5, the pH was increased back to 6.5. Interestingly, the HSQC collected after this pH increase matched the original HSQC spectrum very well, suggesting that this pH-induced effect is reversible (Figure 5.6).

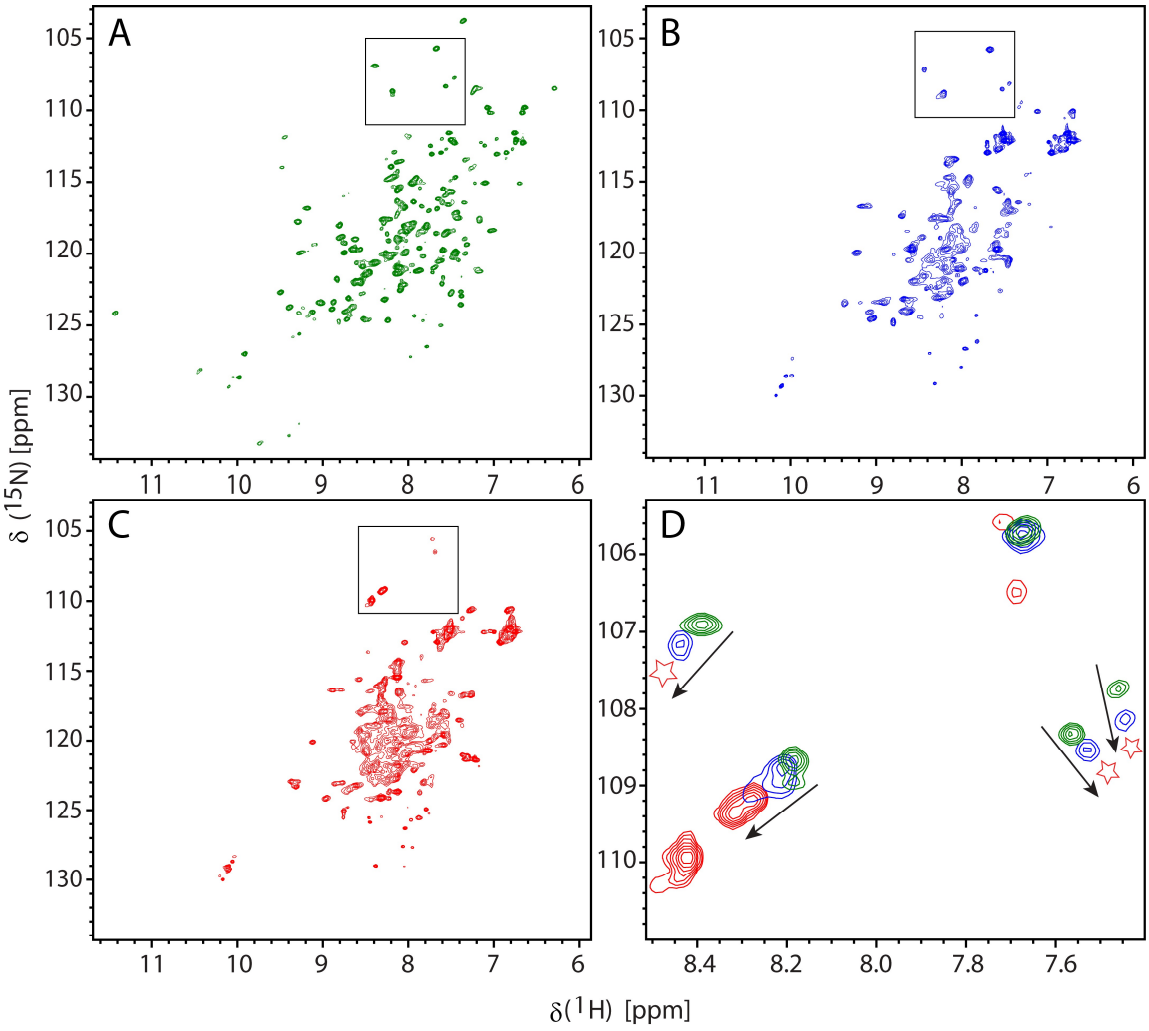

Figure 5.5: Two-dimensional [^{1}H, ^{15}N] HSQC spectra obtained on 600 µM OnubPBP3 at different pH. (**A**) OnubPBP3 at pH 6.5. (**B**) OnubPBP3 at pH 5.5. (**C**) OnubPBP3 at pH 4.5. (**D**) Overlay expanded view of the boxed region from three spectra collected at different pH values: 6.5 in green, 5.5 in blue, and 4.5 in red; arrows indicate the direction of the chemical shift change, red stars refer to missing peaks at pH 4.5.

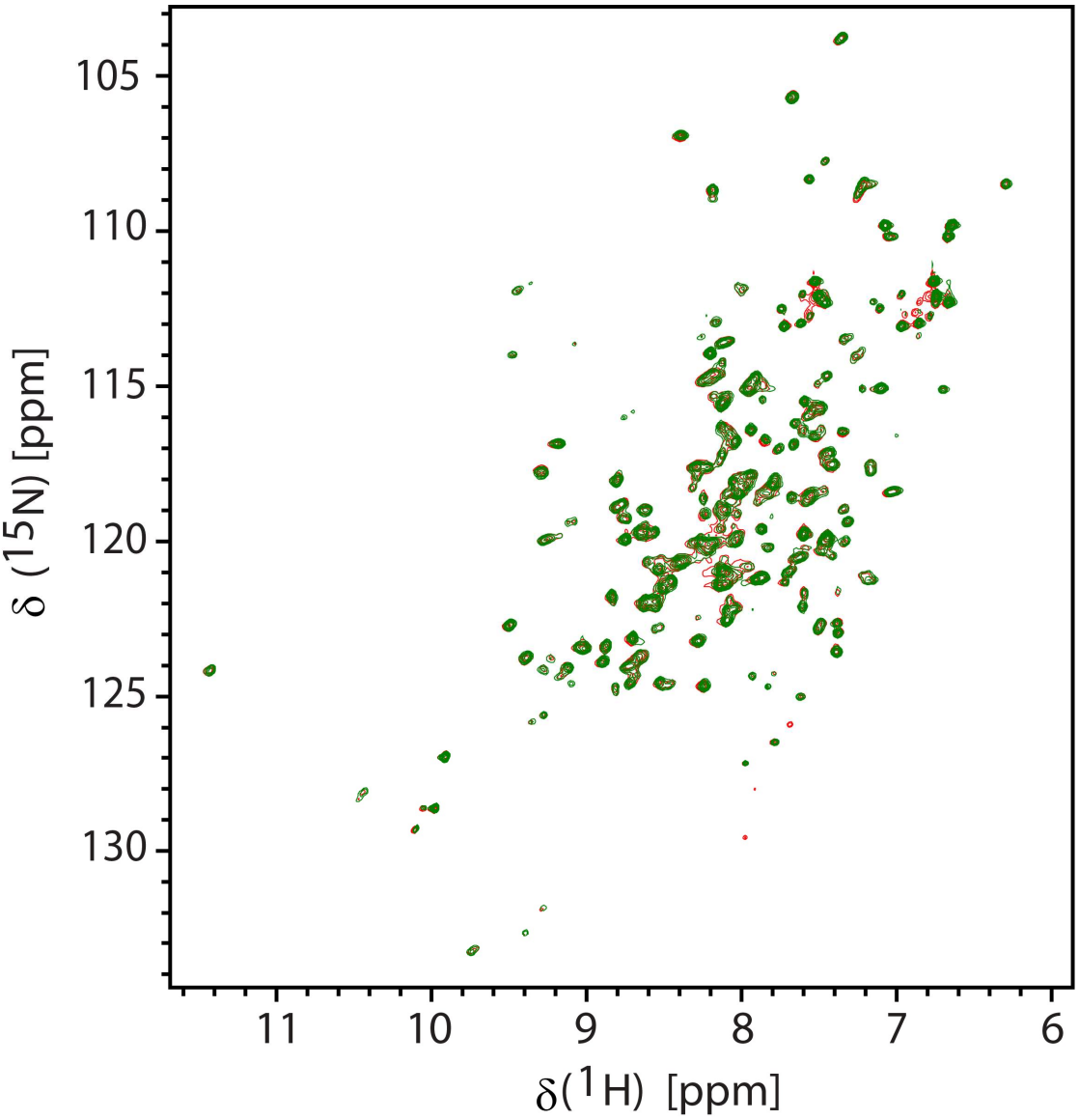

Figure 5.6: Overlay of two-dimensional [^{1}H, ^{15}N] HSQC spectra acquired on 600 µM OnubPBP3 at pH 6.5 before pH change (green) to pH 4.5 and after restoring to pH 6.5 (red).

5.3.3 Small Angle X-ray Scattering (SAXS)

To gain more information about the pH dependence of OnubPBP3 structure, SAXS experiments were performed on OnubPBP3 solutions at various pH values 8.0, 6.5, 5.5, and 4.5. The background-subtracted scattering profile of the protein at each pH is shown from angular range of q_{min} = 0.026 Å^{-1} to q_{max} = 0.3000 Å^{-1} (Figure 5.7). The SAXS curves demonstrated a marked change at high q value around 0.226 Å^{-1} with pH change from acidic to basic. The radii of gyration were

