STRUCTURE AND DYNAMICS STUDIES OF MEMBRANE AND NON-MEMBRANE PROTEINS USING NMR AND MD SIMULATIONS

By

HEM BAHADUR MOKTAN TAMANG

Master's of Science in Physics

Tribhuvan University, Department of Physics

Lalitpur, Nepal

2009

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
December, 2017

STRUCTURE AND DYNAMICS STUDIES OF MEMBRANE AND NON-MEMBRANE PROTEINS USING NMR AND MD SIMULATIONS

Dissertation Approved:

Dr. Donghua H. Zhou (Dissertation Adviser)

Dr. Aihua Xie

Dr. Mario Borunda

Dr. Wouter D. Hoff

ACKNOWLEDGEMENTS

I would like to express my deep gratitude toward

My advisor, Dr. Donghua H. Zhou, who shared with me his research experience and provoked me to good work. I would like to thank him for his superior advisory role, support, guidance, advice and encouragement throughout my PhD career.

My advisory committee members Dr. Aihua Xie, Dr. Mario Borunda and Dr. Wouter Hoff for their help and support of many kinds.

Mom, Dad, and my entire family members who continuously supported me emotionally and financially.

My wife and beloved daughter bringing joy in my life.

My colleagues, Penghui Lin, Lian Duan, Amanda Taylor, Maryam Mashayekhi, and Dependra Bhandari who made the journey fun and enjoyable.

Last but not the least, the entire Physics department family, Oklahoma State University.

Name: HEM BAHADUR MOKTAN TAMANG

Date of Degree: DECEMBER, 2017

Title of Study: STRUCTURE AND DYNAMICS STUDIES OF MEMBRANE AND NON-MEMBRANE PROTEINS USING NMR AND MD SIMULATIONS

Major Field: PHYSICS

Abstract:

The protein structural knowledge is essential in defining molecular recognition rules that power the understanding of basic biological phenomenon. The structures of most proteins are determinable due to advancement in technology and method development. Nuclear Magnetic Resonance (NMR) is one of the most versatile tools designed for this purpose. Proteins are flexible entities and dynamics play key role in their functionality therefore structures alone may provide only partial view on their functions. The experimental techniques have been used to study protein thermodynamics, but computer simulations have evolved to become the most convenient way to obtain the complete picture of protein dynamics. The central aim of this research is to study the structure and DNA binding dynamics of homologous pairing protein 2 (HOP2). In the first phase, the structure of N–terminal domain of HOP2 was investigated using NMR. It was identified with winged–helix DNA–recognition structural motifs. Furthermore, the DNA binding properties of this protein was investigated by NMR chemical shift perturbation method. It was found to bind to double–stranded DNA with considerable affinity, where structural motifs helix 3 (H3) and wing 1 (W1) were responsible for DNA recognition. Additionally, the site directed mutagenesis studies suggested H3 as the major contributor in DNA recognition. In the second phase, the DNA binding dynamics of HOP2 was investigated using classical MD simulations. Complexes of protein HOP2 and its mutants with DNA were constructed and then simulated using software GROMACS. Simulation results revealed the atomic level interactions between HOP2 and DNA, where H3 and W1 motifs engaged with DNA at major and minor grooves respectively. The effects on DNA binding due to point mutations in W1 and H3 were also observed. These effects were accessed in terms of changes in complex stability, binding free energy, and total number of interactions. The simulation results we obtained suggested that the motif W1 is also important as H3 in DNA binding. The NMR experimental and simulation protocol designed in this work will be useful in studying structure and dynamics of protein–protein or protein–ligand systems.

TABLE OF CONTENTS

Chapter Page #

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The central dogma of molecular biology states that genetic information is stored in DNA, which is a linear sequence of four nucleotides. The genetic information is transcribed into RNA that in turn is translated into proteins (1). Proteins perform complex and diverse sets of functions in cells and are involved in almost all vital processes like hearing, vision, smell, metabolism, immune response, and cell division etc. Some of these macromolecules are very small and catalyze chemical reactions and some are parts of large molecular machines, or complexes, capable of transcribing DNA to mRNA and translating them into proteins. Proteins can also polymerize into fibers that build up the cytoskeleton. Additionally, they efficiently perform their functions in a crowded environment of cell. Therefore, the knowledge of such a versatile set of chemicals can provide massive information about biological phenomena. Interestingly, the general knowledge of protein functions comes from its three dimensional (3D) structure so the study of its structural architecture would be the natural next step in understanding its role in biological phenomenon.

Proteins are composed up of several building blocks called amino acids put together by a large molecule, the ribosome (1). As the amino–acid chain or nascent protein also known as primary structure emerges from the ribosome, it rapidly folds to an energy minimum, the specific structure referred to as native state or native fold or tertiary structure (2). This native state of protein is a three–dimensional structure where the structural elements (secondary structures) are arranged in thermodynamically favorable conformations. Although the folding pathway is also critical in understanding protein in general, it is not the scope of this work. Once the protein is folded into its native form, the overall conformation of the secondary structural elements can be extremely useful in characterizing its functions. After the fold, proteins may find the physical association among similar or other different proteins to form stoichiometrically stable complex. This complex structure, also known as quaternary structure acts as a machine to carry out specific cellular function. At the next level, these complexes interact with individual proteins or other complexes to form functional modules and pathways that carry out most cellular processes. Even the limited numbers of proteins are capable of carrying out many kinds of cellular functions through this hierarchical structure.

Currently, many techniques capable of obtaining high–resolution protein structures are available due to advances in technology and method development. Among them, X-ray crystallography is one of the leading techniques in structural biology but the requirement of specific sample condition makes it unsuitable for many protein systems. Another technique, Nuclear magnetic resonance (NMR), has been very versatile in determining structural information of proteins in variety of sample conditions. As for the smaller soluble proteins, NMR has distinct edge due to its robustness and accuracy in elucidating high–resolution structures. This method is based on the resonance of nuclei to a sequence of radio frequencies that can be detected as an electrical signal. The resonance frequencies can be analyzed and used to determine the local conformations of atoms that are close to each other. For protein samples in solution, NMR exploits the tumbling motion of protein. The faster tumbling of smaller proteins produce non–overlapping NMR signals that are extremely useful in generating high–resolution structures.

The solution NMR can be insufficient in determining the high–resolution structures of larger proteins where tumbling motion of protein in solution is relatively insignificant. Additionally for membrane proteins, getting solution sample is tricky and sample needs to be prepared in detergent solution. The choice of detergent is critical because detergents can partially or totally denature the protein. Moreover, the generation of real membrane environment for such proteins is not possible with detergents. The structural fold of proteins can be different when they are in real membrane. The complexity of membrane proteins and the shear size of the system make solution NMR not ideal for studying such proteins. However, the spectral resolution of solid–state NMR does not depend on the size of the system so is an excellent choice for membrane protein systems. These experimental techniques have their strengths and weaknesses, but together they have taken study of biomolecular structure to a complete new height. The protein structures calculated by these methods have dominated the protein data bank till this date.

The 3–D structure of proteins have certainly enhanced our understanding of their functions in biological processes. However, protein molecules are dynamic machines that explore complex energy landscapes and several conformations while performing their functions. Also, proteins (and nucleic acids) are flexible entities, and dynamics play a key role in their functionality so the structures alone provide only a partial view on their functions. X–ray crystallography and NMR have been used to generate protein conformational ensemble, but these are certainly not robust for protein dynamics studies. Researchers have used several other experimental techniques as well but theoretical techniques have evolved to become the most convenient way to obtain the picture of protein dynamics. The advanced simulation algorithms with appropriate force–fields, including specific strategies to increase the conformational sampling are strikingly successful in generating the conformational ensemble. If properly built, ensembles can be utilized to derive accurate thermodynamic properties of a system and to reconstruct complex conformational transitions or even folding events.

The aim of this research is to use Nuclear Magnetic Resonance (NMR) to investigate the structure of homologous pairing protein 2 (HOP2) and use molecular dynamics (MD) simulations to study its DNA binding dynamics. Together with these methods, we try to understand the functional role of this protein in homologous recombination process, which is critical in meiosis cell division. Additionally, the implementation of MD simulations for generating structural models of membrane proteins and verification by solid–state NMR in two separate projects will also be discussed.

1.1 Biological background

1.1.1 Meiosis and homologous recombination

Sexually reproducing organisms rely on strongly conserved and complex developmental process that generates haploid progeny from diploid precursors (3). In 1905, Farmer and Moore termed this specialized form of cell division meiosis (derived from the Greek word **μειωσις**, meaning "reduction") during which chromosome numbers reduce precisely into half the original (23–pair). In this process, reduction in chromosome number in gametes is achieved in different intermediate phases, one round of DNA replication followed by two rounds of chromosome segregations (4). This serves a fundamental function by halving the number of chromosomes, which is required for sexual reproduction. Prophase–I in meiosis–I is considered as a unique event as chromosomes find their homolog and pair up together. The arrangement of maternal and paternal homologous chromosomes into pairs allows them to act as a single unit when microtubules attach and align them on the meiotic spindle. This ensures their orderly segregation to opposite poles of the cell at the first meiotic cell division so that each gamete receives only one copy of each chromosome. In most organisms homologous chromosomes pairs are stabilized by a physical link provided by the crossovers, the products of homologous recombination (HR), which are viewed cytologically as chiasmata. Pairing of homologous chromosomes to form bivalents is an essential feature of meiosis that promotes high levels of genetic recombination and ensures accurate homolog segregation (5, 6). The genome–wide homology search and linkage among the homologs is achieved by rather complicated homologous recombination (HR) mechanism (7). Consequently, mutations that cause loss or misregulation of recombination are invariably associated with increased errors in meiotic chromosome segregation and the generation of aneuploid gametes. Besides its meiotic role, homologous recombination is also a crucial process for DNA damage repair and genetic diversification (8).

Although HR varies widely among different organisms and cell types, most forms are initiated by the formation and processing of DNA double strand break (DSB) that is introduced by protein Spol1 (contains tyrosine which ligates and dissociates with DNA to promote break formation) (9) which is immediately followed by resection (sections of DNA around 5' ends of the break are cut away). DSBs in chromosomal DNA also occur when cells are exposed to various DNA damaging agents like ionizing radiations, cross–linking reagents and oxidative stress (10). Processing the DNA ends of DSBs by exonucleases generates 3' single–stranded DNA tails. Two recombinases, RAD51 and DMC1, then bind to 3' single–stranded tails, which are called nucleoprotein filaments. These filaments invade the intact DNA in search for homologous pair. After invasion, events may follow either of different pathways, mainly double–strand break repair (DSBR) or synthesis–dependent strand annealing (SDSA) (11). The DSBR pathway commonly results in crossover, though there is certain probability of non–crossover; enhancing the genetic diversity. The DSBR pathway is the likely model of crossover HR during meiosis (8).

Homologous Recombination is a universal biological mechanism and its dysfunction has been strongly associated with increased susceptibility to several types of cancer. Also studies have confirmed that the defects in recombination and meiosis cause aneuploidy syndromes such as Downs (1) and azoospermia in men (8). So, the study of proteins that take part in this process is necessary to understand this critical and crucial stage of cell division. Among many proteins that take part in this process, RAD51 and DMC1 recombinases are the major players (12). Efficient action of Dmc1 and Rad51 however, requires the assistance of auxiliary proteins such as Homologous Pairing Protein 2 (HOP2) and MND1 (13-15). Indeed, a heterocomplex formed by HOP2 and MND1 physically interacts with DMC1 and RAD51 to stimulate DNA strand exchange promoted by these proteins, which is required for successful progression of homologous recombination.

1.1.2 HOP2 protein

The homologous–pairing protein 2 (HOP2) is a 217AA protein that belongs to HOP2 family, which is highly expressed in testis and colon and plays an important role in meiotic homologous recombination. Several studies have shown that HOP2 protein is essential for efficient double–strand break repair and normal progression of meiotic cell division (12). It has also been shown *in Vitro* that HOP2 stimulates DMC1–mediated strand exchange necessary for homologous pairing and HOP2 alone is proficient in promoting strand invasion (12, 16). The recombination can proceed normally up to the stage when Rad51 and DMC1 are loaded on the ends of DNA DSBs but without HOP2 and/or MND1, the further progression is debilitated (7). Also, HOP2 in complex with MND1 plays significant role for the alignment of homologous chromosomes and galvanize the recombinase activity of DMC1 (17). Experimental observations indicate that the HOP2–MND1 complex binds DMC1 with a significantly increased affinity with respect to isolated proteins and they work together at the same step during recombination (7). Hence it is evident that the formation of HOP2–MND1 heterodimer results in a new interface which is responsible for the interaction and stimulation of DMC1 (15). According to recent studies, HOP2 in complex with MND1 take part in two critical functions: the synaptic complex formation promoted by DMC1 requires HOP2–MND1 complex and stimulation of DMC1–ssDNA filament to capture duplex DNA (17).

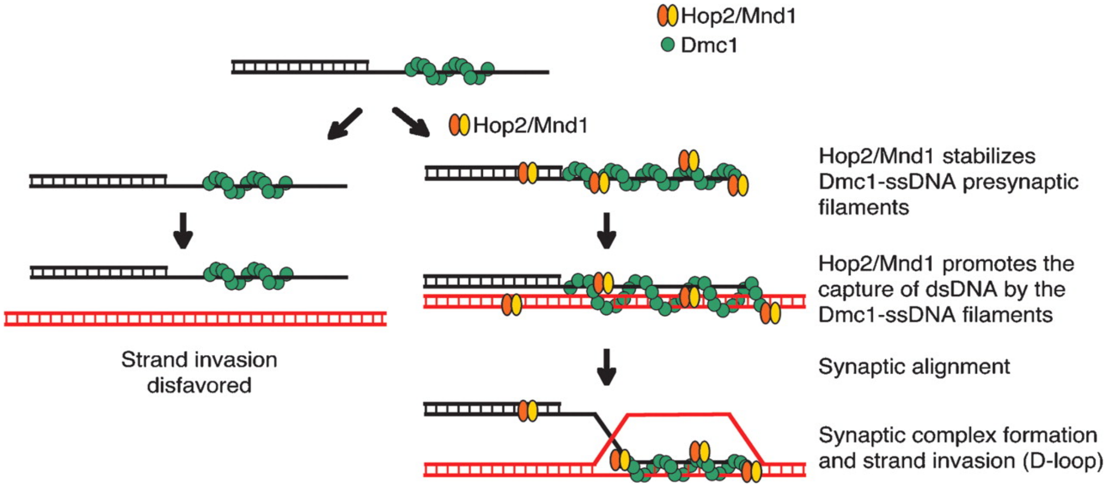


Figure 1.1 A crude model explains the functional role of HOP2 during initial phases of homologous recombination (16).

1.2 Overview of the thesis

Following is an overview of the rest of the chapters in this thesis:

Chapter II: NUCLEAR MAGNETIC RESONANCE (NMR)

Nuclear magnetic resonance (NMR) is one of the most versatile methods in determining the protein structure. We chose this method due to its versatility and the freedom in the choice of sample conditions. The basic concepts and theories pertaining to this method are discussed in this chapter. We used solution NMR to study the structure of HOP2 and our method takes the following input:

a. The protein of interest (N-terminal HOP2), which is relatively small 85 amino–acid long (~9.6 kDa) showed to exist in monomer conformation in the given buffer conditions.

b. Evidence of folded state with compact geometry.

Chapter III: MOLECULAR DYNAMIC (MD) SIMULATIONS

Molecular dynamic simulations have evolved to become one of the most convenient methods to study the dynamics of bio–molecular system. Although NMR is able to provide thermodynamic properties of the system to some extent, the study of atomic resolution dynamic is appropriate only with properly built ensembles by MD simulations. The double–stranded DNA binding dynamics of HOP2 and most other binding properties can be considered as *non–equilibrium dynamics*, the study of which is attainable with MD simulations. We chose GROningen MAchine for Chemical Simulations (GROMACS) originally developed by University of Groningen due to following reasons:

a. It is free and compatible with all operating systems.
b. Parallel computing possible.
c. It is based on classical mechanics so it is fast and reliable.
d. As our system does not contain charge transitions like electron transition, classical mechanics is adequate to study its dsDNA binding dynamics.

Chapter IV: MOLECULAR DYNAMICS SIMULATION OF MEMBRANE PROTEINS

In this chapter, we discuss the general outline of the MD simulation methods used to simulate membrane proteins. We used this method to simulate two membrane proteins and generate their structural models.

Chapter IV: SOLUTION STRUCTURE AND DNA–BINDING PROPERTIES OF WINGED HELIX DOMAIN OF THE MEIOTIC RECOMBINATION HOP2 PROTEIN

This chapter is dedicated to our work published in Journal of Biological Chemistry (2014). This work was completed in collaboration with Dr. Roberto J. Pezza and his research group at Oklahoma Medical Research Foundation (OMRF). We used the solution NMR methods to determine the structure of DNA binding domain of HOP2 (N-terminal HOP2). The method we used is slightly different than conventional method, which is considerably fast and free of ambiguities, especially for smaller proteins like N-terminal HOP2. The NMR chemical shift perturbation method used to study the dsDNA binding of N-terminal HOP2 is also discussed in this chapter.

Chapter VI: WING 1 OF PROTEIN HOP2 IS AS IMPORTANT AS HELIX 3 IN DNA BINDING BY MD SIMULATION

This chapter is dedicated to our work published in Journal of Biomolecular Structure and Dynamics (2017). In this work, we used molecular dynamic (MD) simulation techniques to characterize the DNA binding dynamics of N-terminal HOP2. The careful design of DNA complex systems of HOP2 and its mutants, implementation of MD simulations, and analysis of trajectories for interactions and dynamics are presented.

Chapter VII: IMPLEMENTATION OF MD SIMULATIONS TO GENERATE STRUCTURAL MODELS OF MEMBRANE PROTEINS AND VERIFICATION BY SOLID–STATE NMR

In this chapter, we discuss the implementation of MD simulations to build structural models of two membrane proteins and verified them with the help of experimental solid–state NMR data. In the first one, the models for membrane attachment motif of lipid droplet storage protein 1 (Lsd1) were built and simulated. The chemical shifts of simulated models were generated and compared with the experimental NMR spectra to verify the validity of models. Similar approach was adapted to construct the structural model of linker region of another membrane protein syntaxin.

1.3 Our Contribution

In this thesis, we provide experimental and computational framework that can be applied to a wide range of protein systems to study their structure and dynamics. The flexibility of the methods we applied makes it possible to integrate experimental and computational methods to study large protein complexes, which is the main challenge given the size of the system. Following is a list of our specific contributions:

a. Fast and reliable high–resolution structure calculation of smaller proteins with less ambiguities.
b. A very efficient set of experimental designs for studying structure and binding properties of proteins.
c. A better way of studying thermodynamics and binding dynamics of proteins.
d. A detailed methodology of studying structure and dynamics of membrane bound proteins by combining experimental solid–state NMR and MD simulations.

CHAPTER II

NUCLEAR MAGNETIC RESONANCE

2.1 Introduction

The certain atomic nuclei exhibit an inherent property (quantum spin) that causes them to behave like tiny magnetic dipoles when placed in a magnetic field. These dipoles precess in the external magnetic field with characteristic frequency called precession frequency. An electromagnetic radiation with sufficient energy to match the precession frequency of a nucleus (resonance) excites it and then radiation is re–emitted inducing a measureable signal while it relaxes back to original state. The spectrum of induced signals is signature of a molecule; which is used to identify the type and concentration of nuclei in the sample. This technique, also known as nuclear magnetic resonance (NMR), got wide popularity only after late 1940s when independent research teams led by Felix Bloch at Stanford and Edward Mills Purcell at Massachusetts Institute of Technology simultaneously demonstrated NMR in condensed matter (18, 19). The development of pulsed Fourier transforms NMR and multidimensional NMR spectroscopy made it more popular among researchers in different fields. Despite its very initial application in physics, it has been rigorously used in biology, chemistry and clinical (MRI) studies.

The X–ray crystallography has been the dominant method in structural biology however, the use of superconducting magnets and implementation of fast fourier transform quickly accelerated the use of NMR as one of the powerful tools in protein structure determination. Although the number of protein structures determined using NMR is significantly low compared to X–ray crystallography, NMR has a distinct advantage in determining high–resolution structures and investigating the nanosecond time–scale dynamics of bio–molecules. Also, NMR chemical shift mapping can be used to study protein–protein or protein–ligand interactions that play key role in various cellular processes. This method can also be very useful for drug molecule screening and optimization. Another major advantage of NMR lies in the flexibility of sample conditions. Unlike X–ray crystallography where crystal sample is necessary, NMR can be applied directly to protein samples in solutions or various solid forms like powders, frozen solutions, micro–crystals, gels or proteoliposome. This versatility is extremely important for some proteins, which can only be studied in only one sample condition.

The majority of NMR researches are performed in liquid or solvated samples, such as proteins that range in size up to about 30 kDa known as solution NMR technique. The rapid tumbling motions of the molecules in solutions allow for spatially independent high resolution NMR spectra. But for large, insoluble proteins and some samples that exhibit anisotropic interactions in the external magnetic field, solid state NMR (SSNMR) is used. Solid state NMR requires the use of certain techniques (e.g., sample rotation and special pulse sequences) to overcome the loss of resolution due to the relatively motionless molecules. With these SSNMR techniques, proteins with higher molecular weight and even intact biological tissues can be studied.

2.2 Brief History of NMR

The first successful demonstration of nuclear magnetic resonance (NMR) in bulk matter were published in 1946 by two independent groups led by Bloch at Stanford University and Purcell at Harvard University (18, 19). The impact of their work was immediate and the applications of NMR have steadily widened from physics and chemistry to many different disciplines. The joint award of the 1952 Nobel Prize for Physics to those two leaders recognized this landmark discovery. Although NMR was first discovered and experimentally demonstrated in ordinary materials in 1946, the basic concepts of electron spin and associated magnetic moments began in early 1920s by number of studies especially Stern–Gerlach experiment. By mid–1920s, became apparent that the atomic nuclei also posses spin and a magnetic moment, which was verified in 1933 by refined Stern–Gerlach experiment. The studies on the resonance of nuclei however began in 1936 when Gorter looked for resonance of Lithium nuclei in crystalline lithium fluoride and of protons in crystalline potassium alum, but without success. In 1939, Rabi and co–workers at Columbia University demonstrated nuclear magnetic resonance when a beam of hydrogen molecules was sent through magnetic field and molecules absorbed radio frequency electromagnetic radiation sharply at defined frequency.

After the successful application of NMR, focus was turned into building efficient NMR spectrometer based on homogeneous magnet. Donald and Phillips in 1967 demonstrated that the higher magnetic field strength could result in improved signal resolution. Further advances in biological NMR spectroscopy were made possible by two technological improvements, Fourier Transformation–NMR and 2D–NMR. The development of computers advanced the FT–NMR methods. The sensitivity could be enhanced by the time averaging in a practical manner and the speed of pulse FT method could be exploited alternatively to study fast processes like chemical reactions and time–dependent NMR phenomena (e.g., relaxation). The FT methods also permitted the study of solids with chemically shifted lines. Magic angle spinning discovered in 1959, could now be used in conjunction with new techniques that transfer magnetization from one species to other (cross polarization) to obtain high–resolution spectra of 13C and other nuclei in solids. The most exciting new area was 2D NMR spectroscopy, in which nuclear magnetization are allowed to precess during an initial time period, various pulse sequences are applied, and an FID is recorded, 2D FT of the two independent time domains results in a spectrum that can be displayed along two orthogonal frequency dimensions. Jenner originated the idea of 2D NMR in 1971, but Ernst was the one developing it into a practical and useful method during mid–1970s. 2D NMR spectroscopy is a very powerful method for assigning lines in complex spectra and for studying interactions mediated by cross relaxation, chemical exchange or other physical factors.