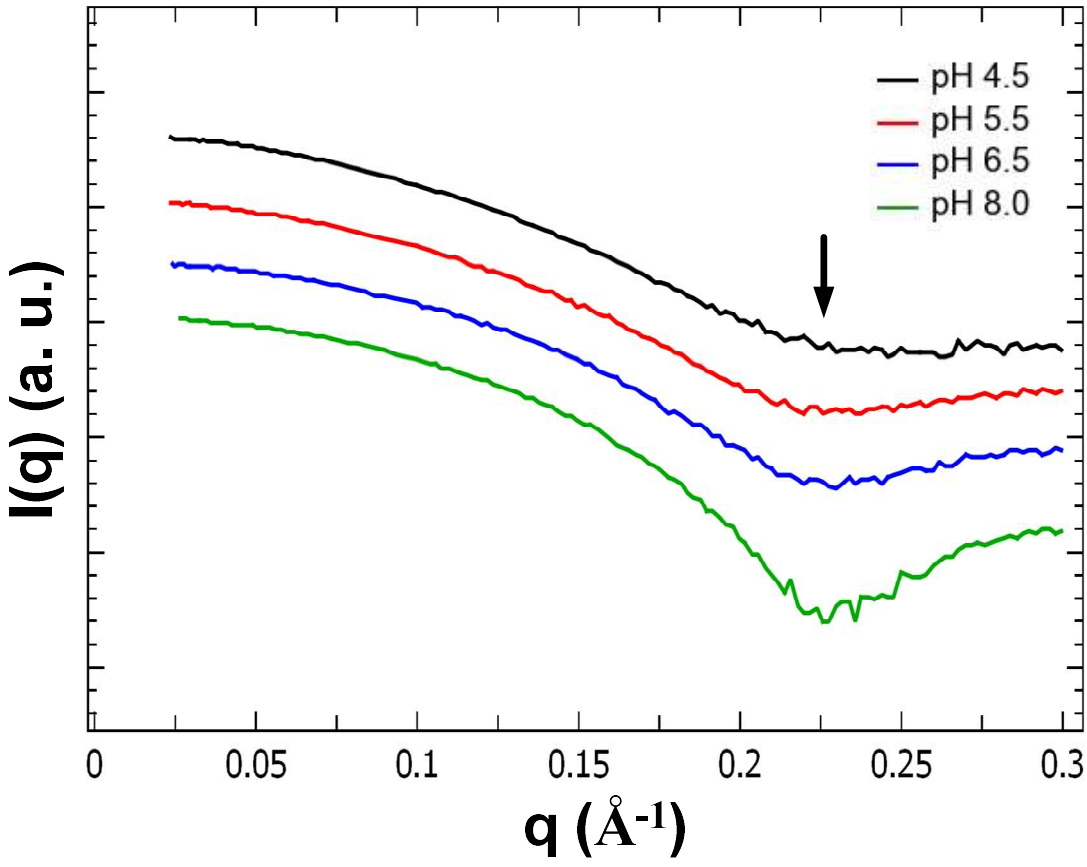

Figure 5.7: Experimental SAXS curves of OnubPBP3 at pH 4.5, 5.5, 6.5, and 8.0. The arrow in the figure is pointing to the most prominent change in the q value with increasing pH.

estimated by using both Guinier analysis and computation of the real-space inter-atomic distance distribution function P(r). The Guinier analysis of the protein samples at different pH values were linear (Figure 5.8), suggesting that the samples were free of aggregation or interparticle interference consistent with acquisition of monodisperse high-quality data.

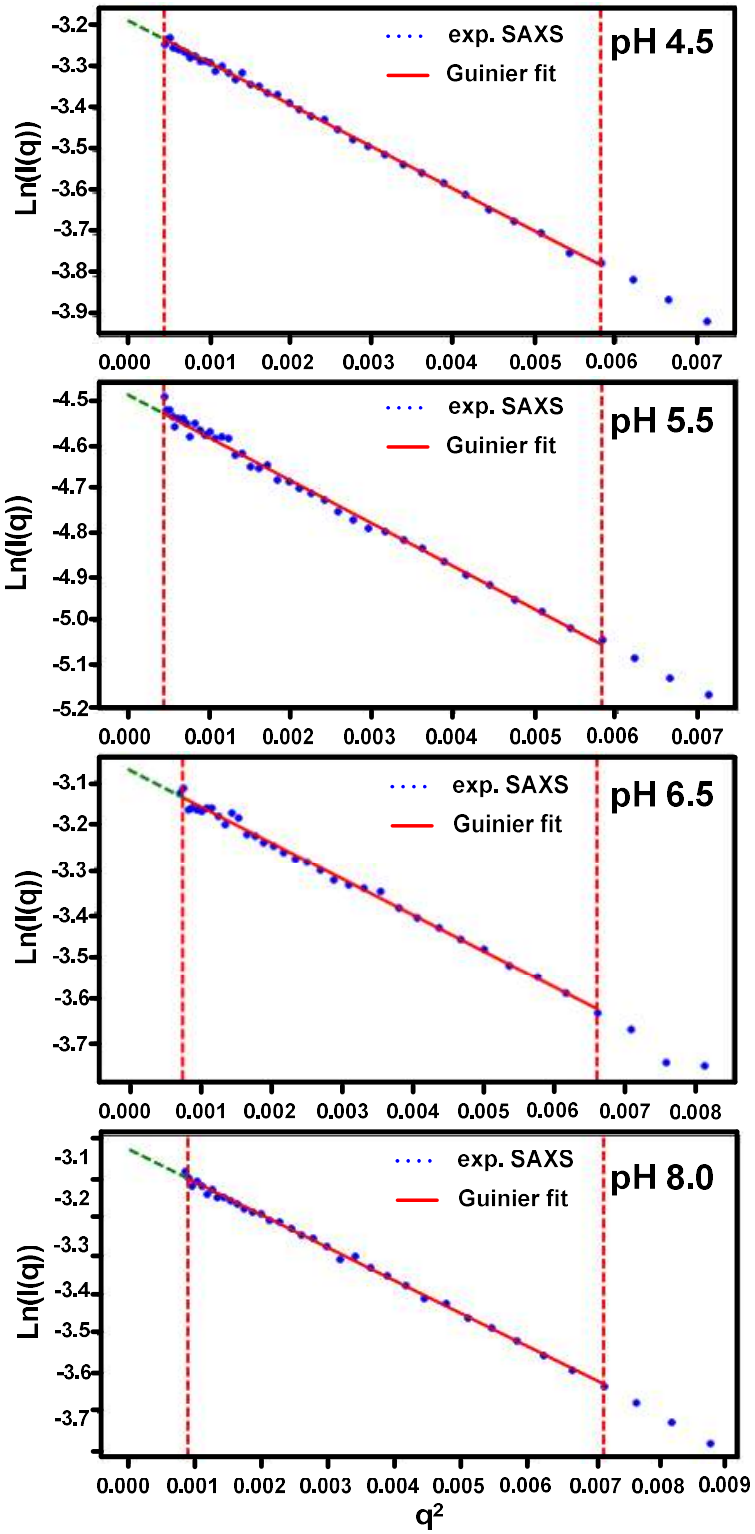

Figure 5.8: Guinier plots of of the scattering curves of OnubPBP3 at various pH values showing linearity over the Guinier range to give accurate Rg values.

The Rg values obtained from Guinier analysis of OnubPBP3 were increased by reducing the pH. The Rg of the protein at pH 8.0, 6.5, 5.5, and 4.5 were 15.84, 15.75, 17.07, and 17.44, respectively. These results showed that OnubPBP3 was less compact at low pH 5.5 and 4.5 than at pH 8.0 and 6.5, indicating the protein was partially unfolded at acidic pH. All Rg values from Guinier analysis were confirmed by P(r) analysis (Table 5.3).

The GNOM program was used not only to calculate the P(r) of the protein at each condition but also to calculate the inverse Fourier transform to evaluate how well the resulting I(q) fits the scattering curve. Calculating the inverse Fourier transform is a vital step to determine the correct values of P(r) and D_{max}. The optimal D_{max} of OnubPBP3 at each pH was determined after finding (1) the Rg value obtained by P(r) was similar to the one extracted from Guinier plot (Table 5.3), (2) the resulting P(r) reached zero smoothly without trailing tail, negative data point, or strong oscillations (Figure 5.9), (3) the resulting I(q) fit well the scattering curve by obtaining a p-value

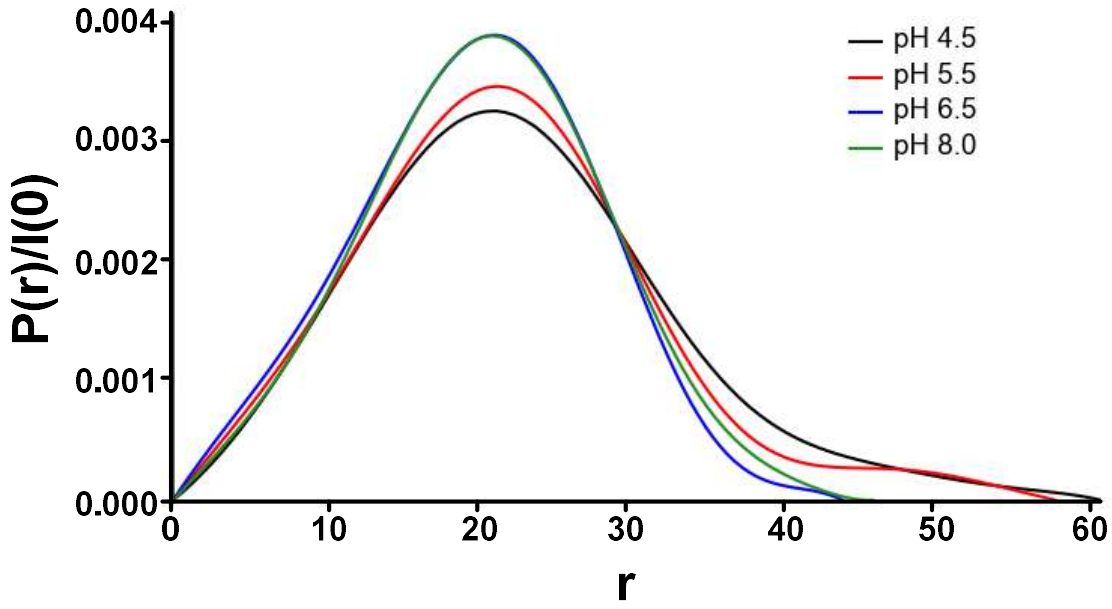

Figure 5.9: Pair distance distribution functions (P(r)). It was calculated from the scattering profiles by the RAW software for OnubPBP3 at the following pH values: pH 4.5 (black), pH 5.5 (red), pH 6.5 (blue), and pH 8.0 (green).

higher than 0.01 (Figure 5.10). The D_{max} of OnubPBP3 showed a dependence on the pH of the solution. The D_{max} values were 46, 44, 58, and 61 Å for pH 8.0, 6.5, 5.5, and 4.5, respectively. This finding suggests significant shifts in the conformation of OnubPBP3 occurred when the pH was reduced (more acidic) due to occurred partial protein unfolding.

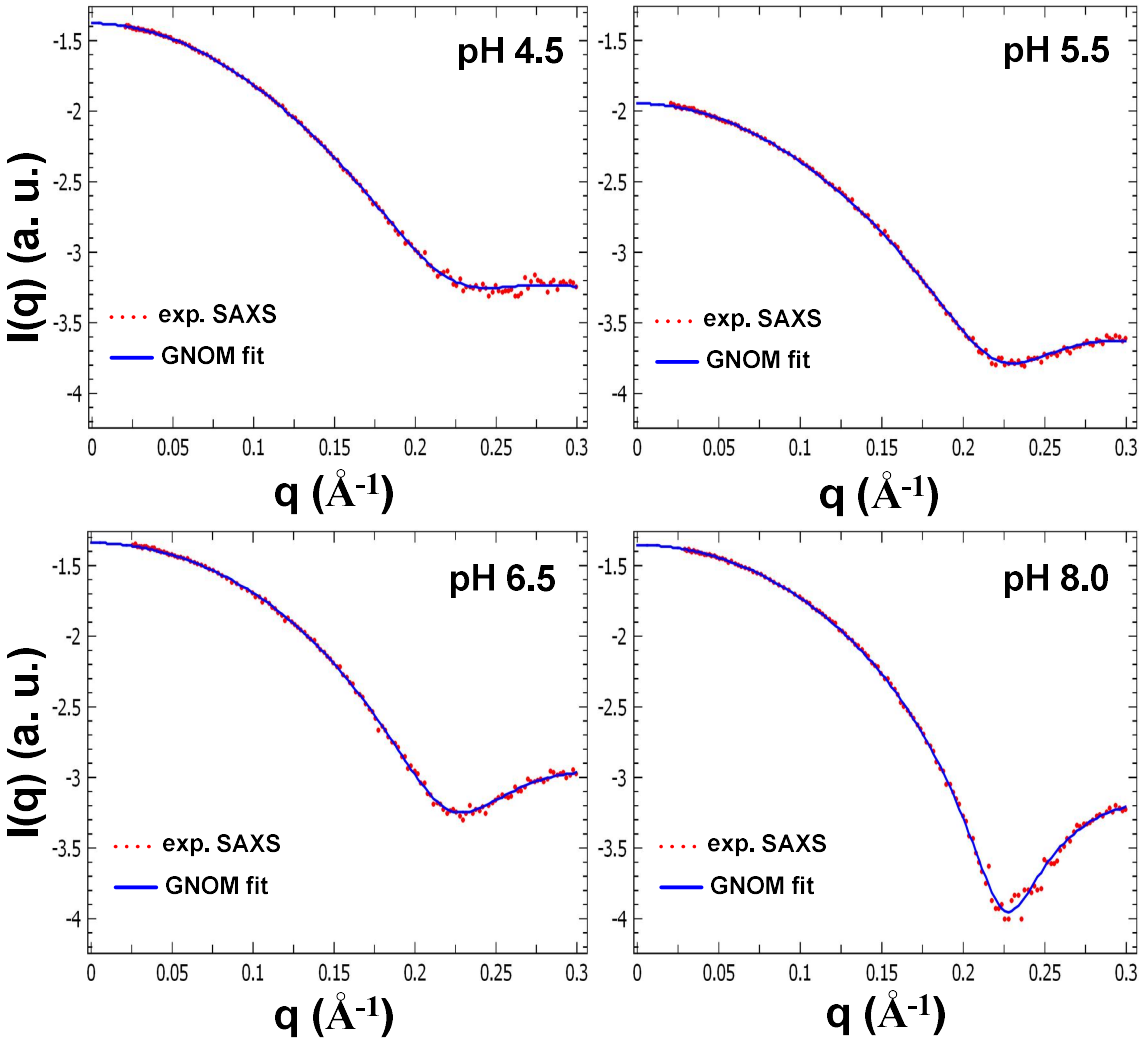

Figure 5.10: Experimental SAXS data (red) of OnubPBP3 at different pH values (4.5, 5.5, 6.5, and 8.0) and fitted curves from the GNOM program (blue) that yielded well-behaved P(r) curves. All p-values of the fitted curves were higher than 0.01.

The normalized Kratky plot was utilized to assess qualitatively the degree of the protein unfolding. This plot showed that OnubPBP3 at pH 5.5 and 4.5 has a peak height and position higher than the ones at pH 6.5 and 8.0 (Figure 5.11). This finding indicates that the protein was partially unfolded at acidic pH.

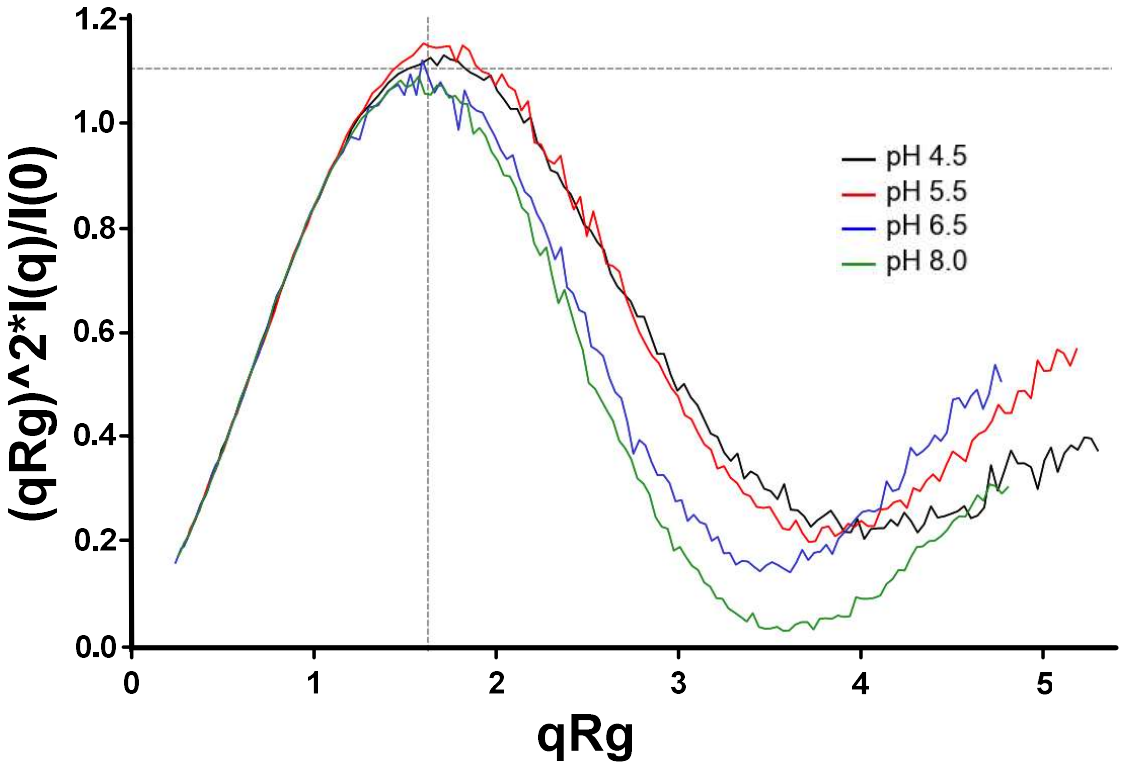

Figure 5.11: Normalized Kratky plots of OnubPBP3 at four pH values, pH 4.5 (black), pH 5.5 (red), pH 6.5 (blue), and pH 8.0 (green).

The three-dimensional models of the protein at all pH values were reconstructed from the distance distribution function P(r). Low-resolution envelope structures that may represent the actual structures of the protein at the particular pH were averaged from 20 ab initio models. The CRYSOL program was employed to evaluate the calculated NMR structure of OnubPBP3 at pH 6.5 and fit it to the experimental scattering curves. The NMR structure of the protein showed an excellent agreement with the experimental scattering curve obtained at pH 6.5 with p-value (0.07), while the experimental scattering curve of pH 8.0, 5.5, and 4.5 showed low fitting quality with p-values 0.001, $4.92*10^{-7}$, and $4.75*10^{-14}$, respectively (Figure 5.12). The NMR structure of OnubPBP3 was fitted to the average envelope structures at pH 6.5 and 8.0 (Figure 5.13A and B). However, average envelope structures of the protein at pH 5.5 and 4.5 were larger than the ones at high pH (Figure 5.14C and D), showing that OnubPBP3 switched its conformation due to the pH reduction to 4.5.

To summarize the results of the SAXS experiments, OnubPBP3 showed that by lowering the pH from 6.5 to 4.5, Rg increased by 10%, D_{max} increased by 38%, and the folding decreased. Combined all together, these results suggest that OnubPBP3 is folded very well at near neutral pH (6.5) and present as a molten globule at acidic pH due to the partial refolding. In contrast to the well-studied lepidopteran PBPs which have well-defined low-pH conformations, OnubPBP3 did not show a clear low-pH conformation. These results suggest that OnubPBP3 may adopt a different mechanism for the ligand release.