The studies of biopolymers received a significant boost with the introduction of high–field magnets that permitted separation of spectral lines caused by chemically distinct nuclei. The real brakethrough came when Nagayama, Wuthrich, Bachmann and Ernst showed that 2D NMR methods could be applied to biopolymers. The spin–coupling connectivity was established with correlated spectroscopy (COSY) and nuclear Overhauser enhancement spectroscopy (NOESY) allowed relaxation effects to be used to estimate intermolecular distances. By 1980s solvent signal suppression methods had been developed to the point where the spectra of proteins in water could be recorded, permitting peptide NH resonances to be included in the spin–coupling and NME pathways. With this, it became possible to determine the complete 3D structure of small protein and in 1985 the structure of 57–residue protein was published (20). By then, NMR has already been established as an alternative to X-ray crystallography method.

2.3 Basics of NMR

2.3.1 Nuclear magnetization

The nucleus of every element and isotope contains protons and neutrons collectively known as nucleons. The number of unpaired nucleons dictates the inherent property of the nucleus known as quantum spin number (I). Nuclei that have unpaired nucleon with an even mass numbers (A) and even atomic numbers (Z) have spin $I = 0$. As the response to external magnetic field depends on I, these nuclei are not affected by external magnetic field and are not NMR active. Nuclei with even mass numbers and odd atomic numbers have integral spin numbers ($I = 1, 2, 3, \ldots$) and those having an odd mass number (regardless of atomic number) have half–integral spin numbers ($I = 1/2, 3/2, 5/2, \ldots$). These nuclei with non–zero quantum spin numbers are in some way affected by external magnetic field and so are NMR active.

Some important NMR active isotopes with their spin numbers as well as other noted properties are listed in Table 2.1 below.

Table 2.1: Nuclear properties of some NMR active isotopes placed in a 14.1 Tesla magnetic field, the absolute value of the Larmor (precession) frequency of each nucleus is shown.

Nucleus	Spin *I*	Gyromagnetic ratio $\gamma\,(\mathrm{T\cdot s})^{-1}$	Natural Abundance (%)	Larmor Frequency in 14.1 T field (MHz)
^{1}H	½	2.6752×10^{8}	99.99	600.34
^{2}H	1	4.107×10^{7}	0.012	92.15
^{13}C	½	6.728×10^{7}	1.07	150.94
^{14}N	1	1.934×10^{7}	99.63	43.39
^{15}N	½	-2.713×10^{7}	0.37	60.87
^{23}Na	3/2	7.081×10^{7}	100.00	158.89
^{31}P	½	1.0839×10^{8}	100.00	243.25

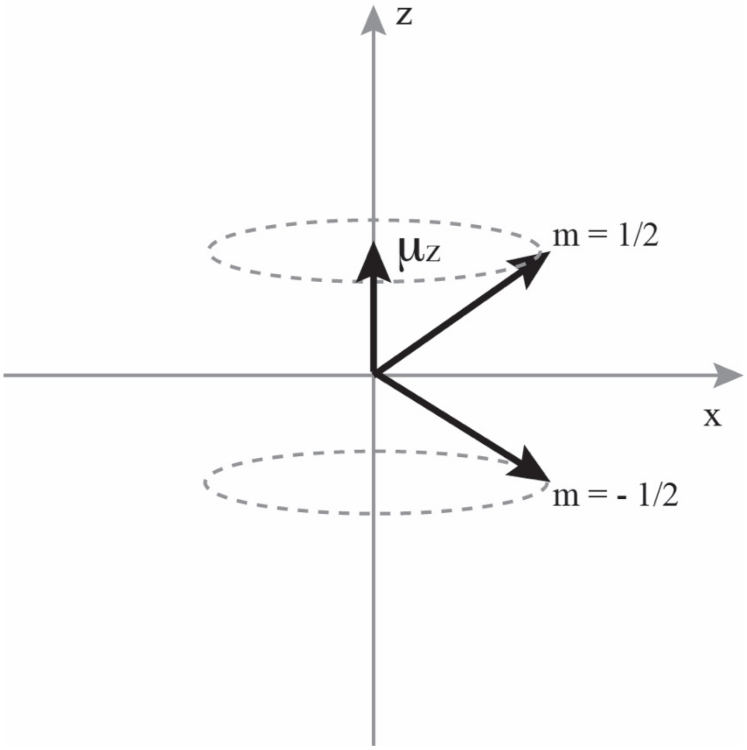


Figure 2.1 Orientation of nuclear angular moment (μ) for spin ½ nuclei. One orientation is pointing to the positive z–axis, the other orientation is pointing to the negative z–axis.

The number of possible spin states for a nucleus with quantum spin number *I* is given by *m* = 2*I*+1; m having values in the range *I, I–1, I–2, ..., –I*. Therefore, the values of '*m*' for a nucleus with spin *I = 3/2* are 3/2, 1/2, –1/2, and –3/2, so this nucleus can have four possible states. The isotopes targeted in this work are ^{1}H, ^{13}C and ^{15}N (spin 1/2 nuclei) and these have just two allowed spin states, m = ±1/2 (Figure 2.1).

Each nuclear spin possess a nuclear angular momentum, the z–component of which is given by:

$$L_z=\hbar m \qquad (2.1)$$

where $\hbar$ is reduced Planck constant (1.054×10^{-34} J·s/rad).

The nuclear magnetic moment in the z direction as shown in Figure 2.1 is given by:

$$\vec{\mu} = |\vec{\mu}_z| = \gamma \hbar m \qquad (2.2)$$

where γ is gyromagnetic ratio of a given nucleus (see Table 2.1).

2.3.2 Energy States and Boltzmann Distribution

When nuclei with spin ½ are placed in an external magnetic field (B_0), the nuclear magnetic moment experiences a torque that tends to align it either parallel or anti–parallel to the field. Nuclei in spin state m = +1/2 (positive μ) align parallel to the field, and those in the negative spin (negative μ) align anti–parallel to the field. Nuclei that align parallel to the field are in lower energy state (α–state (E_1)) and that align anti–parallel to the field are in higher energy state (β–state (E_2)). These energy states represent two energy levels of the entire system. The energy of a spin in external magnetic field is given by the equation:

$$E = -\vec{\mu}_z \cdot \vec{B}_0 = -m\hbar\gamma B_0 \qquad (2.3)$$

The two energy levels are given by:

$$E_1 = -\tfrac{1}{2}\gamma \hbar B_0 \tag{2.4}$$

and

$$E_2 = +\tfrac{1}{2}\gamma \hbar B_0 \tag{2.5}$$

This separation of nuclear spins into energy levels in a static magnetic field is known as *Zeeman effect* or *Zeeman splitting* shown in Figure 2.2.

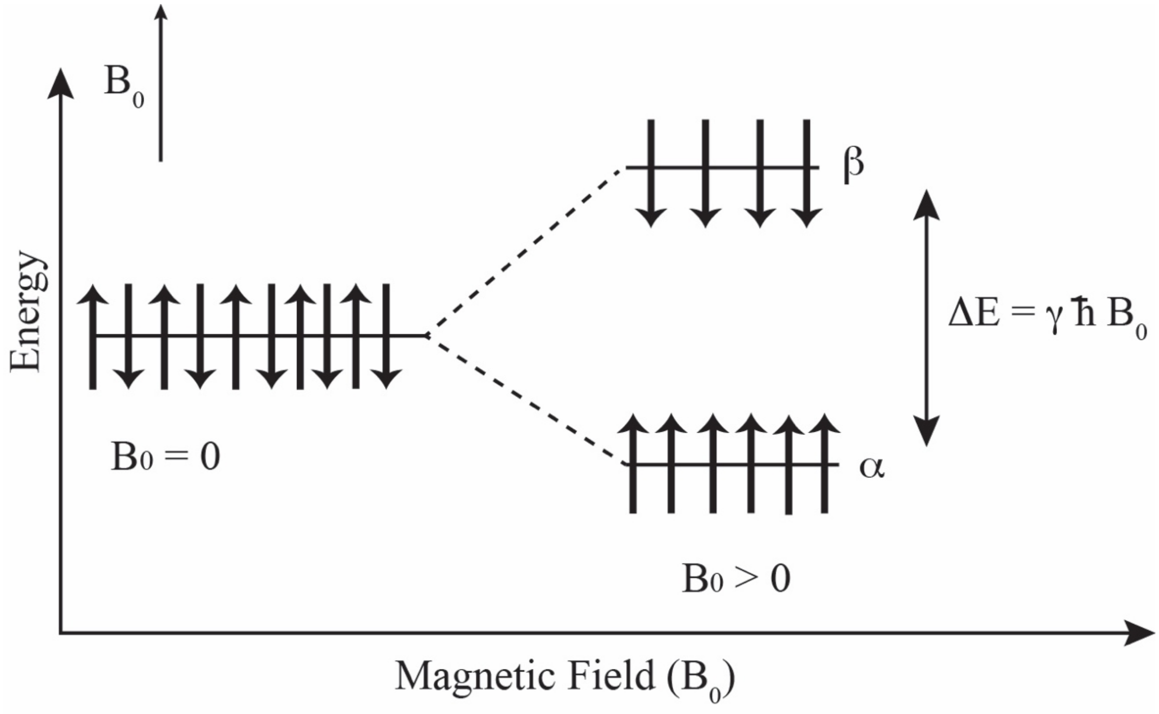


Figure 2.2: Zeeman splitting of nuclear energy levels E_1 and E_2 in the presence of a static magnetic field, B_0. The difference in populations of parallel and anti–parallel spins correlates to the strength of B_0, and factors into the sensitivity of NMR measurements.

The energy difference between these two states is given by

$$\Delta E = \gamma \hbar B_0 \tag{2.6}$$

By definition, the energy difference is $\Delta E = \hbar\omega$. Therefore, the transition frequency ω_0 is given by:

$$\omega_0 = \gamma B_0 \tag{2.7}$$

where ω0 is the Larmor frequency of nuclear spins. Larmor frequency of a specific nucleus is linearly proportional to the static external magnetic field B_0. Instead of angular frequency ω_0, which has a unit of radians per second, Larmor frequency can also be represented in hertz by linear frequency ν:

$$\upsilon = \frac{\gamma B_0}{2\pi} \tag{2.8}$$

In thermodynamic equilibrium, the populations of nuclei in each energy state are nearly equal with the nuclei in parallel state (lower energy) only slightly more. The ratio of populations in these two energy states is governed by the Boltzmann equation and with Taylor expansion,

$$\frac{N_\beta}{N_\alpha} = e^{\frac{-\Delta E}{k_B T}} = e^{\frac{-\gamma \hbar B_0}{k_B T}} = \frac{1}{e^{\gamma \hbar B_0 / k_B T}} \approx 1 - \frac{\gamma \hbar B_0}{k_B T} \tag{2.9}$$

where, k_B is the Boltzmann constant (1.380×10^{-23} J/K), and T is the absolute temperature in Kelvin. For instance, the population ratio for protons (^{1}H nuclei) at 800MHz and at room temperature (300 K) is 0.99987. This indicates that only small fraction of the spins contribute to the signal intensity. This is why NMR is intrinsically very insensitive but sensitivity increases with the magnetic field.

2.3.3 Bulk magnetization

The vector sum of all nuclear angular moments (μ) of a sample is called the bulk magnetization (M_0). The bulk magnetization is associated with the small population difference of spins between the higher and lower energy spin states. The net magnetic moment (the vector sum of individual moments) associated with this very small population difference (ΔN) is given by,

$$M_0 = \gamma \, \Delta N \tag{2.10}$$

This net magnetic moment of tiny population difference determines the sensitivity of NMR spectroscopy, which is one of the major limits of NMR. At equilibrium in an external uniform magnetic field, the net magnetic moment is aligned with the field

$$M_z = M_0 \tag{2.11}$$

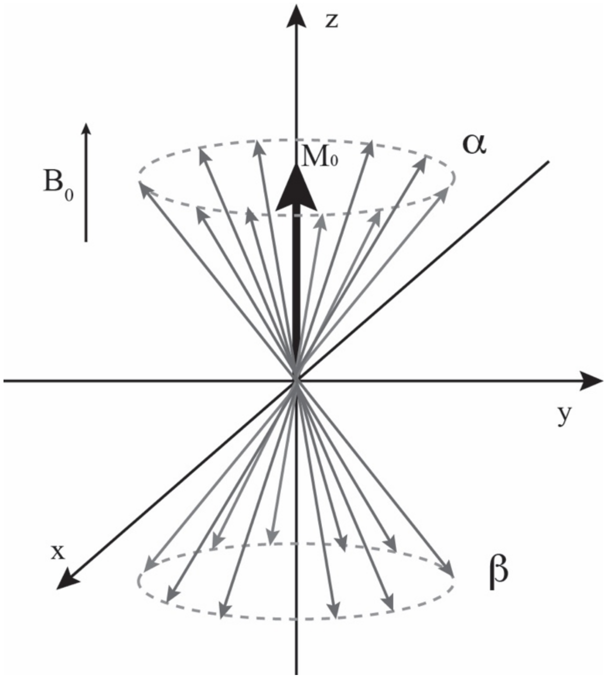


Figure 2.3 The bulk magnetization M_0 of spin ½ nuclei ($\gamma>0$) is represented by a thick arrow. Individual nuclear moments are represented by thin arrows and have equal probability of being in any direction in the *xy* plane.

The net moment in the x–y plane is zero. The magnetic moment of a nucleus precesses at an angle θ around the z–axis as shown in Figure 2.2. The frequency of the precession is known as the Larmor frequency, which is given by,

$$v_0 = \frac{-\gamma B_0}{2\pi} \tag{2.12}$$

2.3.4 NMR pulse and Free Induction Decay (FID)

In NMR experiments, the nuclear magnetic moment aligned to external B_0 field is excited and forced out of alignment by a pulse of electromagnetic radiation with frequency equal to its Larmor frequency. This is actually achieved by applying a field B1 that rotates about z–axis of laboratory frame at the same or near to the frequency of Larmor frequency of nuclei. This oscillating field B1 is generated by the oscillating current passing through the probe coil. The rotation frequency of B1 field is ω_{rf}, known as carrier frequency or transmitter frequency. Resonance between the carrier and nuclear frequencies are achieved when $\omega_{rf} = \omega_0$ which is capable of exciting low–energy (parallel) nuclear spins into higher energy levels. These excited spins then transition back to original energy levels inducing measureable current in the receiver coil that form the NMR signal. The rotation frequency of bulk magnetization M0 due to B1 field is given by

$$\omega_1 = -\gamma B_1 \tag{2.13}$$

Theoretically, same nuclei should have similar resonance frequencies but in reality that is not the case because the local electronic environments of nuclei in a sample are not always similar. The electrons around the nuclei of interest can induce magnetic field that increases (or decreases) the local effective field also known as shielding (or de–shielding) effects. Due to this difference in local electronic environment, nuclei possess different resonance frequencies, which give the sample its characteristic chemical shift spectrum. The resolution of the spectrometer is determined by how well each of the nuclear frequencies is separated (resolved) in the chemical shift spectrum.

Conventionally, Larmor frequency of a proton (1H) is also used to characterize the spectrometer; for example, a spectrometer equipped with a 14.1 T magnet is called a 600 MHz spectrometer since the proton precesses at ν_0 = 600 MHz. All of the experiments reported in this work are performed on a VARIAN 600 MHz spectrometer.

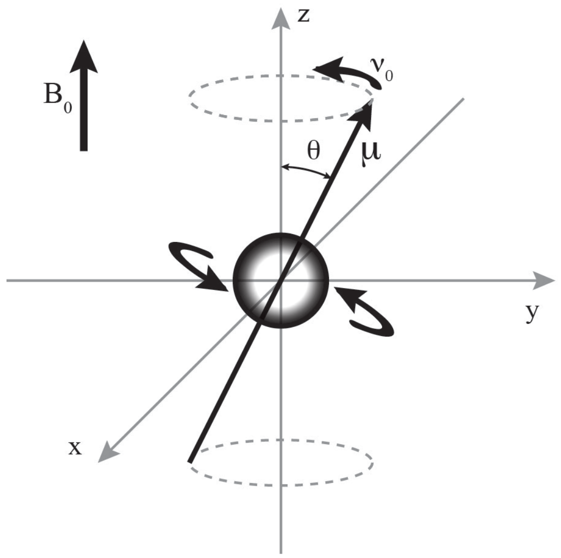


Figure 2.4: Model of a precessing nucleus tilts by an angle θ around the external magnetic field, B_0, axis. The rate of the precession is shown as ν_0, also known as the “Larmor frequency.” The spinning charge creates its own magnetic field, producing a tiny magnetic moment vector $\boldsymbol{\mu}$. The net magnetic moment of an ensemble of spins points along the z–axis.

The net magnetic moment, M_z can be rotated to a desired angle, in radians, given by

$$\theta = -\gamma B_1 \tau_p = -2\pi \nu_1 \tau_p = -\omega_1 \tau_p \tag{2.14}$$

where ω_1 is the angular frequency of applied pulse, and τ_P is the pulse duration or “pulse width” (seconds). A $\pi/2$ (or 90°) pulse rotates the net magnetic moment vector to the x–y (transverse) plane. A π (or 180°) pulse rotates the vector to the negative z–axis. After a pulse, the magnetization vector relaxes back to its equilibrium position, M_z. During its return, the net magnetic moment is coherent (individual moments are in phase) giving a non–zero moment that precesses in the x–y plane at the Larmor frequency. The precessing moment re–radiates absorbed energy and induces a current in the receiver coil of the spectrometer. The received signal is a direct result of Faraday’s Law, where the induced electromotive force (EMF) on a single wire receiver loop with magnetic flux Φ is given by

$$EMF(t) = -\frac{d\,\Phi(t)}{dt} = \omega_0 \frac{\mu_0}{2\,r_{coil}} \left|\vec{\mu}\right| \sin\theta \, \sin(\omega_0 t) \tag{2.15}$$

This induced EMF depends on the radius of the receiver coil r_{coil}, the net magnetic dipole strength $|\vec{\mu}|$, the θ angle from the z–axis, and the precessing angular frequency ω_0 that depends on the static magnetic field strength. The induced signal is an exponentially decaying function of time also known as the *free induction decay* (FID). The FID is then Fourier transformed (FT) to the frequency domain to obtain the characteristic NMR spectrum.

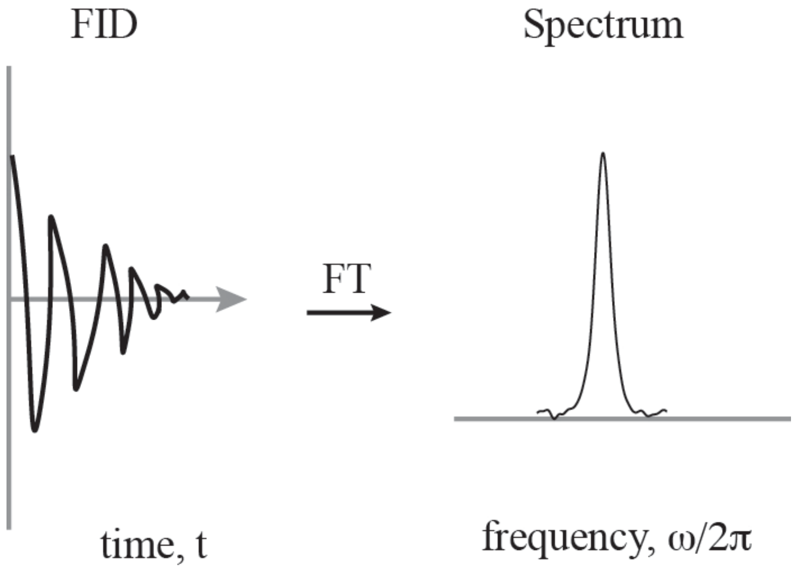


Figure 2.5 FID transformed from time domain to the frequency signal.

In NMR experiments, the frequency of a sample is usually compared to the reference frequency of a standard sample. The magnitude of the frequency shift depends on the strength of the spectrometer: stronger static magnetic fields produce greater frequency shifts. In order to compare the frequency shifts between spectrometers, one may use the relative *chemical shift*, which would be standard across spectrometer sizes. Chemical shift, δ, is calculated by the expression,

$$\delta(ppm) = \frac{\upsilon_{sample} - \upsilon_{reference}}{\upsilon_0} \times 10^6 \tag{2.16}$$

with units of parts per million (ppm).

Table 2.2 Average chemical shifts of nuclei in amino–acids of proteins (in ppm) (21).

Residue	$^{1}H^{N}$	^{15}N	$^{13}C'$	$^{13}C\alpha$	$^{1}H\alpha$	$^{1}H\beta$
Ala	8.15	122.5	177.6	52.2	4.33	1.39
Arg	8.27	120.8	176.6	56.0	4.35	1.89, 1.79
Asn	8.38	119.5	175.6	52.7	4.74	2.83, 2.75
Asp	8.37	120.6	176.8	53.9	4.71	2.84, 2.75
Cys	8.23	118.0	174.6	56.8	4.54	3.28, 2.96
Gln	8.27	120.3	175.6	56.0	4.33	2.13, 2.01
Glu	8.36	121.3	176.6	56.3	4.33	2.09,1.97
Gly	8.29	108.9	173.6	45.0	3.96	.
His	8.28	119.1	174.9	55.5	4.60	3.26, 3.20
Ile	8.21	123.2	176.5	61.2	4.17	1.90
Leu	8.23	121.8	176.9	55.0	4.32	1.65
Lys	8.25	121.5	176.5	56.4	4.33	1.85, 1.76
Met	8.29	120.5	176.3	55.2	4.48	2.15, 2.01
Phe	8.30	120.9	175.9	57.9	4.63	3.22, 2.99
Pro	.	128.1	176.0	63.0	4.42	2.28, 2.02
Ser	8.31	116.7	174.4	58.1	4.47	3.88
Thr	8.24	114.2	174.8	62.0	4.35	4.22
Trp	8.18	120.5	173.6	57.6	4.66	3.32, 3.19
Tyr	8.28	122.0	175.9	58.0	4.55	3.13, 2.92
Val	8.19	121.1	176.0	62.2	4.12	2.13

Note: The chemical shifts of ^{1}H, ^{13}C and ^{15}N are referenced to DSS, DSS and liquid NH3, respectively.

The direction of magnetization can also be manipulated by phased pulse that follows the right–hand rule. If the magnetic field B_1 is applied along the +y–axis, then the magnetization M_0 will rotate from the +z–axis around the y–axis by an angle θ as in Figure 2.6 (A). The resulting magnetizations are

$$M_{0z} \xrightarrow{\theta_+ y} \begin{array}{l} M_x = M_0 \sin\theta \\ M_y = 0 \\ M_z = M_0 \cos\theta \end{array} \tag{2.17}$$

Equilibration of the net magnetization results in the decay of the transverse magnetization, shown in Figure 2.6 (B), which induces the FID that is used to create spectrum.