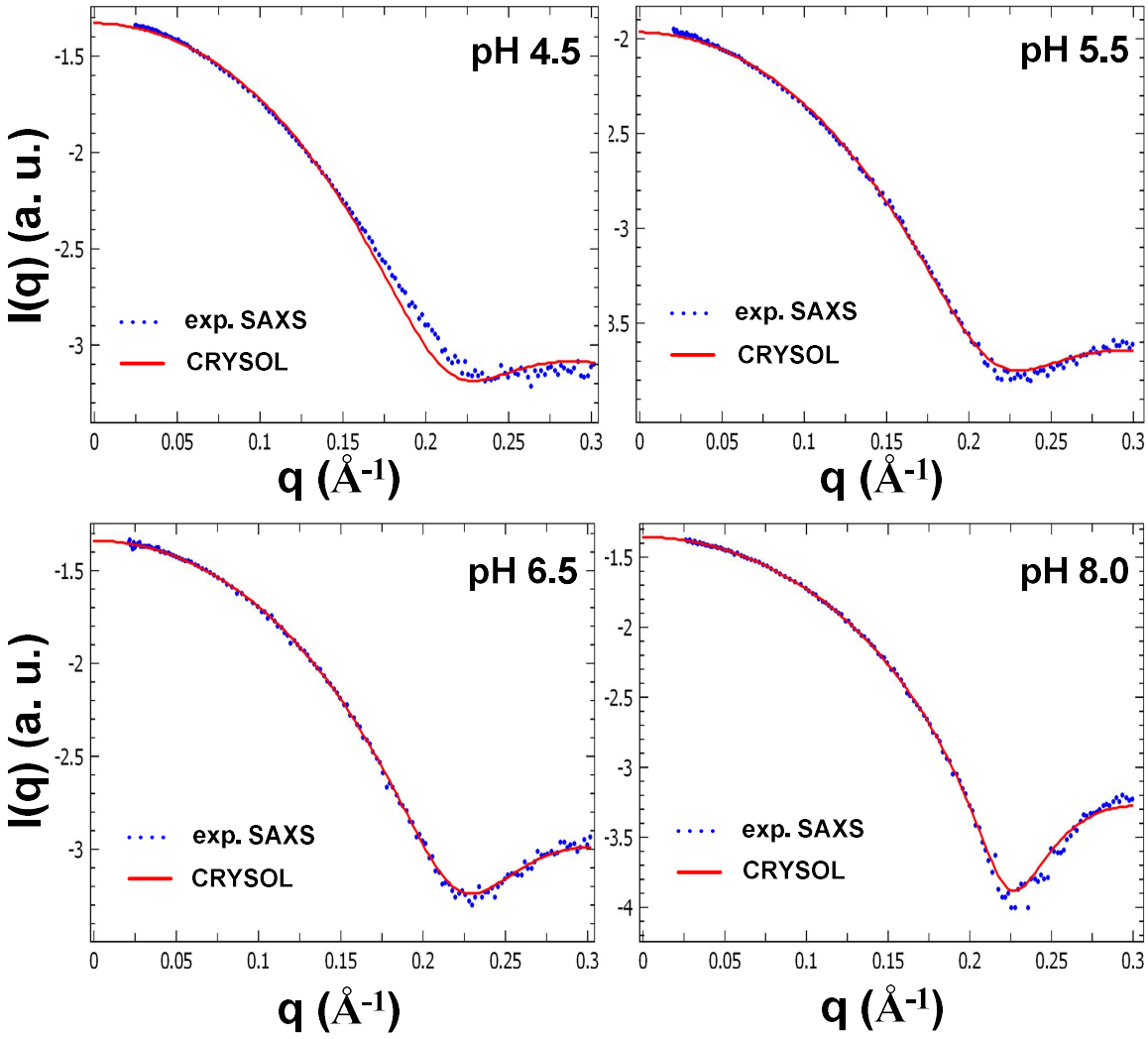

Figure 5.12: Experimental SAXS data (blue) of OnubPBP3 at different pH values (4.5, 5.5, 6.5, and 8.0) and fitted curves from CRYSOL program (red). p-values of the fitted curves at pH 8.0, 6.5, 5.5, and 4.5 were 0.001, 0.07, $4.92*10^{-7}$, and $4.75*10^{-14}$, respectively.

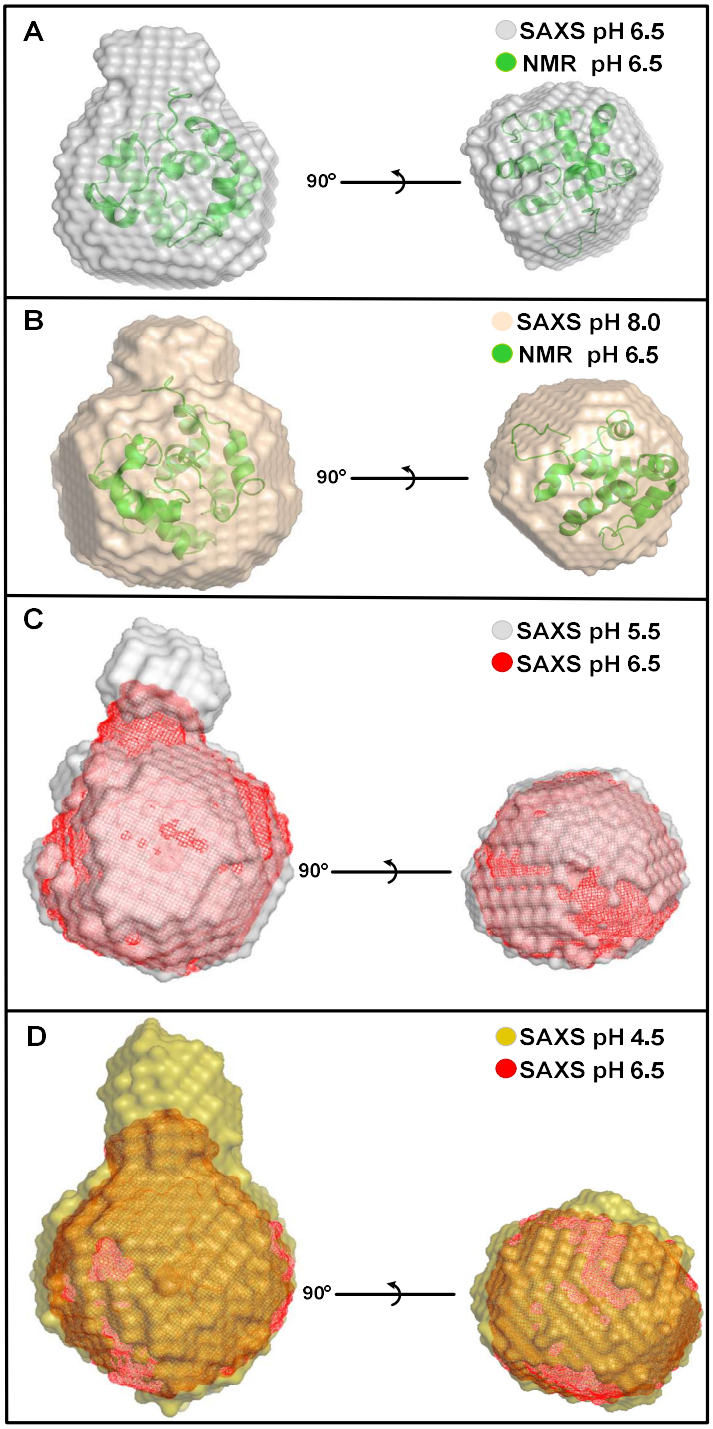

Figure 5.13: The averaged SAXS envelopes for OnubPBP3 at different pH values. (**A**) The SAXS envelope of the protein at pH 6.5 (gray) and the NMR structure at pH 6.5 (green) are superimposed. (**B**) The SAXS envelope of the protein at pH 8.0 (wheat) and the NMR structure at pH 6.5 (green) are superimposed. (**C**) The SAXS envelopes of the protein at pH 5.5 (gray) and at pH 6.5 (red) are superimposed. (**D**) The SAXS envelopes of the protein at pH 4.5 (olive) and at pH 6.5 (red) are superimposed.

Table 5.4: Data collection and SAXS-derived parameters

Instrument	The BioCat Beamline 18ID (Argonne National Laboratory)			
Detector	Pilatus 2M			
Beam geometry μm^2	200 (H) x 40(V)			
Wavelength (Å)	1.03			
q range ($Å^{-1}$)	0.002-0.5			
Sample to Detector Distance	4m			
Structural Parameters	**pH 8.0**	**pH 6.5**	**pH 5.5**	**pH 4.5**
q_{max} ($Å^{-1}$)	0.3	0.3	0.3	0.3
I (0), Guinier analysis	0.0438 ± 0.0008	0.0462 ± 0.0010	0.0112 ± 0.0001	0.0413 ± 0.0002
Rg (Å), Guinier analysis	15.842 ± 0.437	15.755 ± 0.588	17.072 ± 0.409	17.438 ± 0.135
I(0) (Å), P(r) analysis	0.0438 ± 0.0006	0.0453 ± 0.0007	0.0120 ± 0.0001	0.0414 ± 0.0002
R(g) (Å), P(r) analysis	15.65 ± 0.224	15.230 ± 0.169	17.000 ± 0.389	17.610 ± 0.146
D_{max} (Å)	46	44	58	61
Software employed				
Data processing	RAW and PRIMUS			
Ab initio analysis	DAMMIF			
Validation and averaging	DAMAVER			
Computation of model intensities	CRYSOL			

REFERENCES

1. Sheppard, N., Edward raymond andrew. *Biographical memoirs of fellows of the Royal Society. Royal Society (Great Britain)* **2003,** *49*, 1-14.

2. Saunders, M.; Wishnia, A.; Kirkwood, J. G., The nuclear magnetic resonance spectrum of ribonuclease1. *Journal of the American Chemical Society* **1957,** *79* (12), 3289-3290.

3. Laukien, D. D.; Tschopp, W. H., Superconducting NMR magnet design. *Concepts in Magnetic Resonance* **1994,** *6* (4), 255-273.

4. Ernst, R. R.; Anderson, W. A., Application of fourier transform spectroscopy to magnetic resonance. *Review of Scientific Instruments* **1966,** *37* (1), 93-102.

5. Kumar, A., Development of two-dimensional NMR. *Resonance* **2015,** *20* (11), 995-1002.

6. Aue, W. P.; Bartholdi, E.; Ernst, R. R., Two-dimensional spectroscopy. application to nuclear magnetic resonance. *The Journal of Chemical Physics* **1976,** *64* (5), 2229-2246.

7. Wüthrich, K., *NMR of proteins and nucleic acids*. 1986.

8. Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Skelton, N. J., *Protein NMR spectroscopy: principles and practice*. Elsevier: 1995.

9. Rule, G. S.; Hitchens, T. K., *Fundamentals of protein NMR spectroscopy*. Springer Science & Business Media: 2006; Vol. 5.

10. Berjanskii, M. V.; Wishart, D. S., Unraveling the meaning of chemical shifts in protein NMR. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2017,** *1865* (11, Part B), 1564-1576

11. Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D., 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *Journal of Biomolecular NMR* **1995,** *6* (2), 135-140.

12. Marion, D., An introduction to biological NMR spectroscopy. *Molecular & cellular proteomics : MCP* **2013,** *12* (11), 3006-25.

13. Gerothanassis, I. P.; Troganis, A.; Exarchou, V.; Barbarossou, K., Nuclear magnetic resonance (nmr) spectroscopy: basic principles and phenomena, and their applications to chemistry, biology and medicine. *Chemistry Education Research and Practice* **2002,** *3* (2), 229-252.

14. Ando, I.; Kobayashi, M.; Kanekiyo, M.; Kuroki, S.; Ando, S.; Matsukawa, S.; Kurosu, H.; Yasunaga, H.; Amiya, S., Chapter 4 - NMR spectroscopy in polymer science. In *Experimental Methods in Polymer Science*, Tanaka, T., Ed. Academic Press: Boston, 2000; pp 261-493.

15. Chary, K. V. R.; Govil, G., Multi-dimensional NMR. In *NMR in Biological Systems: From Molecules to Humans*, Chary, K. V. R.; Govil, G., Eds. Springer Netherlands: Dordrecht, 2008; pp 69-111.

16. Schmieder, P., Multidimensional NMR Spectroscopy. In *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*, Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; pp 1204-1208.

17. Markwick, P. R. L.; Malliavin, T.; Nilges, M., Structural biology by NMR: structure, dynamics, and interactions. *PLoS computational biology* **2008,** *4* (9), e1000168-e1000168.

18. Ohki, S.-y.; Kainosho, M., Stable isotope labeling methods for protein NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2008,** *53* (4), 208-226.

19. Cai, M.; Huang, Y.; Sakaguchi, K.; Clore, G. M.; Gronenborn, A. M.; Craigie, R., An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli. *J Biomol NMR* **1998,** *11* (1), 97-102.

20. Emwas, A.-H. M.; Merzaban, J. S.; Serrai, H., Chapter 3 - Theory and applications of NMR-based metabolomics in human disease diagnosis. In *Applications of NMR Spectroscopy*, ur-Rahman, A.; Choudhary, M. I., Eds. Bentham Science Publishers: 2015; pp 93-130.

21. Mandal, P. K.; Majumdar, A., A comprehensive discussion of HSQC and HMQC pulse sequences. *Concepts in Magnetic Resonance Part A* **2004,** *20A* (1), 1-23.

22. Matte, A.; Kozlov, G.; Trempe, J.-F.; Currie, M. A.; Burk, D.; Jia, Z.; Gehring, K.; Ekiel, I.; Berghuis, A. M.; Cygler, M., Preparation and characterization of bacterial protein complexes for structural analysis. In *Advances in Protein Chemistry and Structural Biology*, Joachimiak, A., Ed. Academic Press: 2009; Vol. 76, pp 1-42.

23. Oschkinat, H.; Cieslar, C.; Holak, T. A.; Marius Clore, G.; Gronenborn, A. M., Practical and theoretical aspects of three-dimensional homonuclear Hartmann-Hahn-nuclear overhauser enhancement spectroscopy of proteins. *Journal of Magnetic Resonance (1969)* **1989,** *83* (3), 450-472.

24. Inagaki, F., Protein NMR resonance assignment. In *Encyclopedia of Biophysics*, Roberts, G. C. K., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 2033-2037.

25. Kanelis, V.; Forman-Kay, J. D.; Kay, L. E., Multidimensional NMR methods for protein structure determination. *IUBMB life* **2001,** *52* (6), 291-302.

26. Lemak, A.; Gutmanas, A.; Chitayat, S.; Karra, M.; Farès, C.; Sunnerhagen, M.; Arrowsmith, C. H., A novel strategy for NMR resonance assignment and protein structure determination. *Journal of biomolecular NMR* **2011,** *49* (1), 27-38.

27. Güntert, P.; Mumenthaler, C.; Wüthrich, K., Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **1997,** *273* (1), 283-98.

28. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M., The Xplor-NIH NMR molecular structure determination package. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **2003,** *160* (1), 65-73.

29. Rieping, W.; Habeck, M.; Bardiaux, B.; Bernard, A.; Malliavin, T. E.; Nilges, M., ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **2007,** *23* (3), 381-2.

30. DePorter, T. L., Chapter 18 - Use of pheromones in feline practice. In *Feline Behavioral Health and Welfare*, Rodan, I.; Heath, S., Eds. W.B. Saunders: St. Louis, 2016; pp 235-244.

31. Vogt, R. G.; Rybczynski, R.; Lerner, M. R., Molecular cloning and sequencing of general odorant-binding proteins GOBP1 and GOBP2 from the tobacco hawk moth Manduca sexta: comparisons with other insect OBPs and their signal peptides. *J Neurosci* **1991,** *11* (10), 2972-2984.

32. Callahan, F. E.; Vogt, R. G.; Tucker, M. L.; Dickens, J. C.; Mattoo, A. K., High level expression of "male specific" pheromone binding proteins (PBPs) in the antennae of female noctuiid moths. *Insect Biochem Mol Biol* **2000,** *30* (6), 507-14.

33. Steinbrecht, R. A.; Laue, M.; Ziegelberger, G., Immunolocalization of pheromone-binding protein and general odorant-binding protein in olfactory sensilla of the silk moths Antheraea and Bombyx. *Cell and Tissue Research* **1995,** *282* (2), 203-217.

34. Vogt, R. G.; Prestwich, G. D.; Lerner, M. R., Odorant-binding-protein subfamilies associate with distinct classes of olfactory receptor neurons in insects. *Journal of Neurobiology* **1991,** *22* (1), 74-84.

35. Vogt, R. G.; Rybczynski, R.; Lerner, M. R., Molecular cloning and sequencing of general odorant-binding proteins GOBP1 and GOBP2 from the tobacco hawk moth Manduca sexta: comparisons with other insect OBPs and their signal peptides. *The Journal of Neuroscience* **1991,** *11* (10), 2972.

36. Lagarde, A.; Spinelli, S.; Tegoni, M.; He, X.; Field, L.; Zhou, J.-J.; Cambillau, C., The crystal structure of odorant binding protein 7 from Anopheles gambiae exhibits an outstanding adaptability of its binding site. *Journal of Molecular Biology* **2011,** *414* (3), 401-412.

37. Tegoni, M.; Campanacci, V.; Cambillau, C., Structural aspects of sexual attraction and chemical communication in insects. *Trends in Biochemical Sciences* **2004,** *29* (5), 257-264.

38. Lartigue, A.; Gruez, A.; Briand, L.; Blon, F.; Bézirard, V.; Walsh, M.; Pernollet, J. C.; Tegoni, M.; Cambillau, C., Sulfur single-wavelength anomalous diffraction crystal structure of a pheromone-binding protein from the honeybee Apis mellifera L. *The Journal of biological chemistry* **2004,** *279* (6), 4459-64.

39. Lartigue, A.; Gruez, A.; Spinelli, S.; Rivière, S.; Brossut, R.; Tegoni, M.; Cambillau, C., The crystal structure of a cockroach pheromone-binding protein suggests a new ligand binding and release mechanism *. *Journal of Biological Chemistry* **2003,** *278* (32), 30213-30218.

40. Zhang, T.; Sun, Y.; Wanner, K. W.; Coates, B. S.; He, K.; Wang, Z., Binding affinity of five PBPs to Ostrinia sex pheromones. *BMC Molecular Biology* **2017,** *18* (1), 4.

41. Atta ur, R., Studies in natural products chemistry, Volume 36 - Bioactive Natural Products. Elsevier: p 393.

42. Karlson, P.; Lüscher, M., 'Pheromones': a new term for a class of biologically active substances. *Nature* **1959,** *183* (4653), 55-56.

43. Rau, P.; Rau, N. M., *The sex attraction and rhythmic periodicity in Giant Saturnid moths*. 1929.

44. Steinbrecht, R. A., Der feinbau olfaktorischer sensillen des seidenspinners (insecta, Lepidoptera). *Zeitschrift für Zellforschung und Mikroskopische Anatomie* **1973,** *139* (4), 533-565.

45. Shiota, Y.; Sakurai, T., Chapter fourteen - silencing of OBP genes: generation of loss-of-function mutants of PBP by genome editing. In *Methods in Enzymology*, Pelosi, P.; Knoll, W., Eds. Academic Press: 2020; Vol. 642, pp 325-344.

46. 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* **2004,** *337* (2), 443-451.

47. Steinbrecht, R. A.; Ozaki, M.; Ziegelberger, G., Immunocytochemical localization of pheromone-binding protein in moth antennae. *Cell and Tissue Research* **1992,** *270* (2), 287-302.

48. Kaissling, K.; Jaenicke, L., Biochemistry of sensory functions. Berlin Heidleberg New York Springer: 1974; pp 243-273.

49. 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* **2000,** *7* (2), 143-51.

50. Horst, R.; Damberger, F.; Luginbühl, P.; Güntert, P.; Peng, G.; Nikonova, L.; Leal, W. S.; Wüthrich, K., NMR structure reveals intramolecular regulation mechanism for pheromone binding and release. *Proc Natl Acad Sci U S A* **2001,** *98* (25), 14374-14379.

51. Lee, D.; Damberger, F. F.; Peng, G.; Horst, R.; Güntert, P.; Nikonova, L.; Leal, W. S.; Wüthrich, K., NMR structure of the unliganded Bombyx mori pheromone-binding protein at physiological pH. *FEBS Letters* **2002,** *531* (2), 314-318.

52. Xu, X.; Xu, W.; Rayo, J.; Ishida, Y.; Leal, W. S.; Ames, J. B., NMR Structure of Navel Orangeworm moth pheromone-binding protein (AtraPBP1): implications for pH-sensitive pheromone detection. *Biochemistry* **2010,** *49* (7), 1469-1476.

53. 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* **2020,** *59* (37), 3411-3426.

54. 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* **2005,** *354* (5), 1081-1090.

55. 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. *J Mol Biol* **2007,** *373* (4), 811-9.

56. 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* **2013,** *8* (2), e53840.

57. 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* **2005,** *335* (4), 1044-50.

58. 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* **2000,** *7* (2), 143-151.

59. Hamiaux, C.; Carraher, C.; Löfstedt, C.; Corcoran, J. A., Crystal structure of Epiphyas postvittana pheromone binding protein 3. *Sci Rep* **2020,** *10* (1), 16366.

60. Vogt, R. G.; Riddiford, L. M., Pheromone binding and inactivation by moth antennae. *Nature* **1981,** *293* (5828), 161-163.

61. 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

Journal of Biological Chemistry **2009,** *284* (46), 32167-32177.

62. 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* **2013,** *52* (6), 1037-1044.

63. 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* **2019,** *58* (45), 4530-4542.

64. Frolov, A. N.; Bourguet, D.; Ponsard, S., Reconsidering the taxomony of several Ostrinia species in the light of reproductive isolation: a tale for Ernst Mayr. *Biological Journal of the Linnean Society* **2007,** *91* (1), 49-72.

65. Meissle, M.; Romeis, J.; Bigler, F., Bt maize and integrated pest management - a European perspective. *Pest Management Science* **2011,** *67* (9), 1049-1058.