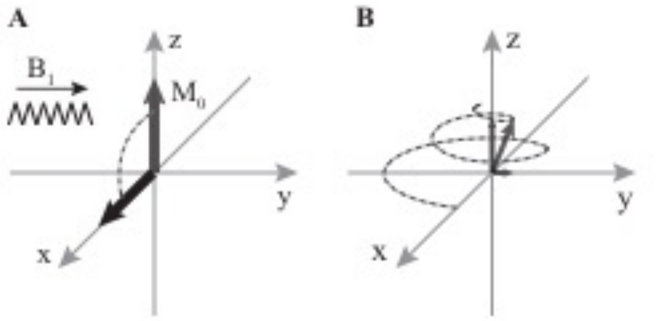


Figure 2.6: (A) In the Bloch model, an on–resonance 90° pulse with y–phase, B_1, rotates the magnetization vector (M_z) away from its equilibrium position on the z–axis through an angle of $\theta = 90°$ to the x–axis, causing a transverse magnetization (M_x). (B) After the pulse, the magnetization vector precesses back to its equilibrium position. During this time, known as the *acquisition period*, the magnetization induces a signal in the receiver coil. This signal is exponentially decaying function of time known as *free induction decay (FID)*.

2.3.5 Relaxation

The time required for the net magnetization to reach their equilibration state is labeled T_1 and known as the *spin–lattice relaxation time* or the *longitudinal relaxation time*. In most NMR experiments, after receiving an excitation pulse, the system of nuclei needs time to re–equilibrate before the next pulse, i.e. the net magnetization is restored along the z–axis. This time is determined by T_1 and allows the nuclear spins to dissipate their energy to the lattice (i.e., non–participating, surrounding nuclei). For a system having a relaxation time constant T_1, the instantaneous net magnetization can be calculated by

$$M(t) = M_0(1 - e^{\frac{-t}{T_1}}) \tag{2.18}$$

Generally, an experiment is repeated after a period of $5 \cdot T_1$, which allows plenty of time for the spins to fully recover the net magnetization, M_0.

The energy absorbed from an applied electromagnetic field is dissipated not only to the lattice as in T_1 relaxation, but can also be exchanged between spins of the system. Interactions between spins (as well as varying nuclear environments) result in the loss of phase coherence in the x–y plane (or *transverse plane*) and is governed by another time constant, T_2, see Figure 2.5. The time required for the net magnetization in the transverse plane to dephase is known as T_2. Spin–spin interactions arise from local magnetic field inhomogeneity, which could be attributed to either an inhomogeneous B_0 field from the spectrometer magnet or from varying magnetic susceptibilities within the sample. The spin–spin interactions are dominated by dipolar coupling (for spin ½ systems) for which the time for energy exchanges is approximately

$$T_2 \sim \frac{\hbar r^3}{\mu^2} \tag{2.19}$$

where *r* is the separation distance between nuclei (19, 22). Measuring T_2 values reveals the nuclear environments present in a sample. For example, rigidly bound protons in proteins have shorter T_2 relaxation times than the more labile protons in free water. Environmental differences cause the protons to relax at different rates that result in the dephasing of the spins in the transverse plane.

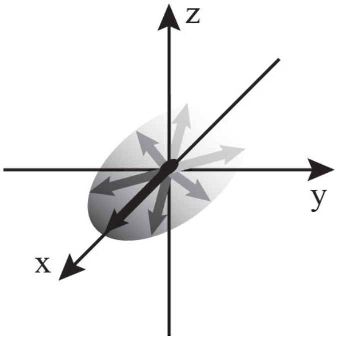


Figure 2.7: Spin–spin (or *transverse* or T_2) relaxation is modeled as decoherence of magnetization in the transverse plane due to local magnetic fluctuations. Magnetization vectors at different nuclear environments will contribute to the transverse relaxation time of a sample. Dephasing of the vectors results in an exponential decay of the net transverse magnetization.

The ideal transverse relaxation could be measured, as it is inversely proportional to the spectral linewidth. However, inhomogeneous local magnetic fields increase the decay rate as individual spins dephase with each other in the x–y plane. The spectral linewidth “artificial” T_2 measurement,

$$LW_{1/2} = 1/\pi T_2^* \tag{2.20}$$

where $LW_{1/2}$ is the spectral peak width at half height, is produced by both the real T_2 value of the sample and the contribution of field inhomogeneity, seen in the equation

$$1/T_2^* = 1/T_2 + 1/T_{inhom}, \tag{2.21}$$

where T_2^* is the spectral linewidth relaxation value, T_2 is the actual transverse relaxation value of the sample, and T_{inhom} arises from field inhomogeneities. Varying magnetic susceptibilities within the sample contribute to the inhomogeneous relaxation. The spectrum from a simple 90° pulse (assuming ideal measurements) is used to find the artificial T_2^* value. This T_2^* value always presents an equal or smaller value than the actual relaxation time.

CHAPTER III

MOLECULAR DYNAMIC SIMULATIONS

3.1 Introduction

The quest to increase the knowledge on how macromolecules are built has been continuous since the beginning of structural calculation in 1950s. The reason is to understand the structural architecture that defines the molecular recognition rules, which power the understanding of basic biological phenomena. The biochemical studies have been successfully driven by the simple visual analysis of 3D structures of protein or nucleic acids obtained by numerous experiments. Despite enormous utility of protein structures in biological studies, the structures may provide only a partial view on their functions because proteins are flexible entities and dynamics play a key role in their functionality. Most proteins go through conformational changes and structural rearrangements while performing their functions. These changes are commonly occurred in catalytic cycle of many enzymes. These conformational changes are attributed by the isolation or the exposure of active sites and allosteric effects that are critical in protein functions.

Traditionally, protein conformational influences on the biomolecular function were studied by calculating the structures that cover the entire conformational space. The generation of conformational ensemble of calcium–binding protein from NMR experiments is an example, where some degrees of conformational variations in loop regions were clearly visible (PDB ID 1A03). Such ensemble of conformations may not provide a complete view on molecular flexibility because structures are merely static view of the molecule. Additionally, the generation of large number of structures is definitely a time consuming and arduous task. Therefore, the recent advances in the performance of simulation algorithms, computing power and strategies to increase the conformational sampling have made the theoretical technique the most convenient way to obtain a picture of macromolecular dynamics.

The macroscopic physical properties of a system can be distinguished by the *static equilibrium properties* (binding constant, the average potential energy or the radial distribution function etc.) and the *dynamic or non–equilibrium properties* (fluid viscosity, diffusion processes in membranes, phase change dynamics, reaction kinetics or the crystal defect dynamics etc.). These properties can be obtained with ensemble averages over a representative statistical ensemble of molecular systems. Traditionally, the *Monte Carlo* (MC) and *molecular dynamics* (MD) methods cover the two classes of stochastic and deterministic simulations respectively that can be used for the generation of a representative equilibrium ensemble. For the generation of non–equilibrium ensembles and for the analysis of dynamic events, only the molecular dynamics (MD) method is appropriate.

Molecular dynamics (MD) simulation is a theoretical method that is used to study the dynamic behavior of a system of particles. This method is a very powerful to attack the many–body problems in several areas of interest including statistical physics, physical chemistry and biophysics. Although sophisticated experiments can be used to achieve some microscopic properties of complex systems, one still has to rely on simulation to study specific aspects in great detail. Also, simulations can be used to accurately characterize the system of interest by providing specific input parameters that come straight from theory and experimental data. Additionally, this technique can often be used to solve theoretical models by using model parameters and provide significant information for further investigation.

There are three steps that essentially define molecular dynamics simulations: generation of a model system, integration of equation of motions and the generation of a statistical ensemble. The system model generation is a crucial step in MD simulation as the success of getting appropriate results depends on the initial conditions of the system. A model contains defined system constituents (atoms, molecules, surfaces etc.) that interact in certain way that can be generated and refined by testing it against real system properties and theoretical predictions. The model system also can be constructed by using parameters from experiments like neutron diffraction or NMR measurements. An integrator propagates particle positions and velocities from time t to t+dt that follows the path to convergence. An ensemble of system is chosen where thermodynamic quantities can be controlled.

The limited range of accessible time– and length–scale is the main drawback of simulation studies. More detailed a simulation technique operates, the smaller the accessibility of time and length scales. For example, if we consider motions of electrons in a system using quantum simulations, the only accessible scale of length and time are of order of Angstrom (Å) and picosecond (*ps*) respectively. Another factor that affects the performance of particle dynamics simulations is the computing power. The shorter time scale MD simulations of smaller systems are achievable with modern desktop computers however, for longer times and larger systems it is necessary to have access to cluster computers with very high computing power.

3.2 Brief History

The use of computer simulations began in early 1950s when N. Metropolis in 1953 applied Monte–Carlo technique to solve physical equations of a system composed of interacting individual molecules. The idea was first to generate a subset of random number to represent conformational space and then use exponent of the energy as probability filter. In 1955 Fermi, Pasta, and Ulam conducted simulation of anharmonic 1D crystal. The proper molecular dynamics (MD) simulation study was however, first reported in 1956 by B.J. Alder and T.E. Wainwright. They simulated an assembly of hard spheres in order to study the phase transition. In 1960, Vineyard group simulated damaged Copper crystal. In 1964, A. Rahman studied correlations in the motion of atoms in liquid Argon using MD simulations. This was the first MD simulation, which was applied to atoms interacting via a continuous potential. In this case, not only binary collisions were taken into account the interactions were modeled by a Lennard–Jones potential and the equations of motion were integrated with a finite difference scheme. It was the first work of its kind where an exact method was used to calculate dynamical quantities of a system. In 1969, Baker and Watts conducted simulation of water using MC, which was followed by MD simulation of water in 1971 by Rahman and Stillinger.

In 1977, the first MD simulation of proteins by McCammon, Gelin, and Karplus was reported. By 1980s, free energy calculations and Protein–ligand docking calculations were done. The emphasis was given to force field development and sampling techniques in 1990s. The 1us MD simulation on folding of Villin headpiece in explicit solvent was done by Duan and Kollman in 1998. Anton supercomputer specialized for MD simulation works were designed in 2009. The computer simulation works were recognized in 2013 with a Nobel Prize in chemistry to Karplus, Michael Levitt and Arieh Warshel for the development of multi–scale models for complex chemical systems.

3.3 Basics of MD simulations and GROMACS

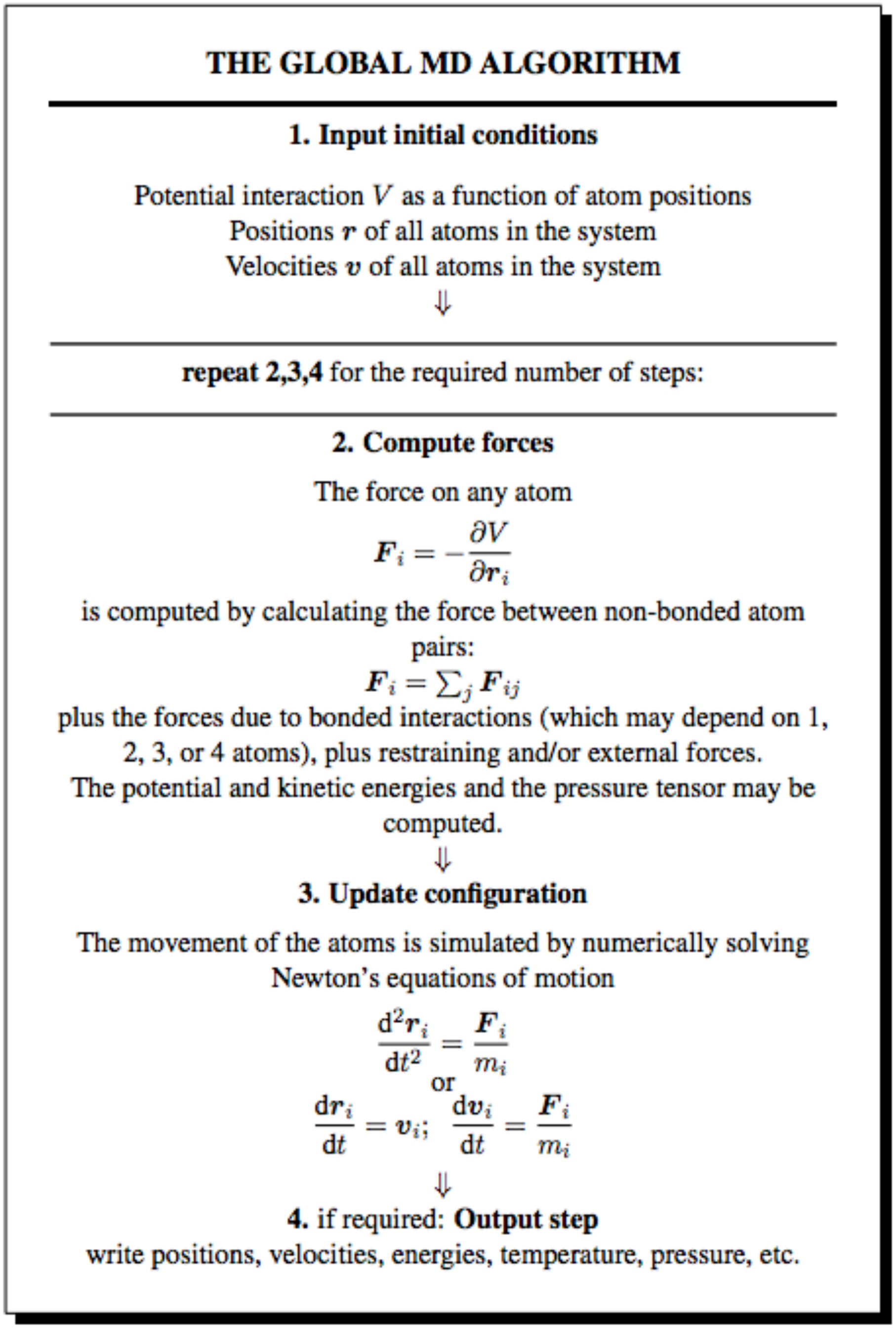
Molecular dynamics (MD) simulation is the modern realization of an old fashioned idea in science; the computation of system behavior is possible if we have a set of initial coordinates for the system and a set of force fields for interactions. MD requires – and provides – complete information of position and momentum of all atoms at all times that can be used to inquire specifics relating to the system of interest. Currently, classical molecular dynamics methods are frequently used to solve many problems, e.g. properties of liquids, defects in solids, fracture, surface properties, friction, molecular clusters, biomolecular dynamics etc. Also the development of MD simulation technique is continuous, as the large groups of researchers have been dedicating their works on the development of software and force fields. GROMACS is one of the freely available and robust software packages for MD simulation works. It is based on the classical mechanics, where Newton's equations of motion for a system of N interacting atoms are solved,

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i \, , \qquad i=1,2,\ldots,N. \qquad (3.1)$$

The forces are the negative derivatives of a potential function $V\ (r_1, r_2, \ldots.., r_N)$:

$$F_i = -\frac{\partial V}{\partial r_i} \qquad (3.2)$$

These equations are solved simultaneously in small time steps. The atomic positions and velocities are updated and written in an output files at every time step by integrator algorithms. These outputs with coordinates and velocities as a function of time represent a trajectory of a system. The temperature and the pressure of the system are forced to remain at required values using specific algorithms to maintain the system for desired ensemble. The macroscopic properties of a system are extracted from the trajectory when system reaches an equilibrium state. The global algorithm that GROMACS uses for the simulation of system of N particles is shown in figure 3.3.1.



THE GLOBAL MD ALGORITHM

---

1. Input initial conditions

Potential interaction V as a function of atom positions
Positions $\boldsymbol{r}$ of all atoms in the system
Velocities $\boldsymbol{v}$ of all atoms in the system
⇓

---

repeat 2,3,4 for the required number of steps:

---

2. Compute forces

The force on any atom

$$\boldsymbol{F}_i = -\frac{\partial V}{\partial \boldsymbol{r}_i}$$

is computed by calculating the force between non-bonded atom pairs:

$$\boldsymbol{F}_i = \sum_j \boldsymbol{F}_{ij}$$

plus the forces due to bonded interactions (which may depend on 1, 2, 3, or 4 atoms), plus restraining and/or external forces.
The potential and kinetic energies and the pressure tensor may be computed.
⇓

3. Update configuration

The movement of the atoms is simulated by numerically solving Newton's equations of motion

$$\frac{\mathrm{d}^2\boldsymbol{r}_i}{\mathrm{d}t^2} = \frac{\boldsymbol{F}_i}{m_i}$$

or

$$\frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t} = \boldsymbol{v}_i ; \quad \frac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} = \frac{\boldsymbol{F}_i}{m_i}$$

⇓

4. if required: **Output step**
write positions, velocities, energies, temperature, pressure, etc.

Figure 3.1 The global MD algorithm (23).

3.4 Use of MD simulations in biology and significance

The original simulation of the bovine trypsin inhibitor (BPTI) provided the fundamental insights concerning the internal motions of proteins, which also dismissed the idea that proteins were rigid structures (24). Since then, molecular dynamics progressed enormously with the use of improved potential functions and simulation algorithms coupled with increased computing powers. Furthermore, MD simulations are firmly established in certain areas of importance in biology. This method has been frequently used from the structure refinement to structure determination using experimental data of X–ray crystallography and NMR spectroscopy. Some examples of use of MD simulations are the generation of structures from NMR chemical shifts, providing structural data on partially folded proteins on lipid bilayers and on viral genome packaging etc.

The internal motions of proteins and resulting conformational changes play an essential role in their functions. Such properties of proteins can be studied using MD simulations that have successfully established the dynamic picture of biomolecules. The availability of simulation software, increasing computing power, and ease of use have fueled this field of study. MD simulation can provide ultimate details concerning individual particle motions as a function of time so it has become one of the most used tools for understanding the physical basis of the structure and function of biological macromolecules. Many specific questions regarding the properties of a model system can be addressed by simulations often more conveniently than experiments on the actual system. However, experiments always play an essential role in validating the simulations. Also, comparisons of simulation and experimental data can serve to test the accuracy of the calculated results that provides criteria for improving the methodology.

As the development continues, the studies using MD simulations have increased many folds. The use of experimental NMR and MD simulations in determining structures and studying dynamics and thermodynamics is now widespread. MD simulations have been successfully implemented in determining role of solvent in protein dynamics and studying conformational changes in the functional mechanism of proteins, dynamic coupling of protein modules, and even protein folding for smaller peptides. With the modern computers and efficient algorithms, the simulation time is extended to a range from 100 ns to microsecond, making it possible to study biological phenomenon as they happen.

3.4 Limitations

The dynamic properties of a protein system obtained by MD simulation can be extremely useful in explaining biological functions however we should also be aware of its limitations. The known experimental properties of the system under study should always be checked to access the accuracy of the results. The approximations adopted in MD simulations are listed as follows:

The simulations are classical: MD simulations are based on classical mechanics and many behaviors like tunneling of electron or protons through potential barriers cannot be explained. Also, the bond and bond–angle vibrations are beyond the classical limit. However, GROMACS considers bond and bond–angle to be constrained that resembles more closely to a quantum oscillator at ground state, which is a better approximation for many systems.

Electrons are in ground state: All the electrons are considered to be in ground state. So MD simulation cannot account any electronic excited state and electron transfer processes.

Force fields are pair-additive and approximate: MD simulation approximates all non–bonded forces by the sum of non–bonded pair interactions. Also, interactions like polarizability, which is non pair–additive are just averaged.

***Boundary conditions are unnatural*:** in MD simulations, periodic boundary conditions are used to avoid real phase boundaries. This may be good for large systems but induces large errors in smaller systems.

CHAPTER IV

MOLECULAR DYNAMICS SIMULATION OF MEMBRANE PROTEINS

4.1 Introduction

Proteins that are partially or completely bound to cell membranes are membrane proteins. These proteins are either partially attached to the membrane (peripheral membrane proteins) or integrated into the membrane and completely cross the bilayer (integral membrane proteins). These proteins carry out variety of biologically important processes like transport of nutrients, toxins and other substances in and out of the cell and consequently play pivotal role in all cellular processes. These proteins are also responsible for cell–cell interactions and membrane fusion.

The human genome is found to encode over 35% of the membrane–associated proteins and so the knowledge of their structures and dynamics is essential for understanding diverse cellular functions. The existing structure calculation techniques may face problems as membrane proteins are naturally associated with lipid bilayer and can only be used in lipid or detergent solutions. The native state of membrane protein is not ideal for structure calculation so far and to make matter worse, some of them can denature in detergent solutions. Additionally, this group of proteins has a very low yield and considered to be very difficult for crystallization. The large conformational changes that these proteins undergo while performing their functions (for example, transporter proteins, which cycle between at least two distinct states) further complicate the problem. These are the reasons why membrane protein structures are underrepresented in Protein Data Bank.

The difficulties in structure calculations of membrane proteins have led the computational studies particularly molecular dynamics (MD) to be used in elucidating useful information of such systems. There has been substantial progress in the simulation of lipid bilayers and membrane proteins embedded within them and MD simulation is the most commonly used method. In this chapter, we discuss how to setup and run membrane protein simulations, focusing on the more practical aspects.

For the simulations of large systems like membrane proteins, a stable simulation is required, usually reflecting the equilibrium state of the system. Also, proper positioning of the protein on the membrane to avoid any unnatural steric clash is critical. The insertion process, if properly done, can be very simple. We place the protein in the bilayer and remove overlapping atoms and/or lipids belonging to overlapping atoms. The positioning of the protein however can be somewhat subjective and that depends on the protein system under study.

4.2 Methods

There are two major components to a membrane protein system: the protein itself and the lipid bilayer.

4.2.1 Preparation of protein

The starting structure for the membrane proteins can be taken from protein data bank (PDB). The sparse availability of such structures and missing components if available in PDB can be a real challenge. However, the practical knowledge on the protein conformation and the use of modeling software can be really useful in designing a reliable starting protein structure model. If few atoms are missing from a small number of residues, one can manually build in the missing atoms using an interactive modeling program such as Visual Molecular Dynamics (VMD) (25) or PyMol. For more complicated case in which a whole loop or units missing, one has to resort to programs that can build random structures that are geometrically correct, such as Modeller. If the structures or structure models are not available, then interactive modeling programs like VMD, Structural Alphabet based protein Backbone Builder from alpha Carbon trace (SABBAC) (26), and simulation software GROMACS can be used to generate hypothesized model structure of full protein or a domain of interest. These hypothesized structures however, have to be refined and validated by some experimental data for reliability and accuracy.

4.2.2 Preparation of lipid bilayer

The simulation of lipid bilayers has matured during last 15 to 20 years and some groups have generously made equilibrated conformations of some lipid bilayer systems freely available in LIPIDBOOK (27). However, sometimes it will be necessary to generate a new bilayer system from scratch and on doing so, one has to take care of the stability of the system. Also the choice of force–field can make huge difference on system reliability. It will be sometimes necessary to modify the existing force–field to incorporate interaction parameters for lipid molecules if it does not include them already.

4.2.3 Protein placement in the membrane

There are different approaches developed by different groups for placing the protein in bilayer. In one approach adopted by Tieleman et al., the lipids and the protein are placed on a widely spaced grid and then shrinking the grid until the bilayer with the protein attains the desired density. The end result would be a system with lipids neatly packed around the protein (28). A different approach proposed by Faraldo–Gomez and colleagues uses a preformed equilibrated bilayer as the starting point (29). According to their method, using the solvent–accessible surface area of the protein as a template, a cavity in a pre–equilibrated lipid bilayer is created. Lipid molecules whose head groups fall within the volume are removed while remaining lipids are subjected to an ever–increasing force acting perpendicular to the surface of the cavity template until the cavity is empty. The protein then simply inserted into the cavity. The advantage of this method is that the pre–equilibrated bilayer can be used, which makes system building process and equilibration extremely fast.