66. Bohnenblust, E. W.; Breining, J. A.; Shaffer, J. A.; Fleischer, S. J.; Roth, G. W.; Tooker, J. F., Current European corn borer, Ostrinia nubilalis, injury levels in the northeastern United States and the value of Bt field corn. *Pest Manag Sci* **2014,** *70* (11), 1711-9.

67. Bohnenblust, E. W.; Breining, J. A.; Shaffer, J. A.; Fleischer, S. J.; Roth, G. W.; Tooker, J. F., Current European corn borer, Ostrinia nubilalis, injury levels in the northeastern United States and the value of Bt field corn. *Pest Management Science* **2014,** *70* (11), 1711-1719.

68. Killer environment. *Environ Health Perspect* **1999,** *107* (2), A62-A63.

69. Aktar, M. W.; Sengupta, D.; Chowdhury, A., Impact of pesticides use in agriculture: their benefits and hazards. *Interdiscip Toxicol* **2009,** *2* (1), 1-12.

70. Filimon, M. N.; Voia, S. O.; Popescu, R.; Dumitrescu, G.; Ciochina, L. P.; Mituletu, M.; Vlad, D. C., The effect of some insecticides on soil microorganisms based on enzymatic and bacteriological analyses. *Romanian Biotechnological Letters* **2015,** *20* (3), 10439.

71. Longnecker, M. P.; Rogan, W. J.; Lucier, G., The human health effects of DDT (dichlorodiphenyltrichloroethane) and pcbs (polychlorinated biphenyls) and an overview of organochlorines in public health. *Annual Review of Public Health* **1997,** *18* (1), 211-244.

72. Anderson, D. W., EFFECT OF DDT ON WILDLIFE. *Epidemiology* **2005,** *16* (5), S163-S164.

73. Renneberg, R.; Berkling, V.; Loroch, V., Chapter 7 - Green biotechnology. In *Biotechnology for Beginners (Second Edition)*, Renneberg, R.; Berkling, V.; Loroch, V., Eds. Academic Press: Boston, 2017; pp 233-279.

74. Schöller, M. E.; Flinn, P. W.; Grieshop, M. J.; Žd'árková, E., Chapter 9 - Biological control of stored-product pests. In *Insect Management for Food Storage and Processing (Second Edition)*, Heaps, J. W., Ed. AACC International Press: 2006; pp 67-87.

75. Razinger, J.; Vasileiadis, V. P.; Giraud, M.; van Dijk, W.; Modic, Š.; Sattin, M.; Urek, G., On-farm evaluation of inundative biological control of Ostrinia nubilalis (Lepidoptera: Crambidae) by Trichogramma brassicae (Hymenoptera: Trichogrammatidae) in three European maize-producing regions. *Pest Management Science* **2016,** *72* (2), 246-254.

76. Lidder, P.; Sonnino, A., Chapter 1 - Biotechnologies for the management of genetic resources for food and agriculture. In *Advances in Genetics*, Goodwin, S. F.; Friedmann, T.; Dunlap, J. C., Eds. Academic Press: 2012; Vol. 78, pp 1-167.

77. Abbas, M. S. T., Genetically engineered (modified) crops (Bacillus thuringiensis crops) and the world controversy on their safety. *Egyptian Journal of Biological Pest Control* **2018,** *28* (1), 52.

78. Vogt, R.; Kohne, A.; Dubnau, J.; Prestwich, G., Expression of pheromone binding proteins during antennal development in the gypsy moth Lymantria dispar. *The Journal of Neuroscience* **1989,** *9* (9), 3332-3346.

79. Kárpáti, Z.; Olsson, S.; Hansson, B. S.; Dekker, T., Inheritance of central neuroanatomy and physiology related to pheromone preference in the male European corn borer. *BMC Evol Biol* **2010,** *10*, 286-286.

80. Anton, S.; Löfstedt, C.; Hansson, B. S., Central nervous processing of sex pheromones in two strains of the European corn borer Ostrinia nubilalis (Lepidoptera: Pyralidae). *The Journal of Experimental Biology* **1997,** *200* (7), 1073.

81. Ishikawa, Y.; Takanashi, T.; Kim, C.-g.; Hoshizaki, S.; Tatsuki, S.; Huang, Y., Ostrinia spp. in Japan: their host plants and sex pheromones. *Entomologia Experimentalis et Applicata* **1999,** *91* (1), 237-244.

82. 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* **2018,** *8* (1), 17105-17105.

83. Yamaguchi, H.; Miyazaki, M., Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules* **2014,** *4* (1), 235-51.

84. Moghadam, M.; Ganji, A.; Varasteh, A.; Falak, R.; Sankian, M., Refolding process of cysteine-rich proteins:Chitinase as a model. *Rep Biochem Mol Biol* **2015,** *4* (1), 19-24.

85. Prestwich, G. D., Bacterial expression and photoaffinity labeling of a pheromone binding protein. *Protein science : a publication of the Protein Society* **1993,** *2* (3), 420-8.

86. Oldham, N. J.; Krieger, J.; Breer, H.; Svatos, A., Detection and removal of an artefact fatty acid from the binding site of recombinant Bombyx mori pheromone-binding protein. *Chemical senses* **2001,** *26* (5), 529-31.

87. Becker, W.; Bhattiprolu, K. C.; Gubensäk, N.; Zangger, K., Investigating protein-ligand interactions by solution nuclear magnetic resonance spectroscopy. *Chemphyschem* **2018,** *19* (8), 895-906.

88. Teilum, K.; Kunze, M. B. A.; Erlendsson, S.; Kragelund, B. B., (S)Pinning down protein interactions by NMR. *Protein science : a publication of the Protein Society* **2017,** *26* (3), 436-451.

89. Zhou, J.-J.; Zhang, G.-A.; Huang, W.; Birkett, M. A.; Field, L. M.; Pickett, J. A.; Pelosi, P., Revisiting the odorant-binding protein LUSH of Drosophila melanogaster: evidence for odour recognition and discrimination. *FEBS Letters* **2004,** *558* (1), 23-26.

90. Siciliano, P.; He, X. L.; Woodcock, C.; Pickett, J. A.; Field, L. M.; Birkett, M. A.; Kalinova, B.; Gomulski, L. M.; Scolari, F.; Gasperi, G.; Malacrida, A. R.; Zhou, J. J., Identification of pheromone components and their binding affinity to the odorant binding protein CcapOBP83a-2 of the Mediterranean fruit fly, Ceratitis capitata. *Insect biochemistry and molecular biology* **2014,** *48* (100), 51-62.

91. Waris, M. I.; Younas, A.; ul Qamar, M. T.; Hao, L.; Ameen, A.; Ali, S.; Abdelnabby, H. E.; Zeng, F.-F.; Wang, M.-Q., Silencing of chemosensory protein gene NlugCSP8 by RNAi induces declining behavioral responses of Nilaparvata lugens. *Frontiers in Physiology* **2018,** *9* (379).

92. Froger, A.; Hall, J. E., Transformation of plasmid DNA into E. coli using the heat shock method. *J Vis Exp* **2007,** (6), 253-253.

93. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **1995,** *6* (3), 277-93.

94. Lee, W.; Tonelli, M.; Markley, J. L., NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2015,** *31* (8), 1325-7.

95. Bahnson, B. J.; Park, D. H.; Kim, K.; Plapp, B. V.; Klinman, J. P., Unmasking of hydrogen tunneling in the horse liver alcohol dehydrogenase reaction by site-directed mutagenesis. *Biochemistry* **1993,** *32* (21), 5503-7.

96. 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* **2006,** *15* (12), 2795-2804.

97. 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* **2009,** *44* (4), 213-223.

98. Güntert, P., Automated NMR protein structure calculation. *Progress in nuclear magnetic resonance spectroscopy* **2003,** *43* (3-4), 105-125.

99. Trainor, K.; Palumbo, J. A.; MacKenzie, D. W. S.; Meiering, E. M., Temperature dependence of NMR chemical shifts: Tracking and statistical analysis. *Protein science : a publication of the Protein Society* **2020,** *29* (1), 306-314.

100. Sharma, D.; Rajarathnam, K., 13C NMR chemical shifts can predict disulfide bond formation. *Journal of Biomolecular NMR* **2000,** *18* (2), 165-171.

101. Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D., Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Science* **2006,** *15* (12), 2795-2804.

102. Havel, T. F.; Wüthrich, K., An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *J Mol Biol* **1985,** *182* (2), 281-94.

103. Roberts, G. C. K., Protein NMR – introduction. In *Encyclopedia of Biophysics*, Roberts, G. C. K., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 2027-2033.

104. Gottstein, D.; Kirchner, D. K.; Güntert, P., Simultaneous single-structure and bundle representation of protein NMR structures in torsion angle space. *Journal of Biomolecular NMR* **2012,** *52* (4), 351-364.

105. Michel, E.; Damberger, F. F.; Ishida, Y.; Fiorito, F.; Lee, D.; Leal, W. S.; Wüthrich, K., Dynamic conformational equilibria in the physiological function of the Bombyx mori pheromone-binding protein. *Journal of Molecular Biology* **2011,** *408* (5), 922-931.

106. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000,** *28* (1), 235-242.

107. McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins. *Nature* **1977,** *267* (5612), 585-590.

108. Purmonen, M.; Valjakka, J.; Takkinen, K.; Laitinen, T.; Rouvinen, J., Molecular dynamics studies on the thermostability of family 11 xylanases. *Protein Engineering, Design and Selection* **2007,** *20* (11), 551-559.

109. Goossens, K.; De Winter, H., Molecular dynamics simulations of membrane proteins: an overview. *Journal of Chemical Information and Modeling* **2018,** *58* (11), 2193-2202.

110. Guterres, H.; Im, W., Improving protein-ligand docking results with high-throughput molecular dynamics simulations. *Journal of Chemical Information and Modeling* **2020,** *60* (4), 2189-2198.

111. Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E., Atomic-level characterization of protein–protein association. *Proceedings of the National Academy of Sciences* **2019,** *116* (10), 4244-4249.

112. Yoo, J.; Winogradoff, D.; Aksimentiev, A., Molecular dynamics simulations of DNA–DNA and DNA–protein interactions. *Current Opinion in Structural Biology* **2020,** *64*, 88-96.

113. Krieger, E.; Joo, K.; Lee, J.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K., Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **2009,** *77 Suppl 9* (Suppl 9), 114-22.

114. 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* **1993,** *26* (2), 283-291.

115. Bhattacharya, A.; Tejero, R.; Montelione, G. T., Evaluating protein structures determined by structural genomics consortia. *Proteins* **2007,** *66* (4), 778-95.

116. Smith, J. A., MOLMOL: A free biomolecular graphics/analysis package. *Genome Biology* **2000,** *1* (2), reports2046.

117. Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J., CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res* **2018,** *46* (W1), W363-w367.

118. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **2004,** *25* (13), 1605-12.

119. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995,** *91* (1), 43-56.

120. Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F., Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal* **2011,** *40* (7), 843.

121. Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E., An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *Journal of Chemical Theory and Computation* **2011,** *7* (12), 4026-4037.

122. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M., LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering, Design and Selection* **1995,** *8* (2), 127-134.

123. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *Journal of Molecular Graphics* **1996,** *14* (1), 33-38.

124. Kabsch, W.; Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22* (12), 2577-2637.

125. Joosten, R. P.; te Beek, T. A. H.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W. W.; Schneider, R.; Sander, C.; Vriend, G., A series of PDB related databases for everyday needs. *Nucleic Acids Res* **2011,** *39* (Database issue), D411-D419.

126. Kumari, R.; Kumar, R.; Lynn, A., g_mmpbsa—A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling* **2014,** *54* (7), 1951-1962.

127. Honig, B.; Nicholls, A., Classical electrostatics in biology and chemistry. *Science* **1995,** *268* (5214), 1144-9.

128. Laskowski, R. A.; Furnham, N.; Thornton, J. M., The Ramachandran plot and protein structure validation In *Biomolecular Forms and Functions*, WORLD SCIENTIFIC / INDIAN INST OF SCIENCE, INDIA: 2012; pp 62-75.

129. Doig, A. J.; Baldwin, R. L., N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. *Protein science : a publication of the Protein Society* **1995,** *4* (7), 1325-36.

130. Serrano, L.; Fersht, A. R., Capping and α-helix stability. *Nature* **1989,** *342* (6247), 296-299.

131. Santos, L. H. S.; Ferreira, R. S.; Caffarena, E. R., Integrating Molecular Docking and Molecular Dynamics Simulations. In *Docking Screens for Drug Discovery*, de Azevedo Jr, W. F., Ed. Springer New York: New York, NY, 2019; pp 13-34.

132. Dannenberg, J. J., An Introduction to hydrogen bonding By George A. Jeffrey (University of Pittsburgh). Oxford University Press: New York and Oxford. 1997. ix + 303 pp. $60.00. ISBN 0-19-509549-9. *Journal of the American Chemical Society* **1998,** *120* (22), 5604-5604.

133. Tokuriki, N.; Tawfik, D. S., Protein Dynamism and Evolvability. *Science* **2009,** *324* (5924), 203.

134. Skopalík, J.; Anzenbacher, P.; Otyepka, M., Flexibility of Human Cytochromes P450: Molecular Dynamics Reveals Differences between CYPs 3A4, 2C9, and 2A6, which Correlate with Their Substrate Preferences. *The Journal of Physical Chemistry B* **2008,** *112* (27), 8165-8173.

135. Münz, M.; Hein, J.; Biggin, P. C., The Role of Flexibility and Conformational Selection in the Binding Promiscuity of PDZ Domains. *PLOS Computational Biology* **2012,** *8* (11), e1002749.

136. da Silva, L. C. A.; Honorato, T. L.; Cavalcante, R. S.; Franco, T. T.; Rodrigues, S., Effect of pH and temperature on enzyme activity of Chitosanase produced under solid stated fermentation by Trichoderma spp. *Indian J Microbiol* **2012,** *52* (1), 60-65.

137. Robinson, P. K., Enzymes: principles and biotechnological applications. *Essays Biochem* **2015,** *59*, 1-41.

138. Zeitlinger, M. A.; Derendorf, H.; Mouton, J. W.; Cars, O.; Craig, W. A.; Andes, D.; Theuretzbacher, U., Protein binding: do we ever learn? *Antimicrobial Agents and Chemotherapy* **2011,** *55* (7), 3067.

139. Ruano, M. L. F.; Pérez-Gil, J.; Casals, C., Effect of acidic pH on the structure and lipid binding properties of porcine surfactant protein A: potential role of acidification along its exocytic pathway *. *Journal of Biological Chemistry* **1998,** *273* (24), 15183-15191.

140. Shaw, K. L.; Grimsley, G. R.; Yakovlev, G. I.; Makarov, A. A.; Pace, C. N., The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein science : a publication of the Protein Society* **2001,** *10* (6), 1206-1215.

141. Schimmel, C. a., Biophysical chemistry - part I: the conformation of biological macromolecules. **1980**.

142. Yu, Y.; Ma, F.; Cao, Y.; Zhang, J.; Zhang, Y.; Duan, S.; Wei, Y.; Zhu, S.; Chen, N., Structural and functional difference of pheromone binding proteins in discriminating chemicals in the gypsy moth, Lymantria dispar. *Int J Biol Sci* **2012,** *8* (7), 979-991.

143. Han, L.; Zhang, Y.-J.; Zhang, L.; Cui, X.; Yu, J.; Zhang, Z.; Liu, M. S., Operating mechanism and molecular dynamics of pheromone-binding protein ASP1 as influenced by pH. *PLOS ONE* **2014,** *9* (10), e110565.

144. Wang, H.; Ma, Y.-F.; Wang, M.-M.; Chen, G.-L.; Dewer, Y.; He, M.; Zhang, F.; Yang, Y.-F.; Liu, J.-F.; He, P., Expression, affinity, and functional characterization of the specific binding of two putative pheromone-binding proteins in the Omnivorous German Cockroach Blattella germanica. *Journal of Agricultural and Food Chemistry* **2020,** *68* (47), 13573-13583.

145. Keil, T. A., Surface coats of pore tubules and olfactory sensory dendrites of a silkmoth revealed by cationic markers. *Tissue and Cell* **1984,** *16* (5), 705-717.

146. Sokolova, A.; Kealley, C. S.; Hanley, T.; Rekas, A.; Gilbert, E. P., Small-angle X-ray scattering study of the effect of pH and salts on 11S soy glycinin in the freeze-dried powder and solution states. *Journal of Agricultural and Food Chemistry* **2010,** *58* (2), 967-974.

147. 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. *Analytical Biochemistry* **2000,** *287* (2), 252-260.

148. Sharma, G. S.; Bhattacharya, R.; Singh, L. R., Chapter 11 - Protein covalent modification by homocysteine: consequences and clinical implications. In *Protein Modificomics*, Dar, T. A.; Singh, L. R., Eds. Academic Press: 2019; pp 281-311.

149. Goto, Y.; Nishikiori, S., Conformational stability of the molten globule of cytochrome c : role of electrostatic repulsion. In *Techniques in Protein Chemistry III*, Angeletti, R. H., Ed. Academic Press: 1992; pp 337-346.

150. Ptitsyn, O. B., Molten globule and protein folding. In *Advances in Protein Chemistry*, Anfinsen, C. B.; Richards, F. M.; Edsall, J. T.; Eisenberg, D. S., Eds. Academic Press: 1995; Vol. 47, pp 83-229.

151. Price, N. E.; Price, N. C.; Kelly, S. M.; McDonnell, J. M., The key role of protein flexibility in modulating IgE interactions *. *Journal of Biological Chemistry* **2005,** *280* (3), 2324-2330.

152. Consalvi, V.; Chiaraluce, R.; Giangiacomo, L.; Scandurra, R.; Christova, P.; Karshikoff, A.; Knapp, S.; Ladenstein, R., Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile Thermotoga maritima. *Protein Engineering, Design and Selection* **2000,** *13* (7), 501-507.

153. 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* **2013,** *110* (46), 18680.

154. Dyson, H. J.; Wright, P. E., Equilibrium NMR studies of unfolded and partially folded proteins. *Nature Structural Biology* **1998,** *5* (7), 499-503.

155. Cinelli, S.; Spinozzi, F.; Itri, R.; Finet, S.; Carsughi, F.; Onori, G.; Mariani, P., Structural characterization of the pH-denatured states of Ferricytochrome-c by synchrotron small Angle X-ray scattering. *Biophysical Journal* **2001,** *81* (6), 3522-3533.

156. Gelamo, E. L.; Itri, R.; Tabak, M., Small angle X-ray scattering (SAXS) study of the extracellular hemoglobin of Glossoscolex paulistus: effect of ph, iron oxidation state, and interaction with anionic sds surfactant *. *Journal of Biological Chemistry* **2004,** *279* (32), 33298-33305.

157. Lipfert, J.; Doniach, S., Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annual Review of Biophysics and Biomolecular Structure* **2007,** *36* (1), 307-327.

158. Mertens, H. D.; Svergun, D. I., Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J Struct Biol* **2010,** *172* (1), 128-41.

159. Franke, D.; Svergun, D. I., Synchrotron small-angle X-ray scattering on biological macromolecules in solution. In *Synchrotron Light Sources and Free-Electron Lasers: Accelerator Physics, Instrumentation and Science Applications*, Jaeschke, E. J.; Khan, S.; Schneider, J. R.; Hastings, J. B., Eds. Springer International Publishing: Cham, 2020; pp 1645-1672.

160. Brennich, M.; Pernot, P.; Round, A., How to analyze and present SAS data for publication. In *Biological Small Angle Scattering: Techniques, Strategies and Tips*, Chaudhuri, B.; Muñoz, I. G.; Qian, S.; Urban, V. S., Eds. Springer Singapore: Singapore, 2017; pp 47-64.

161. Jacques, D. A.; Trewhella, J., Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. *Protein Science* **2010,** *19* (4), 642-657.

162. Hopkins, J. B.; Gillilan, R. E.; Skou, S., BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J Appl Crystallogr* **2017,** *50* (Pt 5), 1545-1553.

163. Semenyuk, A. V.; Svergun, D. I., GNOM– a program package for small-angle scattering data processing. *Journal of Applied Crystallography* **1991,** *24* (5), 537-540.

164. Franke, D.; Svergun, D. I., DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* **2009,** *42* (Pt 2), 342-346.

165. 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* **1995,** *28* (6), 768-773.

166. Franke, D.; Jeffries, C. M.; Svergun, D. I., Correlation map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nature Methods* **2015,** *12* (5), 419-422.