We adopted the direct method of placement of protein in the membrane based on the second approach explained above. We implemented this procedure using VMD (25) in combination with GROMACS. This method is very interactive and exploits the fact that enough simulation time is available to adequately equilibrate the system. Our procedure can be summarized in the following four steps:

a. Obtain or create pre–equilibrated lipid bilayer
b. Align protein in the lipid bilayer in random position and correct orientation
c. Take a note of overlapping lipid molecules and remove them
d. Equilibrate the new system

The alignment of the protein with the pre–equilibrated lipid bilayer in this process is essentially something that is performed by visual inspection guided by pre–set fact or hypothesis on the orientation and position on the membrane. The removal of the overlapping lipid molecules means that the resulting system will have vacuum in between the protein and the lipid molecules. This may create unnatural system where water molecules can get into the hydrophobic core of the bilayer. However, this is avoided during the addition of water where the van der waal radius for carbon atoms are made large enough to prevent water molecules from getting into the interior of bilayer. As the system is equilibrated, the lipid molecules relax around protein molecule.

4.2.4 Equilibration

The equilibration of the protein–bilayer system is essential to allow lipid molecules pack uniformly with the desired density around the protein. The positional restraints are imposed on protein but lipids are allowed to move freely during equilibration process. The equilibration phase is normally determined by monitoring the area per lipid as a function of time. This value can be checked against the experimental data to maximize the reliability of the system.

For the equilibrium purposes, we confined the protein–bilayer system in a periodic box to avoid any real phase boundaries and finite boundary effects. We then added water to fill the remainder of the box, and counter-ions were also added to maintain electroneutrality. The system was energy minimized by using steepest descent algorithm. Two phases of equilibrations, constant volume (NVT ensemble) and constant pressure (NPT ensemble) were then subsequently carried out, each with 1 fs time steps, to maintain the system temperature at 300 K and pressure semi-isotropically at 1 bar respectively. In the first phase, the systems were coupled to a strong temperature bath using V-rescale coupling with temperature coupling constant of $\tau_T = 0.1$ ps. In the second phase, Parrinello Rahman pressure coupling with coupling constant $\tau_P = 5$ ps and a weak Nose-Hoover temperature coupling with a coupling constant $\tau_T = 0.5$ ps was used to ensure the true NPT ensemble.

4.2.5 Running the simulation

The time dependent MD trajectory was obtained by running atomic level simulation, which we used to derive physical properties of the system. All MD simulations were performed using GROMACS 4.5.5 on a Linux cluster using 50 out of 252 standard compute nodes, each with dual Intel Xenon E5-2620 “Sandy Bridge” hex core 2.0 GHz CPUs and 32 GB of 1333 MHz RAM, with an effective calculation rate of 2 days/100 ns for the system with up to 300,000 atoms. The GROMOS96 53A6 force–field combined with lipid interaction parameters with an extended simple point charge (SPCE) water model were used. The LINear Constraint Solver (LINCS) algorithm was used for constraining all bonds. The non-bonded van der Waals interactions were approximated by Lennard-Jones potential. Unless specified, all the interaction range cutoffs were set to 1.2 nm. Electrostatic forces and their contributions to the energies were calculated using Particle-Mesh Ewald (PME) summation algorithm. The production simulations were carried out with 2 fs time steps and the temperature and pressure were maintained using weak coupling methods (Nose-Hoover with τ_T = 0.5 ps and Parrinello-Rahman with τ_P = 2 ps). The MD optimized structures were extracted from the trajectory.

4.2.6 Simulated chemical shifts generation and comparison with NMR spectra

For a given structure model, backbone and β-carbon chemical shifts were predicted by shiftX. Based on these chemical shifts, ^{13}C-^{13}C 2D spectra were simulated by program peaks2ucsf in the Sparky package (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) with assistance of a custom computer scripts.

CHAPTER V

SOLUTION STRUCTURE AND DNA–BINDING PROPERTIES OF WINGED HELIX DOMAIN OF THE MEIOTIC RECOMBINATION HOP2 PROTEIN

This chapter is dedicated to the work we published in Journal of Biological Chemistry in 2014. This was a collaborative work with the team led by Dr. Roberto J. Pezza at Cell Cycle and Cancer Biology Program, Oklahoma Medical Research Foundation. This work resulted the structure of the N–terminus of HOP2 protein, which was found to have a winged head DNA binding structure (PDB ID: 2MH2). The solution structure of the winged head DNA binding domain integrates biochemical and functional aspects of HOP2 recombinational function. The significance of this work is determining the 3D structure of HOP2 is crucial to understand the mechanism of HOP2 action (12).

5.1 Abstract

The HOP2 protein is required for efficient double–strand break repair, which ensures the proper synapsis of homologous chromosomes and normal meiotic progression. We previously showed that *in vitro* HOP2 shows two distinctive activities: when it is incorporated into a HOP2–MND1 heterodimer it stimulates DMC1 and RAD51 recombination activities, and the purified HOP2 alone is proficient in promoting strand invasion. The structural and biochemical basis of HOP2 action in recombination are poorly understood, therefore they are the focus of this work. Herein, we present the solution structure of the amino terminal portion of mouse HOP2, which contains a typical winged helix DNA binding domain. Together with NMR spectral changes in the presence

of double–stranded DNA, protein docking on DNA, and mutation analysis to identify the amino acids involved in DNA coordination, our results on the 3–D structure of HOP2 provide key information on the fundamental structural and biochemical requirements directing the interaction of HOP2 with DNA. These results, in combination with mutational experiments showing the role of a coiled–coil structural feature involved in HOP2 self–association, allow us to explain important aspects of the function of HOP2 in recombination.

5.2 Introduction

In meiosis, two rounds of chromosome segregation follow only one round of DNA replication. This serves a fundamental function by halving the number of chromosomes, which is required for sexual reproduction. The arrangement of maternal and paternal homologous chromosomes into pairs allows them to act as a single unit when microtubules attach and align them on the meiotic spindle. This ensures their orderly segregation to opposite poles of the cell at the first meiotic cell division so that each gamete receives only one copy of each chromosome. In most organisms homologous chromosomes pairs are stabilized by a physical link provided by the crossovers, the products of homologous recombination, which are viewed cytologically as chiasmata (reviewed in (4, 30-32)). Consequently, mutations that cause loss or misregulation of recombination are invariably associated with increased errors in meiotic chromosome segregation and the generation of aneuploid gametes.

Critical functions in homologous recombination are provided by the ubiquitous RAD51 and the meiotic specific DMC1 recombinases. These enzymes search for homologous DNA sequences to accomplish template repair by promoting the invasion of intact double–stranded DNA (dsDNA) by single–stranded DNA (ssDNA) ends arising from resection of dsDNA (33). Remarkably, efficient function of the recombinases requires interactions with accessory proteins. Several lines of evidence have demonstrated that HOP2 and MND1 are indispensable for meiosis via their interactions with DMC1 and RAD51 (16, 33-38). A *Saccharomyces cerevisiae HOP2* deletion mutant exhibits a profound failure in meiosis due to a uniform arrest at meiosis I with chromosomes engaged in synapsis with non–homologous partners (39). Moreover, *HOP2*$^{-/-}$ mouse spermatocytes show meiotic arrest and limited chromosome synapsis, consistent with a failure in double–strand break repair (40). Consistent with these data, others and we observed that the purified mouse HOP2–MND1 heterodimer physically interacts with DMC1 and RAD51, greatly stimulating the recombinase activities of these proteins (16, 34-38). An additional function for HOP2 as a recombinase, independent of DMC1/RAD51, has been proposed as purified HOP2 catalyzes the formation of homologous DNA pairing in meiotic recombination (35, 36, 41). Taken together, these results indicate that HOP2 may have dual role in mammalian meiotic homologous recombination (41).

Here, we present the NMR solution structure of an amino terminal fragment of HOP2 which reveals a DNA binding domain folded in a typical winged helix conformation, a common motif among the helix–turn–helix DNA binding proteins (42). In addition, we propose a model for the HOP2–DNA complex based on results obtained from chemical shift perturbations in the presence of DNA oligonucleotide, protein docking on DNA, and mutational analysis. Along with revealing the role of a coiled–coil structural feature in HOP2 involved in protein self–association, our results help explain important aspects of the molecular mechanism of recombination mediated by HOP2.

5.3 Materials and methods

Protein expression and purification – The plasmid pET22b was used as vector for the over–expression of mouse (His) 6–$^{1–84}$HOP2 in E. coli BL21 (DE3). The uniformly ^{15}N–labeled $^{1–84}$HOP2 sample was prepared by growing cells at 37°C to an A600 nm of 0.7. The culture medium (optimal M9 minimal) contained 1 g/l 15NH4Cl, 2 g/l mM unlabeled glucose, 50 μg/ml of kanamycin, 1X BME vitamins, 13 μM FeSO4, 0.5 mM MgSO4, and 1X trace elements. Protein expression was stimulated by addition of isopropyl β–D–galactosidase (IPTG) to a final concentration of 1 mM. After 18 hours, cells were harvested by centrifugation, lysed, and $^{1-84}$HOP2 was purified from the soluble fraction as previously described (38). The final yield of purified ^{15}N–labeled protein was approximately 6 mg per liter. The uniformly ^{13}C, ^{15}N–labeled $^{1-84}$HOP2 sample was produced according to the same procedure, apart from the fact that ^{13}C glucose was used at a concentration of 2 g/l. Purification of full–length, truncated versions, and point mutants of HOP2 used in DNA binding and chemical cross–linking experiments were performed as previously described (38).

NMR spectroscopy – A 450 µl sample of 6 mg/ml ^{13}C, ^{15}N–labeled $^{1-84}$HOP2 in a buffer of 120 mM NaCl, 10 mM imidazole, 5% glycerol, and 5% D_2O was used for NMR experiments. All NMR experiments were performed at 20 ºC on a Varian INOVA 600 MHz spectrometer with a Nalorac 5–mm ^{1}H, ^{13}C, ^{15}N PFG triple resonance probe and using software VNMRJ with the BioPack suite of pulse programs. ^{15}N–HSQC, HNCO, CO(CA)NH, HNCA, HN(CO)CA, HNCACB, and CBCA(CO)NH experiments were performed for protein backbone–assignment. ^{13}C–HSQC, H(CCO)NH and TOCSY–NHSQC, H(CC)H–TOCSY, C(CO)NH experiments were carried out for side chain assignments. NOESY–NHSQC was obtained for evaluation of structural quality (*vide infra*). Data were processed with NMRPipe (43) and peaks were picked using Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). The NMR peaks were first assigned using automatic assigning program PINE (44). The assignments were then verified or corrected manually.

Structure calculation – For structure resolution we first obtained ^{15}N–heteronuclear single quantum coherence (HSQC) spectrum of $^{1-84}$HOP2 (Fig. 1A). Peaks in the ^{15}N–HSQC spectra are well dispersed and the proton chemical shifts span a large range (7.8 to 9.8 ppm), indicating a well–folded protein. Besides the two prolines that do not contribute to ^{15}N HSQC signal, the first four residues, A12, and G66 were not observed, likely due to unfavorable dynamics. For the 82 non–proline residues, 74 backbone amide ^{15}N, ^{1}HN (90%), 78 Hα (95%, not double counting the two glycine protons), 76 C’ (93%), 78 Cα (95%), and 72 Cβ (88%) were assigned (Table 1).

Furthermore, 48 Cγ, 27 Cδ, 8 Cε, 114 Hβ, 62 Hγ, 35 Hδ and 23 Hε were assigned. Random coil index (RCI) order parameters (S^2) (45), secondary elements propensity, and backbone dihedral angle restraints were predicted from chemical shifts using program TALOS+ (46).

NMR structures were calculated using program CS–ROSETTA hosted on a web server at the Biological Magnetic Resonance Data Bank (BMRB), utilizing the protein backbone ^{15}N, ^{1}HN, ^{1}Hα, ^{13}Cα, ^{13}Cβ and ^{13}C' chemical shifts (47, 48). Unlike conventional NMR structure determination methods, CS–ROSETTA does not require any NOE distance restraints, which are often ambiguous and very cumbersome to obtain (49). Experimental chemical shifts are important to narrow down the selection of peptide fragments for building trial structures and to guide the effective sampling of the conformation space (47). CS–ROSETTA has been demonstrated to fold structures accurately for proteins up to 12 KDa using only chemical shifts (47, 49, 50). The side chain packing in structures determined by CS–ROSETTA has been found to be very accurate when compared to crystal structures, even though side chain constraints are not explicitly used in the procedure (47, 50). The side chain accuracy has been attributed to the ROSETTA all atom force field that includes hydrogen bonding, side chain packing, polar solvation, and backbone and side chain torsional energy (47). The energy–RMSD plot obtained shows convergence toward the lower–left corner with many low energy structures clustered with their Cα RMSD less than 2 Å from the lowest energy structure (Fig. 2A and Table 1). This indicates that the calculated final structures are highly reliable (47). Next, 1,000 lowest–energy structures were validated for consistence with NOE distances. This was done by calculating RPF DP scores for each of these structures against a manually refined peak list from an unassigned NOESY–NHSQC spectrum (49, 51). The DP scores of all 1,000 structures are found to be greater than 0.73 (Table 1), which confirms that the folds are correct and structures are of high quality. The twenty best structures were chosen to represent the final structural ensemble of the protein by setting the thresholds for a combination of CS–ROSETTA all–atom energy, RMSD, and DP scores (Table 1). The quality of these 20 structures was assessed using PROCHECK (52) and WHAT CHECK (53).

NMR measurements of protein–oligonucleotide interactions – For protein–DNA interaction studies, dsDNA (Table S1) was dissolved in the same buffer as the protein, reaching a concentration of 6.8 μg/μl. The control ^{15}N–HSQC NMR spectrum was first acquired for 450 μl of 5 mg/ml ^{15}N–labeled protein sample, followed by titration of the DNA solution at 10 μl increments for a total of 70μl. The ^{15}N–HSQC NMR spectra were acquired at each titration point and the scaled chemical shift perturbation was calculated using the equation, $\delta(i)=((\delta_{Ni}-\delta_{N0})^2+(10*(\delta_{Hi}-\delta_{H0}))^2)^{1/2}$, where, δ_{N0}, and δ_{H0} represent ^{15}N and ^{1}H chemical shifts for protein only and δ_{Ni}, and δ_{Hi} represent those at the ith titration point.

HOP2–dsDNA model – To model an interaction between HOP2 and DNA duplex, the apo–HOP2 structure was superimposed upon the BLAI plus DNA NMR structure (PDB accession number: 2P7C, (54)). The region of the HOP2 main chain path was then modeled on the BLAI structure from the beginning of helix 3 (residue 45) to conform to an optimal path for DNA binding. The joining residues (residues 36–44) between helix 2 of the N–terminal portion and helix 3 were modeled and minimized using the FALC–Loop modeling server (55).

Generation of $^{1–84}$HOP2 point mutant proteins – Polymerase chain reaction (PCR) based site–directed mutagenesis was used to generate single and multiple point mutations in $^{1–84}$HOP2 gene. Briefly, primers (oligonucleotide #5–#23, Table S1) were designed to replace the codon of a target amino acid by an alanine; PCR was used to amplify the mutant gene followed by assembly into the plasmid pRSFDuet1 using the Gibson assembly method (56). Transformation was done in XL2–Blue ultra–competent cells (Agilent Technologies). Point mutations were confirmed by DNA sequencing. Single point mutants (Q30A, K38A, K44A, K49A, D52A, E57A, E62A, K67A, Q68A, K69A, and Y71A) and triple point mutants (Q30A–K38A–K44A, Q30A–K44A–K49A, K63A–Y65–K67A, and Y65A–K67A–Q68A) were developed.

Gel shift assay measurements of protein–DNA interactions – The protein (concentration as indicated) was incubated with 720 nM (nucleotides) ^{32}P–labeled 30bp oligonucleotide in the following buffer: 20 mM Tris–HCl (pH 7.4), 2 mM $MgCl_2$ and 50 mM NaCl in a volume of 20 μl for 10 min at 37 °C. The samples were mixed with 3 μl of loading buffer (30% glycerol, 0.1% bromphenol blue) and analyzed by electrophoresis in 14% polyacrylamide gels in 1× TAE buffer at 5 V/cm for 5 h. The gels were exposed to radiation sensitive films and band corresponding to free dsDNA oligonucleotide was quantitated using the Image J program.

Chemical cross–linking experiments – The protein samples (15 μM $^{1–217}$HOP2, $^{1–84}$HOP2, $^{1–125}$HOP2, $^{126–217}$HOP2, and $^{144–217}$HOP2) were equilibrated for 10 min at room temperature in the presence of reaction buffer (30 mM HEPES (pH 7.4), 50 mM NaCl, and 10% glycerol). The samples were then further incubated with Ethylene glycol bis (succinic acid N–hydroxysuccinimide ester, EGS) (Sigma) for 20 min at room temperature in a total reaction volume of 20 μl. The cross–linkers were then quenched through the addition of 1 μl of 1 M Tris–HCl (pH 8) followed by incubation for 10 min at room temperature. Finally, the samples were resolved in 5–10 % and 15 % SDS–PAGE for full length and HOP2 fragments respectively.

5.4 Results and discussion

The HOP2 amino terminal DNA binding domain belongs to the winged helix family – The amino terminal domain (amino acids 1–84) structure of mouse HOP2 was solved using two– and three–dimensional NMR spectroscopy, making use of uniformly ^{13}C, ^{15}N–labeled proteins (Fig. 5.1A). We then used the backbone chemical shifts to estimate secondary structure (TALOS + (57)). As shown in Figure 5.1B (top panel), there are three α–helices (amino acids 12–23, 29–39 and 44–56) and two β–strands (amino acids 61–65 and 69–72). In addition, Y27 also has a propensity to be in β–strand conformation. In the determined three–dimensional structure (vide infra), both Y27 and S28 are in a short β–strand. The chemical shifts also allowed estimation of protein backbone mobility since the random coil chemical shifts are often associated with highly flexible regions (45). The predicted order parameters (S^2) indicate that the regions of definite secondary structure are well–defined, while both ends (M1–A10 and Q75–A84) show low degree of local order indicating significant sub–nanosecond timescale dynamics (Fig. 5.1B, lower panel).

The backbone and Cβ chemical shifts were utilized to calculate protein structure using the CS–ROSETTA software (Fig. 5.2A) (47). With the thresholds for a combination of CS–ROSETTA all–atom energy, Cα RMSD, and DP scores (energy ≤ –110, RMSD ≤ 0.415 Å, and DP ≥ 0.74) (Table 5.1), 20 structures were chosen to represent the final structural ensemble of HOP2 (Fig. 5.2B–D). The overall quality of these structures was assessed using PROCHECK (52) and WHAT CHECK (53). These structures are found to have a backbone RMSD value of 0.3 Å, and all atoms RMSD of 0.8 Å from the average structure. The structural Z–scores from WHAT CHECK procedure are good except for the backbone conformation, but much better RMS Z scores indicate that the local geometry of the structure is reliable (58). Additionally, all the Psi and Phi angles fall either in the most favored or additional allowed regions in the Ramachandran plot. The good geometry scores are partially due to the facts that structures obtained by CS–ROSETTA are built from homology peptide fragments augmented with chemical shifts and that these structures are not directly constrained by NOE distances as in conventional NMR methods.

Inspection of the final NMR ensemble (see Table 5.1 for statistics) reveals a well–ordered core region folded in a typical winged head/fork head structure (Fig. 5.2). This winged helix motif is a compact α/β structure consisting of two wings (W1 (Tyr 61–Gln 64) and W2 (Ala 69–Asn 72)), three α helices (H1 (Ala 8–Gln 19), H2 (Ala 25–Glu 35)), and H3 (Lys 40–Glu 53) and three β strands (S1 (Arg 21–Ser 24), S2 (Ile 56–Tyr 61), and S3 (Gln 64–Ala 69)), arranged in order H1–S1–H2–H3–S2–W1–S3–W2 (Fig. 5.2E). The amino terminal half of the motif is largely helical, whereas the carboxyl terminal half is composed of two of the three strands forming the twisted anti–parallel β sheet and the two loops of wings, W1 and W2 (42). Wing W1 connects strand S2 and S3, and wing W2 extends from strand S3 to the carboxyl terminal of the DNA binding domain. With some variations, the winged helix structure classifies HOP2 as a member of the winged helix protein family, which is a member of the DNA recognition helix–turn–helix super–family (42). Individual amino acids and the predicted secondary structure of the

HOP2 winged helix domain are highly similar among species (Fig. 5.2E). This indicates that the HOP2 winged helix DNA binding domain mode has been well–conserved throughout evolution.

NMR analysis of the interactions of $^{1–84}$HOP2 with DNA – We began analyzing the $^{1–84}$HOP2s DNA recognition mode by following changes on the chemical shift of HOP2 amino acids induced by interaction with DNA. Chemical shift perturbation is a sensitive method that can provide DNA binding location. ^{15}N–HSQC spectra of $^{1–84}$HOP2 were recorded with increasing amounts of a 30 bp dsDNA oligonucleotide (oligonucleotide #1 and #2, Table 5S1). Since the ^{15}N–HSQC spectra of DNA–free $^{1–84}$HOP2, and of dsDNA–bound proteins share many similarities (Fig. 5.3A), the amide ^{1}H and ^{15}N chemical shifts of the amino terminal domain of DNA–bound $^{1–84}$HOP2 were assigned with reference to the DNA free protein assignments. This global comparison also indicates that $^{1–84}$HOP2 does not undergo major structural rearrangements upon binding to dsDNA. The ^{15}N, ^{1}H chemical shift changes (perturbation defined in the experimental section) for backbone amide groups show the highest shift or intensity changes in helix H3 (amino acids Ala 45, Ala 46, Lys 49 and Asp 52), and in the β sheet, especially close to the wing W1 (amino acids Gln 68 and Tyr 71) (Fig. 5.3A–C). The residues affected by dsDNA binding (highlighted on the HOP2 structure in Fig. 5.3C) define a possible DNA–binding surface located on a single side of the HOP2 winged helix domain. Similar regions are typically implicated in coordinating DNA for a large number of the winged helix family members (42). These results suggest the importance of H3 and W1 structure in DNA binding. The titration curves of perturbation versus DNA concentration were analyzed for the 10 most perturbed sites (data not shown), showing that the perturbation values are saturated at the final few titration points. The apparent dissociation constant that gives half–maximum chemical perturbation Kd=33±10 μM.

HOP2 winged helix domain DNA recognition mode – The N–terminal of HOP2 has been shown to be required for efficient formation of the DMC1–mediated synaptic complex (38). Here, we investigate the DNA recognition mode of $^{1–84}$HOP2 by comparing the 3D structure of HOP2 with protein–dsDNA complexes reported for other winged helix proteins. We first compared the structure of $^{1\text{–}84}$HOP2 with other winged helix proteins structures obtained in the absence of DNA using the VMD software package (59). We included in our analysis eight winged–helix proteins (BlaI, Ahrc, Fox04, LexA, FOX03a, PhoB, Genesis, and WRN) whose structures in the presence or absence of dsDNA have been studied by X–ray crystallography, NMR, or computational tools (Fig. 5.4A and B) (60-67). The most similar 3D structure is that of the BlaI DNA recognition domain (RMSD: 1.62 Å, QH: 0.72, % identity: 10.84). Although no significant sequence homology could be detected by primary sequence similarity searches, BlaI presents a fold very similar to that of $^{1\text{–}84}$HOP2 (Fig. 5.4A). Notably, the length of the helices and the loop in the wing regions are similar and the angles between the three helices are conserved. We propose a model for the interaction between the $^{1\text{–}84}$HOP2 winged head domain and dsDNA by comparison with the structure of the BlaI–dsDNA complex (Fig. 5.4C and D), as described in Methods. Although this method is unable to resolve the precise position of the amino acids and the base pairs contacts, it suggests that the HOP2 winged helix domain uses a canonical mode for DNA binding (42). In this mode, helix H3, called the recognition helix, is presented to the major groove of the DNA. The wings (in particular W1) and the amino terminal part also make contacts with the minor groove of the DNA. This has been observed in the majority of winged helix proteins (reviewed in (42)) and is unlike RFX1, which makes most of the contacts with the DNA major groove via wing W1 with H3 overlying the minor groove (68).

Effect of point mutations on DNA binding by the $^{1\text{–}84}$HOP2 fragment – Results of the chemical shift perturbation in the presence of dsDNA and the protein docking on dsDNA suggest that H3 and W1 structures have functional importance in DNA binding. Prompted by these observations and further to dissect the complex's DNA binding sites, we constructed $^{1\text{–}84}$HOP2 point mutants to be deficient for DNA binding. $^{1\text{–}84}$HOP2 shows a marked preference for dsDNA (apparent Kd (dsDNA): 32μM) (Fig. 5.4F); therefore, we tested the effect of these mutations on HOP2–dsDNA complex formation. We began by identifying conserved regions within the HOP2 winged head domain enriched in basic amino acids, which may be involved in DNA engagement via ionic interaction, and aromatic residues, predicted to participate in stacking interactions with DNA bases. We replaced the highly conserved K63, Y65 and K67 residues encompassing a β–hairpin at W1 to alanine. The KYK mutation has a significant effect on the DNA binding ability; however, association of $^{1-84}$HOP2 to DNA is not abolished (Kd: 52μM) (Fig. 5.4E). This is consistent with previous DNA mutational studies showing that a full–length HOP2 mutant carrying three point mutations within the winged head domain had substantially reduced DNA binding ability (38). Similar results were obtained with a HOP2 mutant in which Y65, K67 and Q68 amino acids were replaced by alanine (not shown). These results indicate that the W1 region participates in DNA binding and that at least one more region of $^{1-84}$HOP2 is involved in dsDNA binding. To test the participation of H3 and H2 in DNA binding we generated two sets of mutants. In one, we changed the highly conserved K44 and K49 (located in the helix H3) and Q30 (located in the helix H2) to alanine (Q30K44K49 mutant). For the second, the amino acids Q30, K38 and K44 were changed to alanine (Q30K38K44 mutant). Q30K38K44 shows a substantial decrease in dsDNA binding (Kd: 58μM) and the Q30K44K49 mutant shows a clear impairment in DNA binding (only 30% of dsDNA substrate bound at 84 μM protein) (Fig. 5.4E). These results indicate the relative importance of helix H3 with notable participation of the highly conserved K49 amino acid in engaging DNA (Fig. 5.4C). Single amino acid changes for any of the mutants described above resulted in wild–type levels of DNA binding.

Charged residues distributed over the surface of HOP2 could potentially act as loci for DNA binding. The GRASP program (69) was used to calculate an accessible surface area and a corresponding surface potential map of the $^{1-84}$HOP2 structure. The protein area proposed to bind DNA can be superimposed nicely with a charged region of the protein electrostatic surface (Fig. 5.4D). The residues K44, K49 and K67 overlap with high potential areas within the DNA binding surface. This is additional evidence suggesting the involvement of these amino acids in coordinating DNA.

In summary, our results are in good agreement with a canonical winged helix domain DNA interaction mode in which helix H3 and W1 are the major responsible structures coordinating dsDNA binding.

HOP2 self–association is mediated by a coiled–coil structure – Amino acid sequence and secondary structure analysis of full–length mouse HOP2 predicts the presence of two coiled–coil structures encompassing amino acids 84–124 and 126–155, respectively (Fig. 5.5A and B). The formation and location of these coiled–coil motifs are highly conserved from yeast to human. We previously showed that mutations affecting only 126–155 but not the N–terminal (amino acids 84–124) coiled–coil structure disrupt the formation of the HOP2–MND1 heterodimer (36). Here, we show that the 126–155 coiled–coil structure is involved in oligomerization of HOP2. We determined the oligomeric composition of full–length (amino acids 1–217) and truncated versions (amino acids 1–84, 1–125, 126–217 and 144–217, (38)) of HOP2 using chemical cross–linking. We observed that full–length HOP2 forms dimers, tetramers and a minor amount of high order oligomers (Fig. 5.5C). Disruptions within the 133 amino acids from the carboxyl terminal of HOP2 ($^{1–84}$HOP2) abolished the formation of dimers and tetramers (Fig. 5.5D). This indicates that the winged head amino terminal region of HOP2 does not promote oligomerization alone. The $^{1–125}$HOP2 mutant showed a low level of dimer formation only at high concentrations of cross linker, indicating that the 84–124 coiled–coil structure has a minor role in HOP2 self–association. In contrast, the $^{126–217}$HOP2 truncation mutant, which contains the 126–155 coiled–coil domain, formed a high level of dimers and tetramers even at lower concentrations of cross linker. This characteristic is strongly affected by protein concentration and ionic strength (data not shown). The role of this presumptive coiled–coil structure in HOP2 oligomerization is confirmed by the reduced levels of oligomers observed in the $^{144–217}$HOP2 mutant, which removes the first 19 amino acids of the domain.

5.5 Conclusions

We and others previously showed that *in vitro* the mouse HOP2 protein show two distinctive activities: (1) When it is incorporated into a HOP2–MND1 heterodimer, it stimulates DMC1 and RAD51 recombination activities (16, 35, 36, 38, 41). (2) HOP2 alone is proficient in promoting strand invasion (35, 36, 41). At present, the structural base supporting these HOP2 biochemical actions is unknown and therefore limits the mechanistic understanding of HOP2 recombination function. In this work, we present the solution structure of the mouse $^{1-84}$HOP2 fragment using NMR techniques. Our results reveal that the HOP2 amino terminal domain displays a compact α/β structure arranged in a typical winged–head DNA binding domain with order H1–S1–H2–H3–S2–W1–S3–W2. This classifies HOP2 as a member of the winged helix protein family of DNA binding proteins (42).

The winged helix proteins constitute a subfamily within the large ensemble of helix–turn–helix proteins. Since the discovery of the winged helix motif (70), different groups of topologically related proteins with diverse biological functions have been characterized. In the most common (canonical) mode of DNA binding, the DNA interactions, including any specificity determining contacts, map to the recognition helix (H3) within the DNA major groove, with other less important interactions involving the loop W1 and the minor groove of the DNA substrate. By NMR solution structure, biochemical, bioinformatics and mutagenic analyses, we provide evidence that helix H3 and the loop W1 of HOP2 are involved in DNA recognition.

Interestingly, a large number of winged helix proteins exhibiting the canonical recognition mode show DNA binding sequence specificity. For example, the BlaI repressor regulates the expression of antibiotic resistance proteins by specific contacts occurring between the base–pairs of the TACA motif in the DNA operator sequence and a conserved amino acid residue of the repressor helix H3 (60). Based on the canonical DNA binding mode of HOP2 and high similarity of the HOP2 winged helix structure with sequence specific DNA binding proteins such as BlaI and a number of other transcription factors (Fig. 5.4B and C), it is tempting to suggest that HOP2 may also have preference for specific DNA sequences. However, further studies, such as the analysis of HOP2 DNA binding sites at specific genomic location or the testing of oligonucleotide library sequences for DNA binding will be needed to test this possibility.

DNA recognition by winged helix proteins may be affected by the oligomerization state of the protein. For example, two molecules of the MotA transcription factor bind cooperatively to their DNA consensus sequence (71). In this case, the formation of the active dimer DNA binding form is mediated by the hydrophobic dimer interface of a coiled–coil interaction. Reminiscent of this DNA binding mode, our results show the importance of a carboxyl terminal coiled–coil structure (amino acids 126–155) but not of the amino terminal coiled–coil domain in HOP2 self–association (Fig. 5.5). This is in agreement with our previous results showing that this structure is required for the interaction of HOP2 and MND1 (36). Notably, recent work in the mouse MND1 protein indicated the presence of a highly conserved winged–head domain required for efficient HOP2–MND1 complex binding to DNA (38). Taken together these results suggest that a coiled–coil structural feature facilitates the formation of an active HOP2–MND1 and HOP2–HOP2 dimer DNA binding form. We speculate that two winged–head domains arranged in either a parallel or antiparallel mode cooperate for efficient DNA binding or may regulate HOP2 DNA binding to certain genomic regions.

Previous biochemical studies have shown that, when incorporated into the HOP2–MND1 complex, HOP2 is critical for efficient homologous DNA pairing mediated by the DMC1 and RAD51 recombinases (16, 34-37). This stimulatory function can be explained by HOP2–MND1 acting in two critical steps of recombinase–promoted homologous pairing (16, 37); one, by stabilizing the DMC1–ssDNA–nucleoprotein filament and second, by facilitating the conjoining of DNA molecules through the capture of dsDNA by the DMC1–ssDNA nucleoprotein filament. Our results showing the solution structure and biochemical properties of a major DNA binding site for HOP2 provide structural support for a recently proposed model in which the N–terminal DNA binding domain of HOP2, which preferentially binds dsDNA, is positioned at a distance to the DMC1–ssDNA nucleoprotein filament to capture dsDNA (38). These HOP2 functions help assemble synaptic complex and permit the efficient sampling of potential homologous DNA sequences.

DATA BANK ACCESSION NUMBERS

Chemical shift assignments of $^{1\text{–}84}$HOP2 have been deposited with BioMagResBank (accession number 19613). The coordinates of the twenty refined structures have been deposited in the Protein Data Bank under the accession code 2mh2.

ACKNOWLEDGEMENTS

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number 1P20GM103636 and by the OCAST grant HR10–48S (to R. J. P.). The authors greatly appreciate financial support from National Institutes of Health (GM097713 to D.H.Z) and Oklahoma Center for the Advancement of Science and Technology (HR12–050 to D.H.Z). We thank Dr. Margaret Eastman (Oklahoma State–wide Shared NMR Facility) for technical assistance.

FIGURE LEGENDS

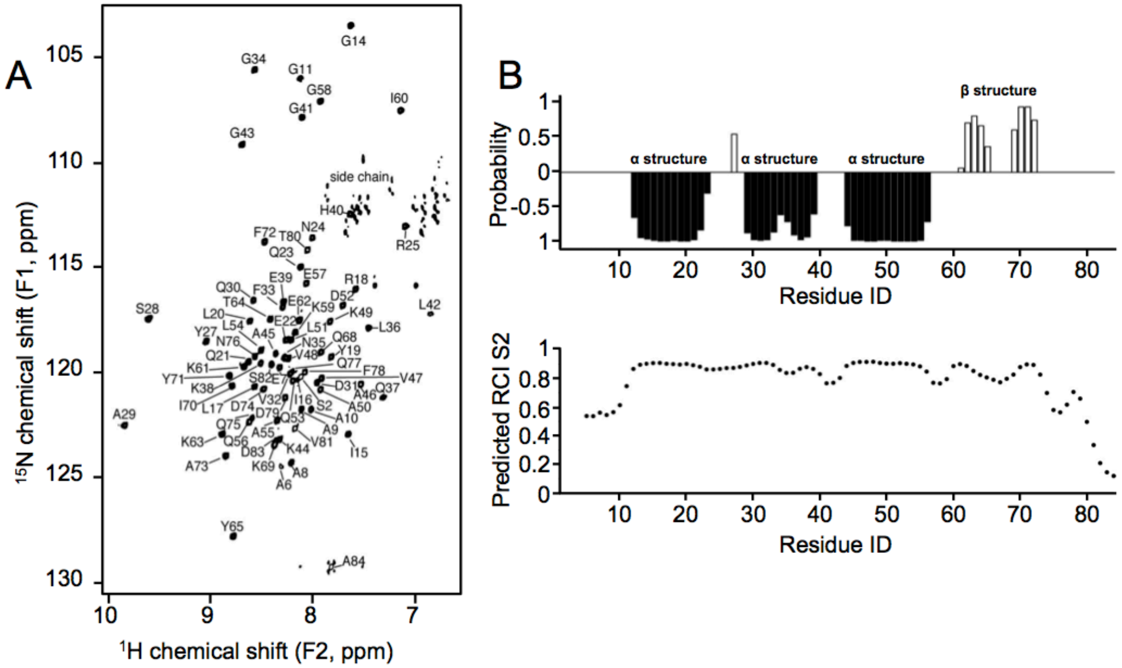


FIGURE 5.1. ^{15}N–HSQC NMR spectrum and secondary structure analysis of $^{1–84}$HOP2. A. Peaks corresponding to chemical shifts were assigned with a set of three–dimensional NMR experiments. A typical ^{15}N–HSQC spectrum is shown. B. Secondary structures and order parameters predicted using assigned chemical shifts. Negative probabilities indicate propensity for α–helical structures, and positive probabilities indicate β–strands (top panel). Chemical shift estimation of protein backbone mobility for $^{1–84}$HOP2 (lower panel).

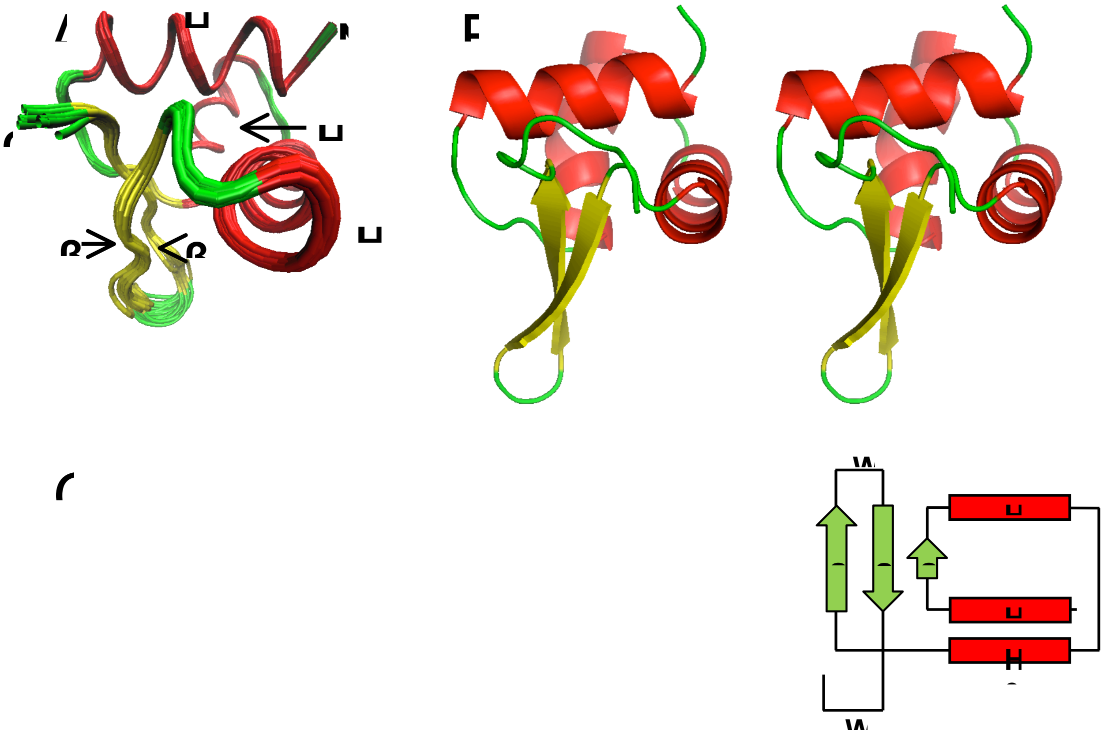


FIGURE 5.2. NMR structure of $^{1–84}$HOP2. A. NMR structure determination of $^{1–84}$HOP2. Plot of chemical–shift–rescored ROSETTA all–atom energy versus Cα RMSD from the lowest energy structure showing convergence of structure calculation. B. Diagram of superimposed 20 refined low–energy structures. C, carboxyl terminal; N, amino terminal; H, helix; β, beta. Secondary structures are colored: Helixes in red, β–sheets in yellow and loops in green. C. Cartoon diagram of the lowest–energy structure. Secondary structures are colored as in B. D. Backbone atom representation of $^{1–84}$HOP2. Carbon, nitrogen and sulfur, grey; hydrogen, blue; and oxygen, red. E. Alignment of the HOP2 amino terminal primary sequence and topology of a typical winged helix fold. Highlighted in red are conserved and in green partially conserved amino acids. The secondary structure prediction for the mouse sequence was obtained using the “nnpredict program” (http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html); H, α–helix. S represents β strands and W1 and W2 are loops or wings.

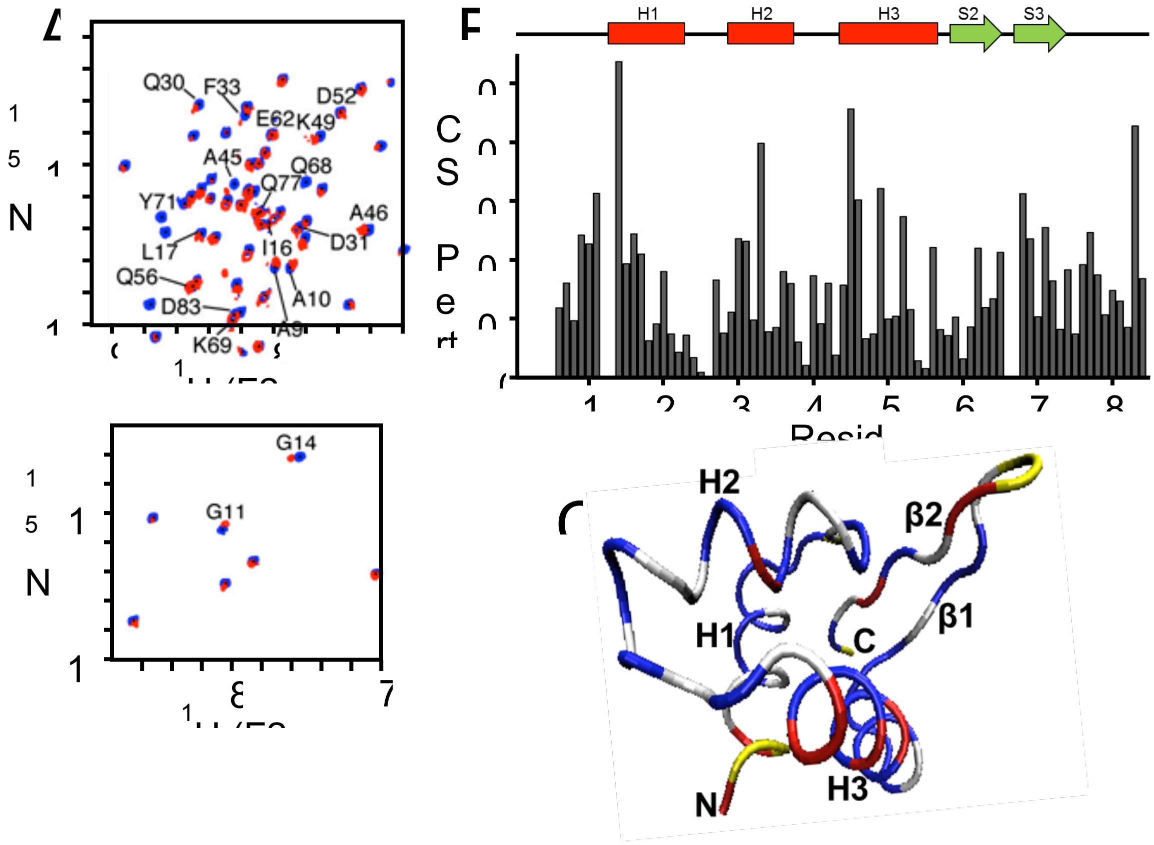


FIGURE 5.3. Chemical shift mapping of $^{1\text{–}84}$HOP2 interacting with DNA. A. DNA titration monitored by ^{15}N–HSQC, showing two regions with significant peak displacements. Blue: initial, Red: final. B. Graph of the chemical shift variations between the DNA free $^{1\text{–}84}$HOP2 and the $^{1\text{–}84}$HOP2/DNA complex plotted against residue number. C. Chemical shift perturbation color coded on the structure of $^{11\text{–}74}$HOP2. Blue: $\delta < 0.15$, gray, $0.15 \leq \delta < 0.25$; red $\delta > 0.25$; yellow: missing due to prolines or unfavorable dynamics. The table contains a list of amino acids involved in the chemical shift perturbation by dsDNA.

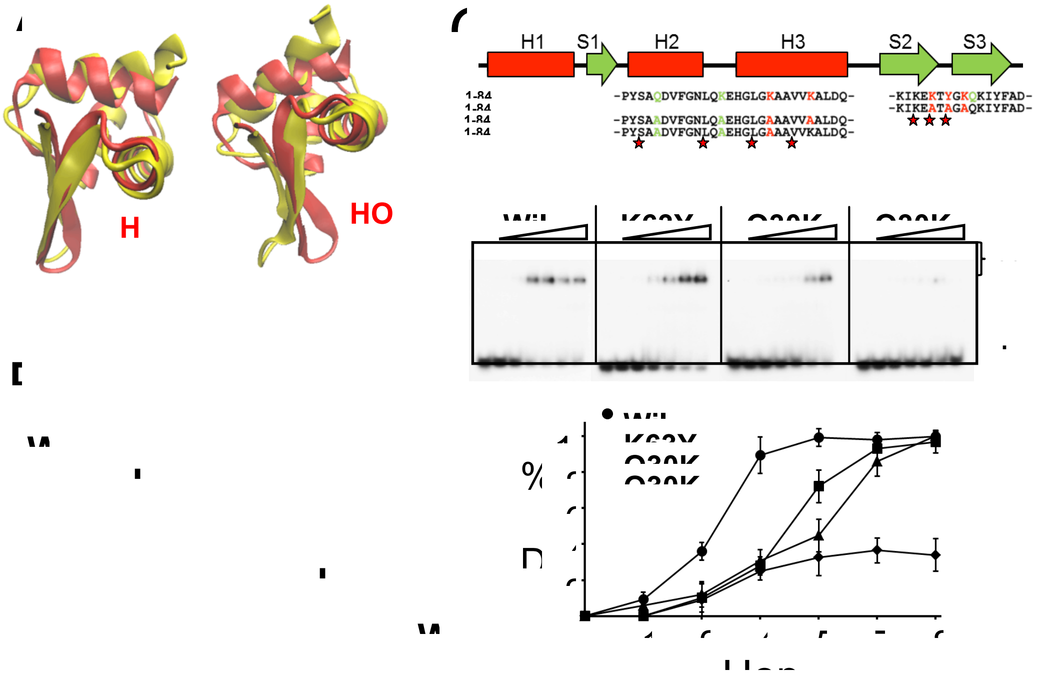


FIGURE 5.4. DNA recognition by $^{1\text{–}84}$HOP2. A. 3D similarity of the NMR structure of $^{1\text{–}84}$HOP2 and the crystal structure of BlaI obtained in the absence or presence of dsDNA. B. Results of 3D comparison of $^{1\text{–}84}$HOP2 and other winged helix proteins. Protein ID corresponds to the protein structure without DNA. C. Surface representation of $^{1\text{–}84}$HOP2–dsDNA complex model. A 30 bp dsDNA oligonucleotide corresponding to the *B. licheniformis* BlaI operator op1 (oligonucleotide #3 and #4, Table S1) was used. Amino acids highlighted in green correspond to helixes H2 and H3 and in orange to β1 and wing 1. D. Electrostatic potential surface (GRASP) representation of $^{1\text{–}84}$ HOP2–dsDNA complex model. Highlighted are amino acids with high potential (5.1 maximum) in blue and low potential (–2.0 minimum) in red. E. DNA binding properties of $^{1\text{–}84}$HOP2 point mutants. The top panel shows the mutants generated in this study. Analysis of wild–type (1–84 amino acids) and mutant variants of $^{1\text{–}84}$HOP2 (K63Y65K67, Q30K38K44, Q30K44K49) (0, 14, 28, 44, 56, 75 and 84 μM) for DNA binding. The mean values ± standard deviations from three independent experiments were plotted (lower panel). F. $^{1\text{–}84}$HOP2 show high DNA binding preference for dsDNA.