167. Salamon, D., Chapter 6 - Advanced ceramics. In *Advanced Ceramics for Dentistry*, Shen, J. Z.; Kosmač, T., Eds. Butterworth-Heinemann: Oxford, 2014; pp 103-122.

168. 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* **2010,** *107* (28), 12535.

APPENDICES

Appendix Table A-1 Backbone chemical shift assignments of OnubPBP3

Residue #	Amino acid	HN	N	CA	CB	CO
2	Gln			57.787	28.065	177.829
3	Thr	7.849	114.889	65.651	67.942	175.512
4	Val	7.387	123.432	65.262	31.300	177.397
5	Met	8.086	118.446	55.778	29.558	179.481
6	Arg	8.667	124.128	60.057	29.357	176.415
7	Glu	7.944	120.741	58.607	29.081	179.452
8	Met	9.106	119.486	59.971	33.971	176.718
9	Thr	8.169	115.127	67.442	68.385	177.927
10	Arg	8.519	120.555	59.459	30.326	176.362
11	Asn	7.329	113.398	55.968	39.425	175.293
12	Phe	8.548	175.293	60.538	40.432	175.354
13	Ile	8.222	112.675	62.410	37.173	176.815
14	Lys	7.359	121.527	58.676	31.811	176.606
15	Ala	7.326	118.931	52.148	17.694	174.957
16	Tyr	8.208	120.056	63.033	37.923	175.429
17	Glu	8.767	118.700	60.065	28.608	178.004
18	Val	7.842	116.640	64.759	30.849	177.175
19	Cys	7.171	121.007	59.717	41.125	174.551
20	Ala	9.105	123.940	54.511	16.680	179.175
21	Lys	7.337	116.380	57.848	31.873	178.022
22	Glu	8.020	119.718	59.063	29.740	176.894
23	Tyr	8.071	113.456	57.818	37.371	173.370
24	Asn	7.579	119.625	53.521	36.547	173.957
25	Leu	8.208	117.514	52.670	40.019	174.249
26	Pro			61.945	31.703	177.017
27	Glu	9.061	124.503	59.177	29.013	177.765
28	Ala	8.565	119.622	53.865	18.173	177.964
29	Thr	7.935	114.917	66.065	67.432	175.020
30	Gly	7.448	107.677	45.117		174.713
31	Ser	7.535	115.677	60.616	62.380	174.745
32	Glu	7.584	121.718	59.006	29.649	177.463
33	Leu	7.596	116.244	56.944	41.959	178.055
34	Ile	8.066	116.513	63.677	36.994	174.671
35	Asn	6.690	115.030	52.713	37.904	173.240
36	Phe	7.673	118.544	62.855	40.245	174.901

37	Trp	9.487	113.964	56.515	28.931	174.329
38	Lys	7.927	124.321	56.948	32.056	177.071
39	Glu	9.389	132.606	57.855	28.601	176.624
40	Gly	9.072	114.076	44.876		172.459
41	His	7.450	119.821	55.746	31.327	172.784
42	Glu	7.795	124.138	53.886	31.693	173.631
43	Leu	7.486	122.560	53.999	41.774	175.361
44	Thr	9.409	111.739	61.656	70.998	174.245
45	Thr	8.749	115.955	60.206	70.752	173.223
46	Arg	8.876	123.339	59.293	29.308	177.171
47	Glu	9.172	116.766	61.375	28.256	176.989
48	Ala	8.033	122.067	55.115	16.889	178.808
49	Gly	7.654	105.573	48.099		174.738
50	Cys	9.234	119.822	54.624	34.638	176.776
51	Ala	8.709	124.010	55.767	17.716	177.205
52	Ile	8.119	118.990	65.318	37.300	177.701
53	Leu	8.450	124.468	58.026	41.803	177.508
54	Cys	8.355	120.495	59.532	40.209	175.115
55	Met	9.232	123.770	59.435	33.919	175.928
56	Ser	8.156	112.897	62.460		174.539
57	Thr	7.842	118.412	65.916	68.082	177.618
58	Lys	8.646	123.631	57.268	30.783	177.333
59	Leu	7.159	117.460	53.926	41.590	174.014
60	Asn	7.907	114.765	53.588	36.584	173.589
61	Leu	7.988	111.869	54.927	40.406	174.889
62	Leu	7.267	114.079	52.038	43.184	176.800
63	Asp	9.492	122.642	51.499	41.388	177.195
64	Val	8.126	115.487	64.158	31.293	175.823
65	Gln	8.004	118.357	55.110	28.426	175.372
66	Gly	8.172	108.667	45.115		172.594
67	Ser	8.271	117.755	56.384	63.466	173.105
68	Val	9.272	125.547	64.641	31.379	175.198
69	His	9.707	133.136	56.415	31.314	174.978
70	Arg	9.275	131.800	60.004	29.684	175.477
71	Gly	8.388	106.827	47.341		176.153
72	Asn	11.415	124.079	54.906	36.633	178.384
73	Thr	8.497	121.424	67.588	66.296	175.167
74	Val	8.884	123.799	66.484	30.306	176.671
75	Glu	7.790	118.103	59.258	28.896	177.396
76	Phe	7.868	121.081	60.714	37.550	175.662
77	Ala	8.614	121.900	55.229	16.380	180.094
78	Lys	8.781	118.757	58.973	30.905	179.733
79	His	8.392	120.617	57.383	27.791	174.465
80	His	7.088	114.978	55.791	29.354	172.541
81	Gly	7.350	103.744	44.674		173.876
82	Ser	8.033	116.690	59.079	64.269	173.487
83	Asp	8.088	122.460	52.063	40.673	176.312
84	Asp	8.597	120.556	58.403	41.088	176.127
85	Ala	8.118	120.890	54.899	17.427	180.081

86	Met	8.734	119.871	58.858	30.640	176.306
87	Ala	7.618	120.409	55.164	19.566	177.848
88	His	8.189	113.860	60.154	30.308	176.908
89	Gln	8.132	121.299	59.022	27.944	177.503
90	Val	8.600	118.934	67.364	30.885	176.268
91	Val	7.402	117.430	66.801	30.426	175.562
92	Asp	8.821	121.702	57.578	39.268	179.156
93	Ile	8.441	121.218	64.785	38.569	177.269
94	Leu	8.103	120.826	58.734	41.206	177.854
95	His	9.264	117.667	57.438	29.827	177.031
96	Ala	8.260	123.102	54.897	17.262	180.539
97	Cys	8.787	117.963	56.976	39.149	175.731
98	Glu	8.542	121.871	59.021	29.581	176.466
99	Lys	7.472	115.641	56.679	32.011	176.360
100	Ala	7.430	119.809	53.661	19.144	177.295
101	Thr	7.224	108.520	57.370	69.820	170.745
102	Pro			62.575	31.496	175.378
103	Asn	8.190	119.981	53.249	41.301	173.565
104	Glu	8.709	124.505	57.340	29.617	175.286
105	Asp	8.619	119.569	52.157	40.687	175.857
106	Lys	8.800	124.695	57.639	31.116	178.672
107	Cys	8.111	117.129	60.953	44.937	174.943
108	Met	7.536	118.428	55.270	29.227	178.687
109	Leu	8.686	123.050	57.376	41.605	176.657
110	Ala	7.477	120.161	55.244	18.141	178.024
111	Leu	7.440	117.125	58.569	42.583	177.100
112	Ser	8.169	114.427	61.482	62.548	174.491
113	Ile	9.004	123.322	65.684	37.453	176.341
114	Ala	8.605	121.869	55.733	17.878	178.502
115	Met	8.228	114.716	56.057	30.172	177.946
116	Cys	8.228	124.572	59.148	39.156	174.362
117	Phe	9.381	123.679	60.617	39.501	175.179
118	Lys	8.750	119.177	59.170	33.068	176.812
119	Ala	7.568	118.700	54.902	17.671	180.288
120	Glu	7.937	117.808	57.963	29.293	177.794
121	Ile	8.521	120.847	60.826	33.717	178.693
122	His	8.644	119.629	60.377	28.145	178.341
123	Lys	7.676	120.851	59.273	31.831	176.985
124	Leu	6.998	118.281	54.465	41.179	174.319
125	Asp	7.919	114.921	54.342	39.368	175.959
126	Trp	7.560	115.867	53.196	30.496	172.963
127	Ala	7.375	122.638	48.511	17.944	173.950
128	Pro			59.253	28.591	177.785
129	Asn	7.777	117.971	55.696	41.670	176.874
130	His	7.749	116.930	56.248	34.713	175.551
131	Glu	8.077	121.136	54.656		175.874
132	Leu	7.881	118.660	54.529	41.696	175.872
133	Met	7.885	118.625	55.273	32.243	174.466
134	Phe	8.033	121.086	56.658	38.502	174.004

135	Glu	8.092	120.582	56.306	31.700	172.968
136	Glu	7.618	124.930	53.763	31.784	173.500
137	Leu	8.276	122.503	54.778	41.875	175.828
138	Val	4.162	122.476	61.789	32.586	174.959
139	Ser	8.250	118.987	57.677	63.369	173.285
140	Asp	8.252	122.469	55.586	40.787	176.195
141	Met	8.047	118.002	57.632	32.913	175.660
142	Trp	7.656	116.724	56.290	29.501	175.069
143	Asn	8.259	119.931	53.009	38.808	173.299
144	Ser	7.896	121.146	59.627	64.470	177.633

VITA

Omar I. Al-Danoon

Candidate for the Degree of

Doctor of Philosophy

Dissertation: OVEREXPRESSION, CHARACTERIZATION, AND STRUCTURE-FUNCTION STUDIES OF PHEROMONE BINDING PROTEINS FROM EUROPEAN CORN BORER, OSTRINIA NUBILALIS

Major Field: Chemistry

Biographical:

Education:

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

Completed the requirements for the Master of Science in Chemistry at Florida Institute of Technology, Melbourne, Florida in 2014.

Completed the requirements for the Bachelor of Science in Chemistry at University of Mosul, Mosul, Iraq in 2007.

Experience:

Advanced NMR spectroscopy
Advanced Atomic Force Microscopy techniques
Fluorescence spectroscopy
Circular dichroism spectroscopy
Protein expression, refolding, and purification
Small angle X-ray scattering
Protein structure calculation from NMR data
Molecular dynamics simulations (protein)

Professional Memberships:

A staff member at Chemistry department, University of Mosul, Iraq