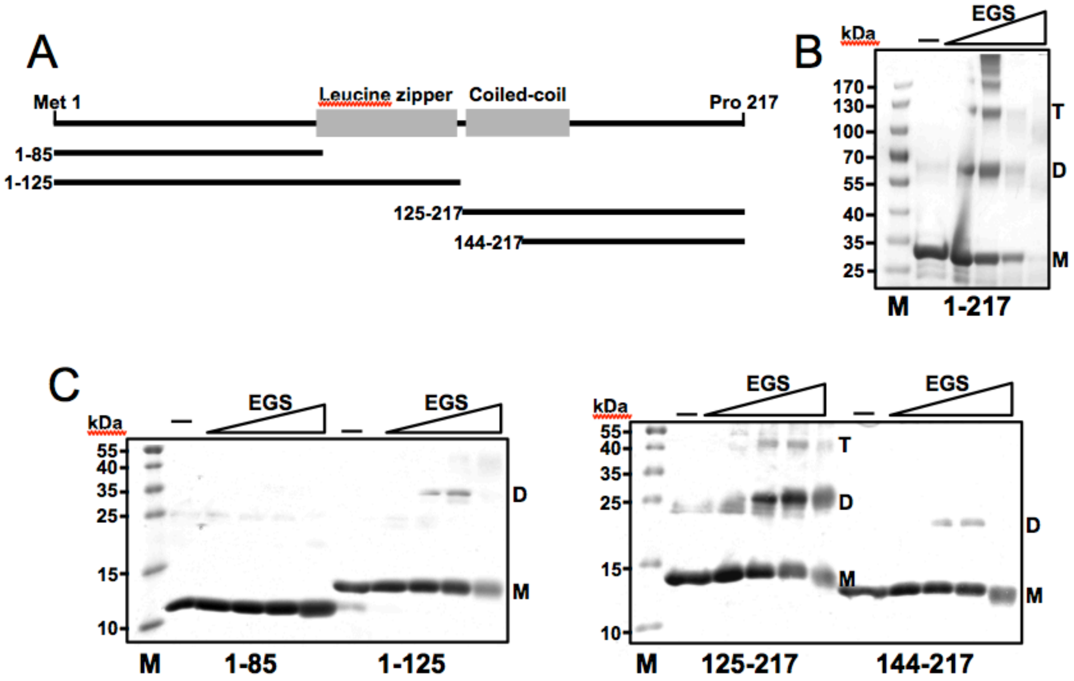


FIGURE 5.5. A coiled–coil structure in the central region of HOP2 mediates protein self–association. A. Scheme of wild–type, carboxyl– and amino–terminal mutants HOP2 used in this study. B. Sequence of HOP2 putative coiled–coil domain and coiled–coil prediction plot by the Multicoil software. C. Products of chemical crosslink of full–length HOP2. D. Products of chemical crosslink of 1–84, 1–125, 126–217 and 144–217 HOP2 truncation mutants. M, monomer; D, dimer; T, tetramer. EGS, chemical cross linker.

TABLE 5.1: Inputs for CS-ROSETTA and structure validation for HOP2 (residues 11-74).

Chemical shifts used in CS-ROSETTA	
Cα	62
Cβ	56
C'	60
N	59
HN	59
Ha (count one for each glycine)	62

Manually refined NOESY-NHSQC peaks	1097
1,000 lowest energy models (out of 40,000)	
CS-ROSETTA scaled energy	-106.1 (2.7)
Cα RMSD	0.8 (0.3)
RPF DP score	0.744 (0.005)
20 models	
CS-ROSETTA scaled energy	-112.7 (2.2)
Cα RMSD	0.3 (0.1)
RPF DP score	0.745 (0.004)
RMSD from average coordinates (Å)	
Backbone atoms	0.28 (0.08)
All atoms	0.8 (0.1)
Ramachandran statistics[a]	
Most favored regions (%)	97.2 (1.1)
Additional allowed regions (%)	2.8 (1.1)
Generously allowed (%)	0 (0)
Disallowed regions (%)	0 (0)
G-factors[a]	
Phi-psi	0.29 (0.03)
All dihedrals	0.42 (0.02)
Covalent	0.68 (0.00)
Overall	0.52 (0.01)

[a] From PROCHECK.

TABLE 5S1: Oligonucleotides used in this study. FWD, forward. REV, reverse.

Primer name	Sequence
#1-DNA chemical perturbation_FWD	TAGTAGTATGAATTTGCCGCAAGATCTGAT
#2-DNA chemical perturbation_REV	ATCAGATCTTGCGGCAAATTCATACTACTA
#3-BlaI operator op1_FWD	GAAAGTATTACATATGTAAGATTTAAATGC
#4-BlaI operator op1_REV	GCATTTAAATCTTACATATGTAATACTTTC
#5-Q30A_FWD	ATCATCCTGAGGTACCTGCAAGAACAAAACCGGCCCTACAGCGCCGCGGACGTG
#6-K38A_FWD	GGACGTGTTCGGAAACCTACAGGCGGAACATGGACTGGGCAAGGCGGCGG
#7-K44A_FWD	CTACAGAAGGAACATGGACTGGGCGCGGCGGCGGTAGTGAAGGCGCTGGATC
#8-K49A_FWD	GGACTGGGCAAGGCGGCGGTAGTGGCGGCGCTGGATCAGCTGGCCCAGG

#9-D52A_FWD	GGCGGTAGTGAAGGCGCTGGCTCAGCTGGCCCAGGAAGGCAAG
#10-E57A_FWD	GAAGGCGCTGGATCAGCTGGCCCAGGCAGGCAAGATCAAAGAGAAG
#11-E62A_FWD	GCCCAGGAAGGCAAGATCAAAGCGAAGACCTACGGCAAGCAGA
#12-K67A_FWD	CAAAGAGAAGACCTACGGCGCGCAGAAAATTTATTTTGCC
#13-Q68A_FWD	CAAGAAGACCTACGGCAAGGCGAAAATTTATTTTGCCGATCATCAGAAC
#14-K69A FWD	AGAAGACCTACGGCAAGCAGGCAATTTATTTTGCCGATCAGAA
#15-Y71A_FWD	GACCTACGGCAAGCAGAAAATTGCTTTTGCCGATCAGAACCAG
#16-Q30A_K44A_FWD	CTACAGAAGGAACATGGACTGGGCGCGGCGGCGGTAGTGAAGGCGCTGGATC
#17-Q30A_K38A_FWD	GGACGTGTTCGGAAACCTACAGGCGGAACATGGACTGGGCAAGGCGGCGG
#18-Y65A_K67A_FWD	GGCAAGATCAAAGAGAAGACCGCCGGCGCGCAGAAAATTTATTTTGCCGAT
#19-Q30A_K38A_K44A_FWD	TACAGGCGGAACATGGACTGGGCGCGGCGGCGGTAGTGAAGGCGCTGGAT
#20-Q30A_K44A_K49A_FWD	CTGGGCGCGGCGGCGGTAGTGGCGGCGCTGGATCAGCTGGCCCA
#21-K63A_Y65A_K67A_FWD	CAGGAAGGCAAGATCAAAGAGGCGACCGCCGGCGCGCAGAAAATTTATTT
#22-Y65A_K67A_Q68_FWD	GATCAAAGAGAAGACCGCCGGCGCGGCGAAAATTTATTTTGCCGATCAG
#23-gRSFDuet_REV	GGTTTCTTTACCAGACTCGAGCCTCAGGGGTCGGGGAGCAAAACATTATG

CHAPTER VI

WING 1 OF PROTEIN HOP2 IS AS IMPORTANT AS HELIX 3 IN DNA BINDING BY MD SIMULATION

This chapter is dedicated to molecular dynamics simulation work that was published in Journal of Biomolecular Structure and Dynamics in 2017. Through MD simulations, we found that wing 1 of protein HOP2 plays equally important role as Helix 3 in DNA binding. The findings of this work help understand the role of HOP2 protein in meiotic recombination (72).

6.1 Abstract

The repair of programmed DNA double–strand breaks through recombination is required for proper association and disjunction of the meiotic homologous chromosomes. Meiosis specific protein HOP2 plays essential roles in recombination by promoting recombinase activities. The N–terminal domain of HOP2 interacts with DNA through helix 3 (H3) and wing 1 (W1). Mutations in wing 1 (Y65A/K67A/Q68A) slightly weakened the binding but mutations in helices 2 and 3 (Q30A/K44A/K49A) nearly abolished the binding. To better understand such differential effects at atomic level, molecular dynamics simulations were employed. Despite losing some hydrogen bonds, the W1–mutant DNA complex was rescued by stronger hydrophobic interactions. For the wild type and W1–mutant, the protein was found to slide along the DNA grooves as the DNA rolls along its double–helix axis. This motion could be functionally important to facilitate the precise positioning of the single–stranded DNA with the homologous double–stranded DNA. The sliding motion was reduced in the W1–mutant. The H–mutant nearly lost all intermolecular interactions. Moreover, an additional mutation in wing 1 (Y65A/K67A/Q68A/K69A) also caused

complete complex dissociation. Therefore, both wing 1 and helix 3 make important contribution to the DNA binding, which could be important to the strand invasion function of HOP2 homodimer and HOP2–MND1 heterodimer. Similar to cocking a medieval crossbow with the archer’s foot placed in the stirrup, wing 1 may push the minor groove to cause distortion while helix 3 grabs the major groove.

6.2 Introduction

Meiotic recombination is initiated by double–strand DNA breaks (DSBs) at multiple sites of chromosomal DNA. Processing the DNA ends of DSBs by exonucleases generates 3’ single–stranded DNA tails. Two recombinases, RAD51 and DMC1, then form nucleoprotein filaments on single–stranded DNA and search for DNA homologous sequences leading to the invasion of the intact homologous sequence. Efficient action of Dmc1 and Rad51 requires the assistance of auxiliary proteins such as HOP2 and MND1 (13-15). Indeed, a heterocomplex formed by HOP2 and MND1 physically interacts with DMC1 and RAD51 to stimulate DNA strand exchange, which is required for successful progression of homologous recombination (15, 73). Importantly, HOP2 by itself, independent of MND1 and DMC1/RAD51, is able to work as a recombinase in strand invasion and synaptic complex formation (7, 17).

There are several proteins or protein complexes involved in the DMC1–mediated strand invasion, but their molecular mechanisms are yet to be fully understood (74). We have recently solved the solution structure of N–terminal DNA binding domain of mouse HOP2 (residues 1–84), which provided an initial structural basis for understanding its biochemical functions (12). This structure has been identified with winged–helix DNA–recognition structural motifs (12) belonging to the helix–turn–helix superfamily (75). It has a compact alpha/beta structure consisting of two wings, three alpha helices, and three beta strands, arranged in the order H1–S1–H2–H3–S2–W1–S3–W2 (12).

The crystal structure of HOP2–MND1 heterodimer of *G. lamblia* has been recently solved by Oh and coworkers (73). The winged helix domains of HOP2 and MND1 were found juxtaposed together in a fixed orientation via hydrophobic and salt bridge interactions. This juxtaposition was believed to introduce a large distortion in double–stranded DNA, initiating the strand opening stage during homologous recombination (73). This work is certainly a key step to understand the molecular action of HOP2–MND1 as a heterodimer during homologous recombination.

The importance of both the C–terminal coiled coil region and the N–terminal winged helix domain of HOP2 during homologous recombination has been reported (12, 17, 73). NMR chemical shift perturbation analysis on the winged–helix domain of HOP2 has shown that it binds to double–stranded DNAs with considerable affinity and that helix 3 and wing 1 are involved in DNA binding (12). The functional importance of helix 3 and wing1 in DNA binding has been further confirmed by site–directed mutagenesis studies (12). Point mutations in wing 1 (Y65A/K67A/Q68A) slightly reduced the DNA binding. In comparison, there was a clear impairment in binding when sites in helices H2 and H3 (Q30A/K44A/K49A) were mutated (12).

These results have been very important to understand the significance of the N–terminal domain of HOP2 and the mechanism by which HOP2 and MND1 work as a complex. However, how HOP2 alone is capable of promoting strand invasion still remains elusive. HOP2 alone is believed to function as a homodimer (17). Although, several studies confirmed the involvement of helix 3 and wing 1 in DNA binding, their contributions in DNA binding have not been investigated at atomic level. Moreover, it is not clear whether different DNA binding abilities of the two mutants indicate that helix 3 is more important than wing 1 in DNA binding. Here we employed classical molecular dynamics simulations on DNA complexes of HOP2 winged–helix domain and its mutants to study the differential effects of point mutations in DNA binding, and to identify functionally important dynamics.

6.3 Methodology

6.3.1 System preparation

The NMR structure of N–terminal HOP2 (PDB code 2MH2), which consists of the well–structured residues 11–74, was taken as a starting structure. To prevent the terminal amine and carboxylic acid groups from forming artificial bonds with the DNA, the N–terminus and C–terminus were acetylated and amidated, respectively, using program VMD (25). This structure was first energy minimized in vacuum and then equilibrated in water for 20 ns using GROMACS (23) with CHARMM36 (76) sets of force fields (see details in a subsequent method section, *Equilibration and molecular dynamics*). B–form structure of a 30–mer double–stranded DNA with sequence ^{1}TAGTAGTATGAATTTGCCGCAAGATCTGAT^{30}, which was identical to the one used in the previous NMR study (12), was generated using 3DNA web server (77) and equilibrated in water for 20 ns. These equilibrated protein and DNA structures were docked using ZDOCK web server (78), which was guided with interacting protein residues (Gly11, Gly14, Phe33, Ala45, Ala46, Lys49, Asp52, Gln68, Tyr71) identified from the NMR chemical shift perturbation study (12). The docked protein–DNA complex (Figure S1A) had helix 3 in the DNA major groove and wing 1 in the minor groove, which is characteristic of the winged–helix domain DNA–binding proteins such as BlaI (79). After successful docking, DNA bases far from the docking site were removed, resulting in a 14–mer with sequence ^{6}GTATGAATTTGCCG^{19} to speed up simulations and analyses. The truncated DNA and the docked protein were taken as initial structures for subsequent MD simulations.

VMD (25) mutator was used to generate two protein mutants. The W1–mutant had residues Tyr65, Lys67, and Gln68 in wing 1 mutated to alanines, and the H–mutant had residues Gln30, Lys44, and Lys49 in H2 and H3 mutated to alanines. Both mutants were equilibrated in water for 20 ns. These equilibrated mutant structures maintained the folded structure of the wild type, with backbone root–mean–square deviation (RMSD) of 0.15 nm and 0.11 nm for the W1– and H–mutants, respectively, when compared to the wild type. The mutant structures were aligned with the protein in wild–type protein–DNA complex system using VMD to obtain similar configurations. The coordinates were then saved to obtain the initial mutant–DNA complex systems. These three protein–DNA complex systems as well as the three free proteins were subjected to complete MD simulations and analyses.

6.3.2 Solvation and energy minimization

Each system was placed at the center of a rectangular box with a distance of at least 1.5 nm from the wall. Periodic boundary conditions were imposed to mimic the infinite system to avoid real phase boundaries and to minimize finite boundary effects (23). All systems were solvated with pre–equilibrated water molecules with Na^+ and Cl^- ions added to maintain overall salt concentration of 0.120 M. All systems were energy minimized using steepest descent algorithm with 1fs time steps and found to converge to a minimum energy with forces less than 100 kJ/mol/nm (23).

6.3.3 Equilibration and molecular dynamics

All molecular dynamics simulations were performed on a Linux cluster supercomputer (using 30 out of 252 standard compute nodes, each with dual Intel Xenon E5–2620 “Sandy Bridge” hex core 2.0 GHz CPUs and 32 GB of 1333 MHz RAM) using software GROMACS 4.5.5 (23) and AMBER03 sets of force field with SPCE water model (80). Two equilibration phases were carried out, each with 1 fs time steps. In the first constant volume (NVT) ensemble phase, each system was strongly coupled to a temperature bath using V–rescale coupling (81) (coupling constant of τ_T = 0.1 ps) for 0.05 ns to maintain system temperature at 293 K. In the second constant pressure (NPT) phase of 5 ns, a strong Parrinello–Rahman pressure coupling (τ_P = 5.0 ps) (82) was used to maintain the pressure isotropically at 1 bar and a weak Nose–Hoover temperature coupling (τ_T = 0.5 ps) was used to ensure a true NPT ensemble (83-85). The total of 5.05 ns equilibration was followed by a 100 ns production run of molecular dynamics in 2 fs step size, during which temperature and pressure were maintained using weak coupling methods (Nose–Hoover τ_T = 0.5 ps and Parrinello–Rahman τ_P = 3 ps). A leap–frog integrator (86) was used during the simulation. The non–bonded van der Waals interactions were estimated using Lennard–Jones potential, and electrostatic forces and energies were calculated using Particle–Mesh Ewald (PME) summation algorithm (87). Unless specified, all the cut–off distances for these interactions were set at 1.2 nm. The bonds were constrained by linear constraint solver (LINCS) algorithm (88). MD trajectories were sampled every 2 ps and used for analysis. Moderate positional restraints (1000 kJ/mol•nm^2) had been placed on heavy atoms of both protein and DNA molecules during the equilibration steps to avoid drastic rearrangement of the macromolecules, while still allowing water and salt molecules to move freely. The positional restraints were removed during the production run.

6.3.4 MD data analysis

All analyses were performed on 100 ns of production MD trajectories using GROMACS built–in analysis tools with the aid of custom computer scripts. RMSD with respect to the initial structure in the production run, the radius of gyration, and root–mean–square fluctuation (RMSF) with respect to time–averaged positions were calculated to analyze relaxation of the complexes and convergence of the simulations. The protein–DNA interactions were analyzed by calculating the direct hydrogen bonds, salt bridges, hydrophobic contacts, and water–bridges. The H–bond cut–off distance for direct H–bond was set to 0.35 nm with an angle cut–off value of 30º. The output was a set of interaction time frames that satisfy the H–bond cut–off criteria, which was processed to obtain the percentage of time for a specific hydrogen bond to exist during 100 ns of production runs. Positively charged residues along the binding surface were analyzed for possible salt–bridges by calculating the minimum distances between charged side–chain nitrogens to DNA phosphate oxygens. The salt–bridge cut–off distance was set to 0.33 nm (89). The hydrophobic interaction was analyzed by tracking minimum distance between non–polar residues (including aliphatic Ala, Val, Leu, Ile, Pro and aromatic Phe, Tyr, Trp) and DNA base carbons.

The free energy landscape (FEL) sampled by the protein–DNA complex was extracted by mapping Gibbs free energies to the corresponding RMSD and radius of gyration. The landscape displays meta–stable conformational states of the system as low energy valleys and energy barriers as peaks. The global landscape minimum was used to extract the minimum–energy conformation of the complex.

The functional motion of the protein–DNA complex was analyzed using principal component analysis (PCA). In this method, a covariance matrix was constructed from the MD trajectory and then diagonalized using GROMACS inbuilt tools, resulting in a set of eigenvectors and eigenvalues. Each eigenvector describes a collective motion of particles and the associated eigenvalue reflects the amplitude of the motion. This analysis reduces molecular motion into a few collective degrees of freedom, which account for the most essential dynamics of the system (90).

6.3.5 Binding free energy analysis

The Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) and the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) are most frequently used methods to calculate binding free energies of macromolecules in computational works. Both methods are computationally efficient and MM/PBSA was shown to perform better in calculating absolute binding free energies (91). In MM/PBSA, binding free energy of a protein and DNA forming a complex can be calculated as follows.

$$\Delta G_{bind} = G_{complex} - (G_{Protein} + G_{DNA}), \qquad (6.1)$$

where $G_{Complex}$, $G_{Protein}$ and G_{DNA} are average free energies of the complex, the isolated protein, and isolated DNA respectively. The free energies are given by

$$G = (E_{bond} + E_{vdw} + E_{elec} + G_{pol} + G_{nonpol}) - TS_S, \quad (6.2)$$

where S_s is solute entropy. Therefore,

$$\Delta G_{bind} = \Delta E_{bond} + \Delta E_{vdw} + \Delta E_{elec} + \Delta G_{pol} + \Delta G_{nonpol} - T\Delta S, \quad (6.3)$$

where ΔE_{bond} (bond, angle, dihedral, and improper), ΔE_{vdw} (van der Waals) and ΔE_{elec} (electrostatic) contribute to the change of the gas phase molecular mechanics energy. ΔG_{pol} and ΔG_{nonpol} are polar and non–polar contributions to the change in solvation free energy respectively, T is the absolute temperature and ΔS_S is the change in solute entropy (92, 93). In the single trajectory approach, ΔE_{bond} is by definition zero (94). The binding free energy without the entropic term can thus be expressed as,

$$\Delta G_{bind} = \Delta E_{vdw} + \Delta E_{elec} + \Delta G_{pol} + \Delta G_{nonpol} \quad (6.4)$$

Different energy terms in Equation (4) were calculated for HOP2 and its mutant systems by using the GROMACS compatible program (g_mmpbsa) (95). The dielectric constant of water solvent was set to 80. The dielectric constants of biomacromolecules were not universally agreed in the field (96). In one work on protein–DNA systems, several dielectric constants (4 for non–polar, 9 for polar, and 10 for charged residues) were used in MD simulations (97). In another study, values of 1, 2, and 4 were used for protein–ligand systems (91). A recent study showed that 6–7 is appropriate for the interior of proteins (98), and another study reported experimental measurement ~8 for DNA (99). Based on these, dielectric constant of 6 was chosen for both macromolecules in this work. The most stable 15 ns (60–75 ns) of MD trajectories were used for energy calculations for all three complexes; the complex with the H–mutant was unstable and it was inappropriate to assess energies for the whole trajectory.

6.4 Results and Discussion

The compactness of protein–DNA complexes and the convergence of MD trajectories as well as the structural flexibilities were first assessed by calculating RMSD, RMSF, and radius of gyration. Protein–DNA interactions in terms of hydrogen bonds, salt bridges, and van der Waals contacts were calculated from 100–ns MD trajectories for wild–type HOP2 and its mutants in complex with DNA. The free energy landscapes were obtained to find the minimum energy configurations. The major molecular motions of protein–DNA complexes were investigated by the principal component analysis.

6.4.1 Structural stability of protein–DNA complexes

RMSD and RMSF together provide insights into stability and flexibility of structures and convergence of MD trajectories (23). For the three protein–DNA complexes, all–atom RMSD for protein molecules and whole complexes were shown in Figure 6.1A and 6.1B, respectively. RMSD values initially increased from 0 to 0.1–0.2 nm in the first nanosecond of the production run. The small sudden initial increase was expected because different pressure coupling constants and positional restraints were used in the equilibration and production runs (see Methods). Wild–type HOP2 and mutant proteins were stable throughout the simulation, which can be seen from the small protein RMSD of about 0.2 nm (Figure 6.1A). The complex RMSD values for the wild type HOP2 and W1–mutant (Y65A/K67A/Q68A mutations in wing 1) were around only 0.24 nm and 0.32 nm, respectively (Figure 6.1B). In contrast, the complex RMSD for the H–mutant (Q30A mutation in helix 2 and K44A/K49A in helix 3) was initially around 0.28 nm until 17 ns. Then it gradually increased to a much higher average of 1.2 nm. The radii of gyration for the three complexes were shown in Figure 6.1C. The complexes with wild–type and W1–mutant HOP2 remained compact with average radius values of 1.55 nm and 1.56 nm, respectively. In contrast, the complex with the H–mutant remained compact with the radius of 1.6 nm only until 17 ns. The radius then gradually peaked to 2.45 nm at 40 ns, suggesting a large separation between the protein and DNA. Taken together with RMSD, these results indicated that the H–mutant dissociated from DNA after 17 ns (Figure 6.1C). The radius value then dropped to an average of 1.8 nm and stabilized, which suggested that the protein approached to DNA again. This process for the H–mutant was illustrated in a movie (Movie S1).

The system stability, compactness, and convergence of the DNA complex of the W1–mutant were found comparable to those of the DNA complex of the wild type. The complex of H–mutant with DNA, however exhibited notable loss of stability. A closer inspection of configurations at 100 ns (Figure 6S1) confirmed that the wild type and W1–mutant were stable with H3 resided in DNA major groove and W1 in minor grooves during the simulation, whereas the H–mutant stayed relatively far from DNA with no significant binding.

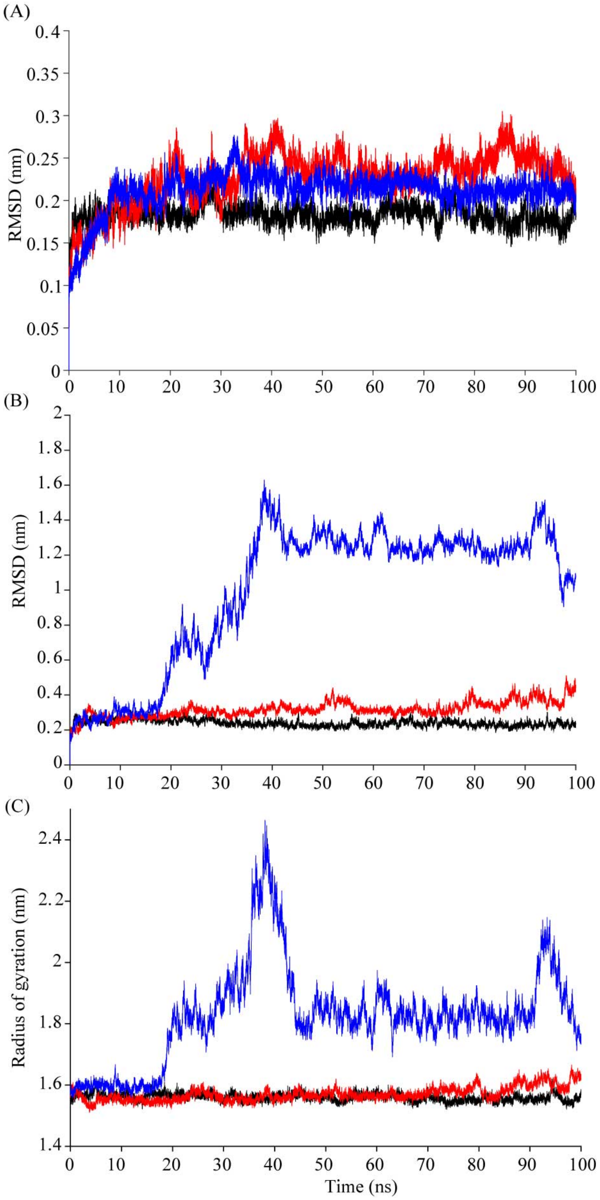


Figure 6.1. All–atom RMSD and radius of gyration analyses for three protein–DNA complexes. (A) RMSD of protein HOP2 and its mutants, (B) RMSD of the whole protein–DNA complexes, (C) The radii of gyration of the complexes. Wild type is shown in black, W1–mutant in red, and H–mutant in blue.

The residue–wise RMSF of protein backbone alpha carbons were calculated from 100–ns trajectories for both free proteins and proteins in the complexes (Figure 6.2). In the absence of DNA, the RMSF patterns were similar for all three cases since the overall structure was maintained in the mutants (see details Methods). However, the region consisting of helix H2 and the loop between H1 and H2 in mutants was more stable than that of the wild type (Figure 6.2A). To understand this observation, long–range hydrogen bonds involving this region were calculated (Figure 6.3 and Table 6S1). The mutants exhibited almost twice as many long–range hydrogen bonds as the wild type. The additional hydrogen bonds helped the mutants to be more rigid in this region. On the other hand, RMSF at E39 increased significantly for the mutants (Figure 6.2A), which was attributed to the loss of the hydrogen bond between E39 and Y19 side chains (Figure 6.3 and Table 6S1).

Upon forming complex with the DNA, RMSF of the wild type was significantly reduced for almost all residues, with the exception of a slight increase for E39 (Figure 6.2B, Figure 6S2). The most significant reductions were observed for wing 1, helices H2 and H3. The mutated regions in respective mutants showed considerable fluctuations when paired with DNA.

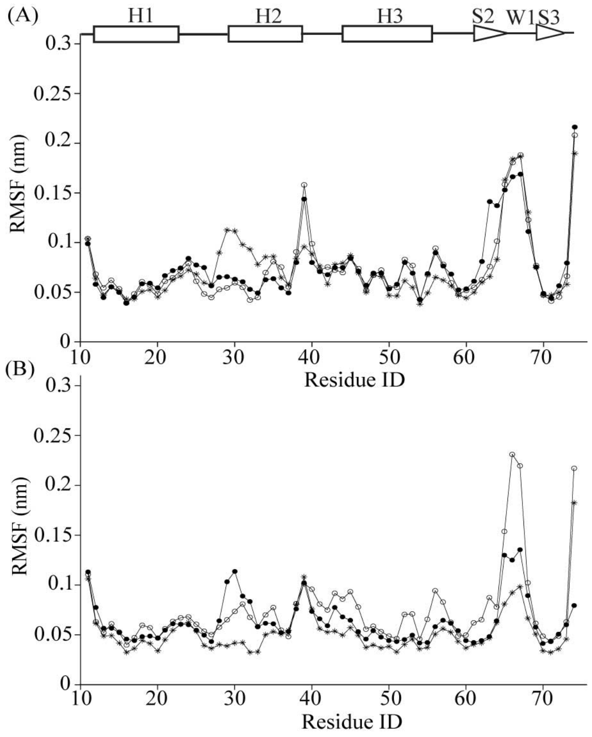


Figure 6.2: Protein alpha–carbon RMSF of wild type HOP2 (asterisk), W1–mutant (open circles), and H–mutant (filled circles) for (A) free protein and (B) protein in complex.

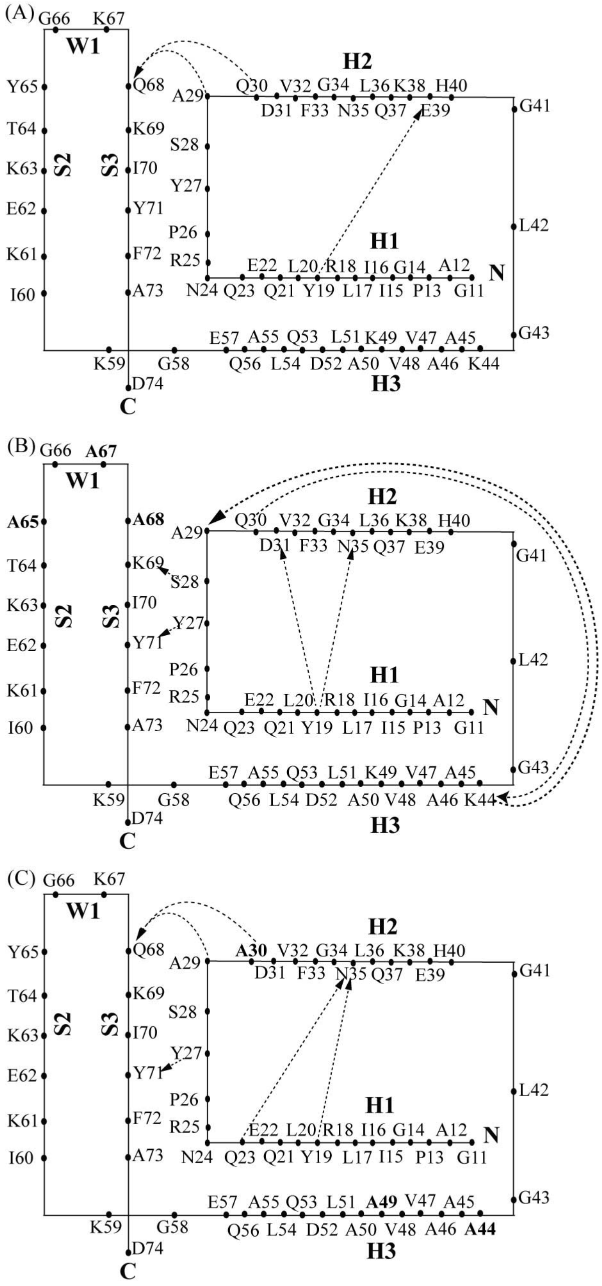


Figure 6.3: Diagrams of long–range hydrogen bonds involving helix 2 and the loop between helices 1 and 2 of (A) wild–type HOP2, (B) W1–mutant, and, (C) H–mutant. Intra–molecular hydrogen bonds here were calculated with the maximum bond distance set to 0.35 nm and the maximum angle at hydrogen atom to 60º (100).

6.4.2 Hydrogen bond interactions

Proteins interact with DNA via hydrogen bonds, salt bridges, and hydrophobic interaction (101). Polar and charged residues of proteins play a central role in coordinating DNA via direct hydrogen bonds (102). In this study, direct hydrogen bonds were analyzed using a GROMACS built–in tool g_hbond. The distance and angle cut–off values for this calculation were set at 0.35 nm and 30º, respectively, according to the default criterion in GROMACS (23). A list of time frames during which specific H–bonds existed was obtained. This output was further processed using a custom computer script to calculate the percent of time frames for a specific hydrogen bond to exist (Table 6.1). Similar approach was used to investigate the water–mediated hydrogen bonds (data not shown), which were found insignificant compared to direct hydrogen bonds.

Table 6.1. Direct hydrogen bonds between protein and DNA for the wild type HOP2, the W1–mutant, and the H–mutant. Low occurrences below 5% existence were omitted.

Protein Donor	**Donor atom**	**Nucleic Acceptor**	**Acceptor atom**	**% Existence**		
				Wild type	**W1–mutant**	**H–mutant**
Ala12	N	G10	O3'	38		
	N	A11	O1P	45.1		9.9
	N	G10	O1P		48.5	
	N	G10	O2P		26.8	
Arg25	NH2	G10	O3'			26.1
	NE	G10	O3'			21.8
Ser28	OG	T9	O3'			10.6
Gln30**	NE2	G44	O2P	6.4		NA
	NE2	C45	O1P	10.4	20.4	NA
	NE2	A46	O1P		17.5	NA
Lys38	NZ	A52	O3'			6.3
	NZ	T53	O3'			5.7
Lys44**	NZ	G44	O1P	30.9		NA
	NZ	C45	O1P	16.9	19.5	NA
Lys49**	NZ	A12	O2P	36.5		NA
	NZ	A47	N6	19.9		NA

	NZ	T9	O4		6.1	NA
	NZ	A48	N6		41.1	NA
	NZ	T49	O4		14.1	NA
Gln53	NE2	A11	O2P	27.9		
Gln56	NE2	A48	O1P	13		
Lys59	NZ	T9	O1P	27.7		
Tyr65*	OH	G10	O1P		NA	8.9
Lys67*	NZ	C18	O2	5.8	NA	
	NZ	C18	O3'	29.9	NA	
	NZ	G19	O1P	6.5	NA	
	NZ	G44	O3'	16.1	NA	
	NZ	C45	O1P	6.2	NA	
	NZ	C18	O4'		NA	15.7
	NZ	G19	O5'		NA	5
Gln68*	NE2	A46	O1P	64.6	NA	
	NE2	G44	O3'		NA	16.9
Lys69	NZ	A47	O1P	47.8	38.6	
	NZ	A47	O2P	45.9	62.3	
	N	A46	O1P	23.1		6.4
	N	A46	O2P	23.4	10.6	
	NZ	A46	O3'		31.4	

*Residues mutated to Ala in the W1–mutant.
**Residues mutated to Ala in the H–mutant.

In wild type HOP2, Ala12 (83%) in helix 1 and Lys44 (48%), Lys49 (56%), Gln53 (28%), and Lys59 (28%) in helix 3 were bound to DNA. Lys67, Gln68 and Lys69 in and around wing 1 were also bound with DNA over 60% of the 100 ns simulation time. Gln30 (17%) in helix 2 and Gln56 (13%) in H3 interacted with DNA but were considered insignificant. The hydrogen bond interactions from residues in helices 1 and 3 helped helix 3 to stay securely in major groove. Wing 1 reached out to interact with phosphate oxygen atoms in minor groove. This result corroborated well with previously proposed docking configuration and the canonical mode of DNA binding for winged–helix domain HOP2 (12).

Hydrogen bonds between protein and DNA bases largely determine the sequence specificity; hydrogen bonds with phosphate backbones are generally considered non–specific, contributing mostly to the stabilization of the protein–DNA complex (103). For wild–type HOP2, DNA backbone phosphate groups were involved in hydrogen bond interactions except for one case involving the adenine base of nucleic acid A47. The side chain of Lys49 in H3, which reached deep into the major groove and facilitated interaction with base nitrogen N6 of A47 for about 20% of the time (Figure 6.4). Moreover, no water–mediated hydrogen bond with the nucleic base was found either. The single amino acid to nucleic base interaction may be insufficient to determine the sequence specificity for DNA recognition (103).

In the W1–mutant, three residues (Tyr65, Lys67, and Gln68) in wing 1 were mutated to alanines, and the hydrogen bonds involving Lys67 and Gln68 side chains were no longer available. The Lys69 sidechain nitrogen NZ formed more than one hydrogen bonds with DNA at times, giving total of 130.5% existence. Lys69 likely made major contribution to the stability of wing 1 region. Indeed, in another simulation with Lys69 also mutated (Y65A/K67A/Q68A/K69A), the protein completely dissociated from DNA after about 25 ns (see Figure 6S3 for RMSD and radius of gyration analyses, and Figure 6S4 for the structure after 100 ns). In the W1–mutant, Lys49, Lys44, Gln30 in helices 3 and 2 and Ala12 in helix 1 were hydrogen–bonded to DNA. Lys49 also interacted with the base nitrogen of A48. Two residues in helix 3 (Ala46 and His40) showed insignificant direct hydrogen bonds with DNA (data not listed in Table 6.1). The bonding pattern slightly changed and several direct hydrogen bonds involving the mutation sites and helix 3 (Gln53, Gln56 and Lys59) were missing. This also supported the mutagenesis study that showed a weakened but incompletely abolished binding for the W1–mutant (13).

The H–mutant, where three residues (Gln30, Lys44, and Lys49) in helices 2 and 3 were mutated to alanines, showed no significant intermolecular hydrogen bonds that involved helix 3 and wing 1. All hydrogen bonds found for this mutant–DNA complex existed for less than 20% of the simulation time except for Arg25 (48%). Combining this result with RMSD and radius of gyration calculations, it can be concluded that there were interactions between the H–mutant and DNA for the first 17–18 ns of the simulation time. These interactions were so weak that the protein separated from the complex after around 19 ns. It flipped around 40 ns, followed by approaching the DNA with Arg25 in contact with the DNA during 45–90 ns of the simulation. However, this single interaction was insufficient to stabilize the complex (Figure 6.1C). Thus, the mutation of residues in helices 2 and 3 had remarkable impact in binding. Lack of direct hydrogen bonds also supported the impairment of H–mutant in DNA binding observed in mutagenesis studies (13).

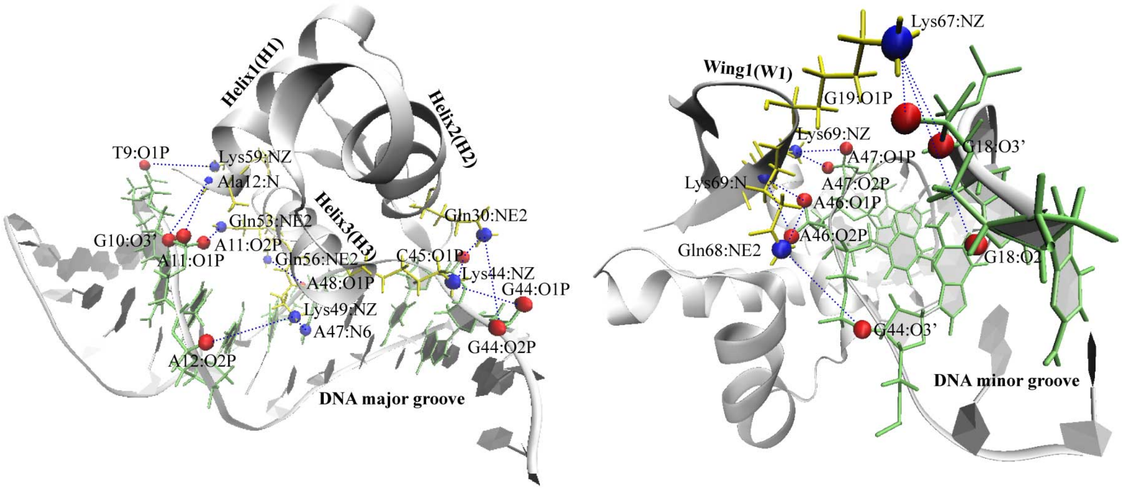


Figure 6.4: The hydrogen bond network for wild type protein–DNA complex. The interacting protein residues (yellow) and the nucleic acids (green) are shown in sticks, whereas donor and acceptor atoms are represented by spheres.

6.4.3 Salt bridge interactions

Salt bridges are commonly observed non–covalent interactions in protein–DNA complexes that contribute stability to entropically unfavourable conformations (104, 105). The salt bridges along with hydrogen bonds and van der Waals interactions make the most contribution to the overall stability of the complexes (106). In this study, the positively charged protein residues (Lys38, Lys44, Lys49, Lys63, Lys67, and Lys69) along the binding surface were analyzed for possible salt–bridge interactions with DNA. Arginines for wild type and W1–mutant were not included in this calculation because none of them was at the binding interface. GROMACS tool g_mindist was used to calculate the minimum distances between charged side–chain nitrogen to DNA phosphate oxygen atoms (Table 6.2). Salt bridge interactions involving Lys–phosphate in complexes were found constantly broken and remade, which is an indication of its dynamic nature (104).

Table 6.2. The average minimum distances from positively charged nitrogens of lysine residues to negatively charged oxygens of DNA calculated for 100 ns of MD trajectories.

Residue	**Minimum distance (nm)**		
	Wild type	**W1–mutant**	**H–mutant**
Lys38	_	_	0.94 ± 0.56
Lys44**	0.32 ± 0.08	0.46 ± 0.14	NA
Lys49**	0.60 ± 0.08	0.86 ± 0.10	NA
Lys63	0.87 ± 0.04	0.88 ± 0.05	_
Lys67*	0.36 ± 0.11	NA	0.74 ± 0.33
Lys69	0.29 ± 0.02	0.29 ± 0.03	0.91 ± 0.43

*Residues mutated to Ala in the W1–mutant and no salt bridge was expected.
**Residues mutated to Ala in the H–mutant and no salt bridge was expected.

Among protein residues of interest in the binding interface, Lys44 and Lys69 in the wild type protein had 0.32 and 0.29 nm distances, respectively, within a 0.33 nm range (89) for salt bridges. For the W1 mutant, only Lys69 was found to be within the range. For the H–mutant however, distances were well above the threshold. Arg25 of the H–mutant, which was found to be close to DNA (Figure 6S1(D)), was also analyzed for possible salt bridge interaction (data not included in the table). The average minimum distance over 100 ns period was found to be 0.99 nm with very high standard deviation of 0.76 nm. The distance after 50 ns was found to average at 0.42 nm with high deviation of 0.28 nm. The absence of salt bridge interaction is one of many reasons why the H–mutant DNA complex was very unstable.

6.4.4 Hydrophobic interactions

Hydrophobic interaction is much stronger than what would be expected from classical theories and it plays an important role in the stability of protein–DNA complexes (107, 108). Although the hydrophobic interaction is long range in nature (109), it is much stronger at short distances (108).

In this study, the hydrophobic interactions were analyzed by calculating minimum distances between DNA base carbons and non–polar aliphatic (Ala, Val, Leu, Ile and Pro) and aromatic (Phe, Tyr, Trp) residues along binding interface. Residues in helices 2 (Phe33) and 3 (Ala45, Ala46, Val47, Val48, Leu51, and Leu54) were analysed for possible hydrophobic interactions (Table 6.3).

Table 6.3. The average minimum distances of non–polar and aromatic residues to DNA base carbons.

	Distance (nm)		
Residue	**WT**	**W1–Mutant**	**H–Mutant**
Phe33	1.03 ± 0.06	1.02 ± 0.05	1.19 ± 0.39
Ala45	0.56 ± 0.04	0.38 ± 0.11	1.14 ± 0.64
Ala46	0.68 ± 0.05	0.47 ± 0.10	1.66 ± 0.66
Val47	0.80 ± 0.03	0.77 ± 0.08	1.58 ± 0.47
Val48	0.30 ± 0.02	0.50 ± 0.13	1.38 ± 0.63
Leu51	0.77 ± 0.04	0.85 ± 0.09	1.65 ± 0.47
Leu54	0.87 ± 0.07	1.12 ± 0.11	2.08 ± 0.57

For the wild–type HOP2, Val48 remained very close to DNA bases throughout the simulation with an average distance of 0.30 nm. Ala45 and Ala46 were found at average distances of 0.56 and 0.68 nm respectively. For the W1 mutant, Ala45, Ala46, and Val48 also have the shortest distances to DNA bases of 0.38, 0.47, and 0.50 nm, respectively. These hydrophobic interactions might have assisted the neighboring Lys49 in the wild type and the W1–mutant to form direct hydrogen bonds with DNA base nitrogen atoms (N6 in Table 6.1), which are deeper in the DNA groove than the backbone phosphate groups. Ala45 and Ala46 in the W1–mutant were closer to DNA than their counterparts in the wild type. This additional hydrophobic interaction might be responsible for the stability of W1–mutant DNA complex, despite of the loss of several hydrogen bonds.

The distances in the case of the H–mutant were all beyond 1 nm, indicating weak or no hydrophobic contact with DNA. The absence of hydrophobic interactions and salt bridges along with insignificant hydrogen bond interactions support the insignificant DNA binding affinity of the H–mutant (12).

6.4.5 Free energy landscape and protein DNA complex configuration

The protein–DNA structures must possess a well–defined free energy minimum to be considered reliable complex configurations. Such configurations can be obtained by constructing free energy landscape from the simulation trajectories. In this study, FEL for complexes of HOP2 and its mutants were obtained by mapping Gibbs free energies to the corresponding RMSDs and radii of gyration (Figure 5).

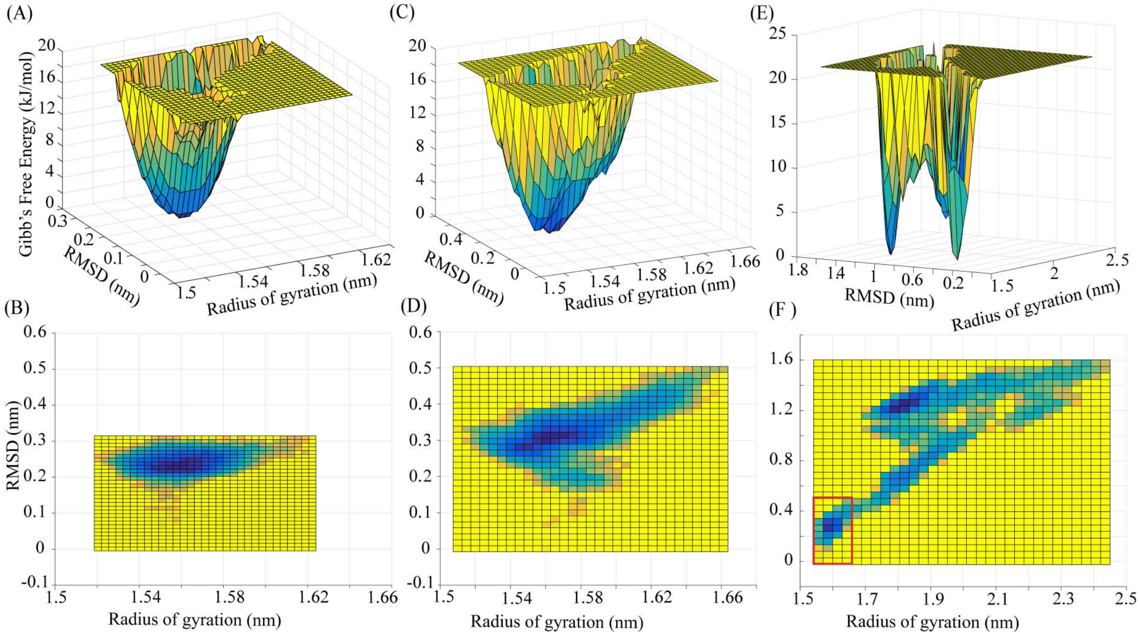


Figure 6.5. The free energy landscape for protein–DNA complexes calculated for 100 ns of MD trajectories for (A) wild type (C) W1–mutant and (E) H–mutant. Free energy contour maps as a function of the RMSD and the radius of gyration of protein–DNA complex for (B) wild type (D) W1–mutant and (F) H–mutant. For comparison, the red rectangle in (F) represents the same range

plotted in (D) for the W1–mutant. Each energy landscape consisted of 1024 (32×32) data points and Matlab was used to generate the plots.

The DNA complexes of the wild type and the W1–mutant were found to converge to a minimum in FEL (Figure 6.5A–D). FEL of the wild–type HOP2–DNA complex exhibited a minimum of zero energy with 1.56 nm radius of gyration and 0.22 nm RMSD (Figure 6.5A, B). The W1–mutant complex also showed a minimum of zero energy with 1.57 nm radius and 0.31 nm RMSD (Figure 6.5C, D). The optimal configuration with minimum free energy was retrieved from the trajectory for the wild–type HOP2 in complex with DNA (Figure 6S5). Helix 3 of the protein lodged in a major groove whereas wing 1 reached to a minor groove of DNA, promoting several interactions that stabilized the complex. The W1–mutant also showed a similar stable configuration (Figure 6S6A). In contrast, the H–mutant complex exhibited a large conformational variation as evident from very large RMSD and radius values (Figure 6.5E, F). The energy landscape showed a local minimum of 0.895 kJ/mol with radius of 1.572 and RMSD of 0.236 nm. This corresponded to the initial stage before the protein dissociated from the DNA. The FEL also exhibited a energy minimum of zero with radius of 1.776 and RMSD of 1.177 nm. The corresponding configuration for the H–mutant DNA complex at energy minimum was shown in Figure S6B. Combined with the calculated interactions (Tables 6.1–6.3), it became obvious that the H–mutant showed very little to no affinity in DNA binding.

6.4.6.Principal component analysis

The functional motion of biological macromolecules has been a key to understand their interactions and dynamics (110). The regions of macromolecules that account for the most significant dynamics can be assessed via Principal Component Analysis by transforming fast local atomic motions from MD trajectories into dominant functional motions (90). In this study, GROMACS inbuilt tools were used to capture large–scale motions for HOP2 and its mutants in complex with DNA. For each protein–DNA complex, a covariance matrix was constructed from its MD trajectory, followed by solution of eigenvectors and eigenvalues for the matrix. The conformational space (Figure 6.6) for the complex was obtained by projecting the trajectory onto two of its eigenvectors (principal components) with the largest eigenvalues (Table 6S2).

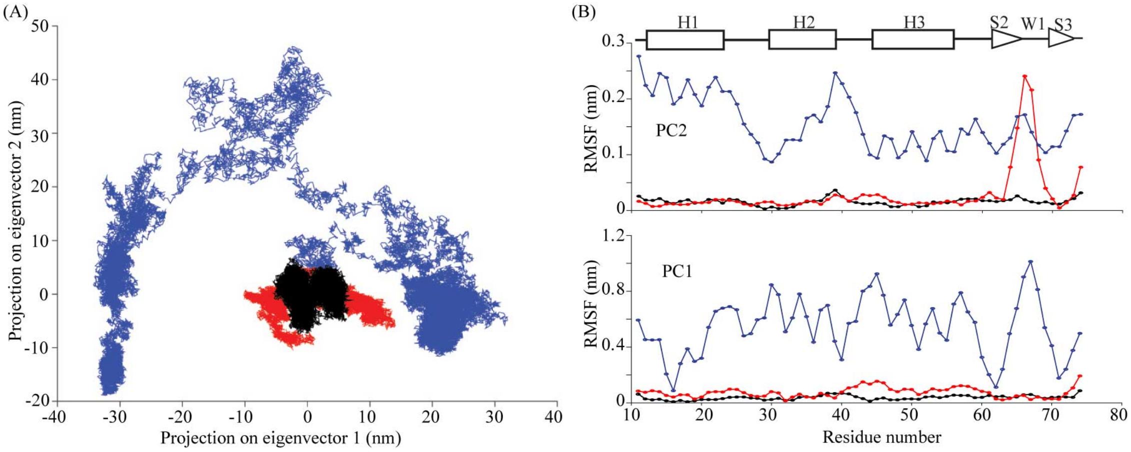


Figure 6.6. (A) Conformational space obtained by projection of the trajectories onto the first two eigenvectors for the the complexes formed between DNA and wild type HOP2 (black), W1–mutant (red), and H–mutant (blue). (B) The RMS fluctuation of each protein α carbons along its principal components PC1 and PC2.

The wild type protein–DNA complex was found to cluster in a small region of the conformational space (Figure 6.6A, black). This suggested that there were very small structural variations, which correlated to their restricted relative motions between the two molecules. The W1–mutant spanned a slightly larger area (Figure 6.6A, red), which indicated that its complex with DNA experienced slightly more conformational changes. On the contrary, the H–mutant explored several widespread regions (Figure 6.6A, blue) separated by energy barriers, indicating high degree of conformational variation.

The mobility of the protein in the complex along the first two eigenvectors was accessed from the backbone α carbon RMS fluctuations (Figure 6.6B). The motion of the wild type was more restricted along both PC1 and PC2 compared to both mutants according its low RMSF. The W1–mutant was slightly less restricted than wild type. The H–mutant, showed the least restricted motion with the highest fluctuations along both components. This is consistent with the observation that the H–mutant DNA complex conformational cluster was not found localized in conformational space.

Two large amplitude dynamics were visualized for the wild type (Movie S2 and S3) and W1–mutant (Movie S4 and S5). For the wild type, motion associated with the largest amplitude (7.1 nm, Table 6S2) exhibited as rolling of the DNA molecule around its double–helix axis (Movie S2). Relative to the DNA molecule, the protein smoothly slid along the DNA major and minor grooves. This explained the change of intermolecular hydrogen bond partners during the simulation (Table 6.1). This sliding motion could be functionally important to facilitate the precise positioning of the single–stranded DNA sequence with the homology double–stranded DNA in the catching and homology matching stages of the strand–invasion model proposed by Kang et al. (73). The motion associated with the second largest amplitude (4.8 nm, Table 6S2) exhibited a side–bending mode of the DNA molecule with little relative motion between the protein and DNA (Movie S3).

For the W1–mutant the largest motion (amplitude 20.0 nm, Table 6S2) demonstrated as a twisting of the DNA molecule (Movie 6S4). The protein molecule has large motion relative to the DNA, with the C–terminal end of the helix 3 being completely disengaged from the major groove. Wing 1 and the N–terminal end of the helix 3 were still in contact with the DNA. The second largest motion (5.0 nm, Table 6S2) is a DNA rolling together with the protein sliding along the DNA grooves (Movie S5), showing smaller amplitude than the wild type. Both weakened binding and the reduced groove sliding motion may adversely affect the functional activities of the W1–mutant.

6.4.7 Interaction energy

The Molecular Mechanics/Poisson–Boltzmann Surface Area (MMPBSA) method is one approach to estimate the binding energy of biological macromolecules. In this method, molecular mechanical energies and solvation free energies for an ensemble of configurations are obtained from a MD trajectory. A GROMACS compatible program g_mmpbsa (95), which is based on this method, was used to evaluate components of free energy for protein–DNA complexes during the most stable 15 ns (60–75 ns, see Figure 6.1) of the simulation trajectories (Table 6.4).

Table 6.4. Comparison of binding energy components among wild type, W1–mutant, and H–mutant HOP2 proteins in complex with DNA.

Energy terms (kJ/mol)	**Wild–type**	**W1–mutant**	**H–mutant**
Electrostatic	–1185±43	–892±39	–482±95
van der Walls	–259±19	–180±19	–100±33
Polar solvation	1191±103	811±78	499±166
Non–polar solvation	–35±2	–19±3	–9±5
Net Binding energy	–287±83	–279±69	–91±104

For the wild–type protein in complex with DNA, the binding energy was found very large (in terms of magnitude). Except for the polar solvation energy, all components including the intermolecular electrostatic, van der Waals, and nonpolar solvation free energies contributed favorably to the complex formation. For the W1–mutant, the reduced number of direct hydrogen bond and salt bridge interactions resulted in smaller electrostatic energy and polar solvation energy. The mutations on wing 1 caused the protein to move away from the DNA to a certain extent. This eliminated some interactions with DNA, resulting in a smaller interface area between the two molecules. These effects were reflected in smaller van der Waals and nonpolar solvation energies. The overall contribution from these components resulted in a slightly smaller binding energy than the wild type, agreeing with the weakened binding activity. In contrast, the H–mutant in complex with DNA exhibited dramatically smaller energy components. The hydrogen bond and salt bridge interactions between the H–mutant and DNA were sparse, leading to significantly smaller electrostatic and polar solvation energies. The large structural fluctuations in the H–mutant DNA complex caused the protein to migrate away from the complex, resulting in lower van der Waals and nonpolar solvation energies. The small binding energy indicated weak to no binding affinity of H–mutant with DNA. This remarkably reduced binding activity of the H–mutant signalled the crucial role of helix H3 in coordinating DNA.

6.6 Conclusions

The elucidation of HOP2 atomic structure has greatly contributed to the understanding of the molecular mechanism behind its action during recombination. The studies on the N–terminal of HOP2 have shown that this domain binds strongly to double–stranded DNA using the canonical DNA binding mode for winged–helix proteins (12). Residues from helix 3 and wing 1 were found actively involved in coordinating DNA, which were further confirmed by mutation studies (12). It was interesting that site–directed mutations of highly conserved residues in these motifs showed different impacts to the DNA binding affinity. Here, atomic level MD simulations of DNA complexes of HOP2 and its mutants were performed to evaluate the contributions of HOP2 motifs in DNA binding.

The wild–type HOP2 in complex with DNA was found stable during MD simulation. Residues from wing 1 and helices 2 and 3 were found to interact with DNA via direct hydrogen bonds, salt bridges, and hydrophobic interactions. Interactions involving recognition helix 3 and another helix 2 with DNA in major groove and wing 1 in DNA minor groove contributed most to the formation of the complex.

Despite having the canonical DNA binding mode and high structural similarity to sequence–specific DNA binding proteins such as BlaI and a number of other transcription factors (75), the N–terminal domain of HOP2 was found to dominantly interact with DNA phosphate backbones rather than DNA bases (103). This has also been observed for full–length HOP2 in complex with MND1 as a heterodimer (73). Non–specificity might be required for HOP2 to bind to a wide range of DNA sequences in order to perform its DNA repair functions. Nevertheless, the DNA sequence used in this study was arbitrary and the data presented here cannot fully exclude the possibility of sequence specificity.

For the W1–mutant DNA complex, most interactions involving helices 2 and 3 were still preserved. However, the number of interactions involving wing 1 were reduced due to mutations. Nonetheless, the DNA complex of W1–mutant was also stable due to slightly stronger hydrophobic interactions. The lowest energy structure of the protein–DNA complex was similar to and the binding energy was only slightly smaller than the wild type.

Mutations in helices 2 and 3 however had dramatic effects on the stability of the complex and on its DNA binding ability. The interactions involving residues in helices 2 and 3 completely disappeared. Helix 3 was no longer able to stay in the major groove. These results suggested that H–mutant failed to bind to DNA in the canonical binding mode. This mutant interacted with the DNA via a completely different interface with a three–fold smaller binding energy.

For the wild type and the W1–mutant, the protein molecule was found to slide along the major and minor grooves while the DNA molecule was undergoing a rolling motion along its double–helix axis. This sliding motion could be functionally important to facilitate the precise positioning of the single–stranded DNA sequence, which is carried by the C–terminal segment of the full–length HOP2 according to the functional model proposed by Kang and coworkers (73), with the homology double–stranded DNA. The sliding motion was relatively smaller for the W1–mutant than the wild type. The W1–mutant also underwent a large–amplitude DNA twisting motion, together with a large relative protein motion with the C–terminal end of the helix 3 being completely disengaged from the major groove. Both weakened binding and the reduced groove sliding motion may adversely affect the functional activities of the W1–mutant.

Simulation on another mutant with an additional mutation in wing 1 (Y65A/K67A/Q68A/K69A) caused the protein to completely dissociate from DNA. This is similar to the separation from DNA in the case of the H–mutant, which also included mutation of two lysines to alaines. From these observations, it became obvious that lysine residues along helix 3 and wing 1 provided the most critical protein–DNA interactions. These results supported the hypothesis that the charged residues distributed over the surface of HOP2 act as anchor points for DNA binding (12).

The crystal structure of the HOP2–MND1 heterodimer has provided significant insight into the mechanism on how these accessory proteins work during homologous recombination. The juxtaposed winged helix domains were found in a specific orientation, which was considered responsible for double–stranded DNA distortion (73). Similar to cocking a medieval crossbow with the archer's foot being placed in the stirrup while the bowstring being pulled, the two wing 1 motifs of HOP2–MND1 heterodimer or HOP2 homodimer push against the minor groove while the two helix 3 motifs pull the major groove on the sides. Therefore, the result that both helix 3 and wing 1 have strong interactions with DNA is important to the strand opening function of HOP2 and MND1, which in turn facilitates the strand invasion during homologous recombination.

Acknowledgements

We appreciate financial support from Oklahoma Center for the Advancement of Science and Technology (HR12–050). The computing for this project was performed at the OSU High Performance Computing Center at Oklahoma State University supported in part through the National Science Foundation grant OCI–1126330. The authors appreciate valuable discussion with Dr. Roberto Pezza at Oklahoma Medical Research Foundation (OMRF).

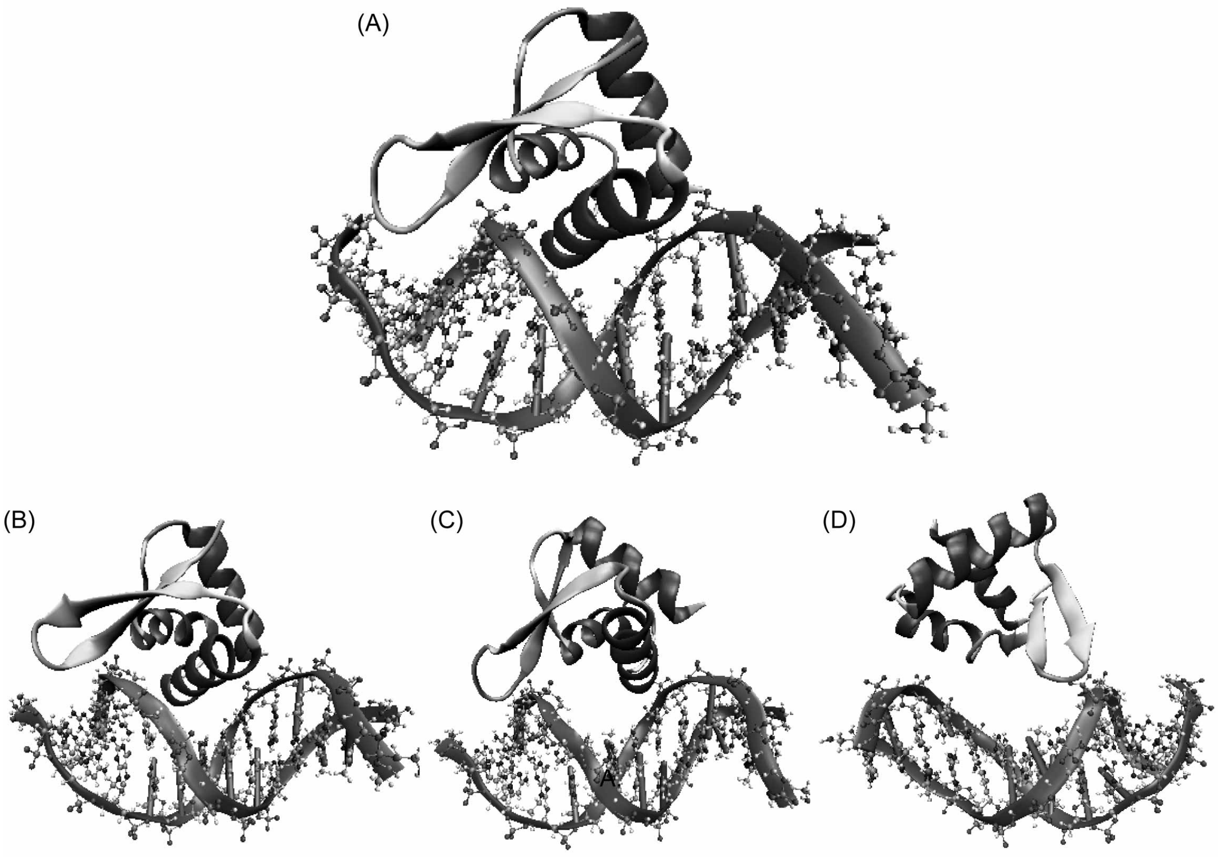


Figure 6S1: Starting configuration of protein-DNA complex (A), and the complex configurations after 100 ns MD simulations for (B) wild type (C) W1-mutant and (D) H-mutant.

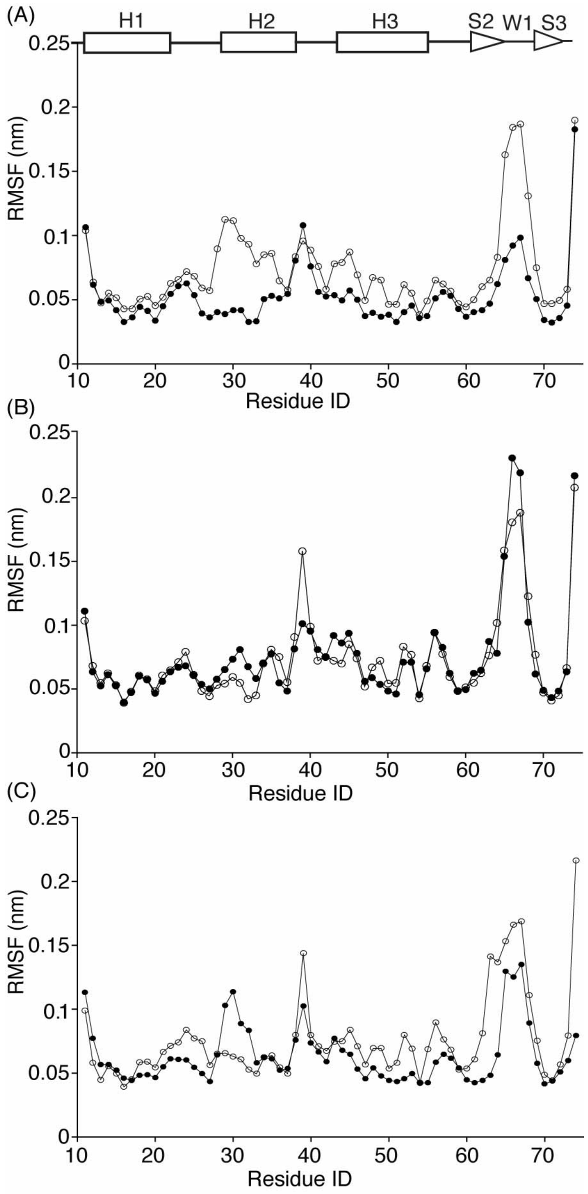


Figure 6S2: Protein alpha-carbon RMSF of protein alone (open circles), and protein in complex (filled circles) for (A) wild type, (B) W1-mutant, and (C) H-mutant.

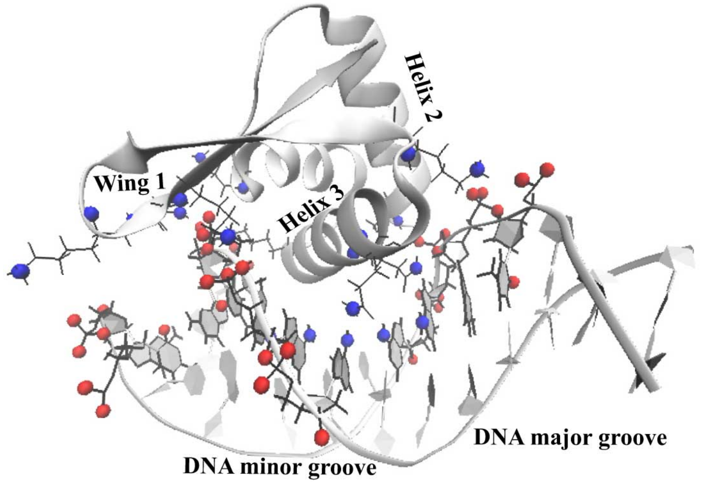


Figure 6S3: The minimum energy configuration of the wild-type protein-DNA complex. Nitrogen and phosphate oxygen atoms participating in intermolecular hydrogen bonds and salt bridges as listed in Tables 1 and 2 were shown in ball represtntations.

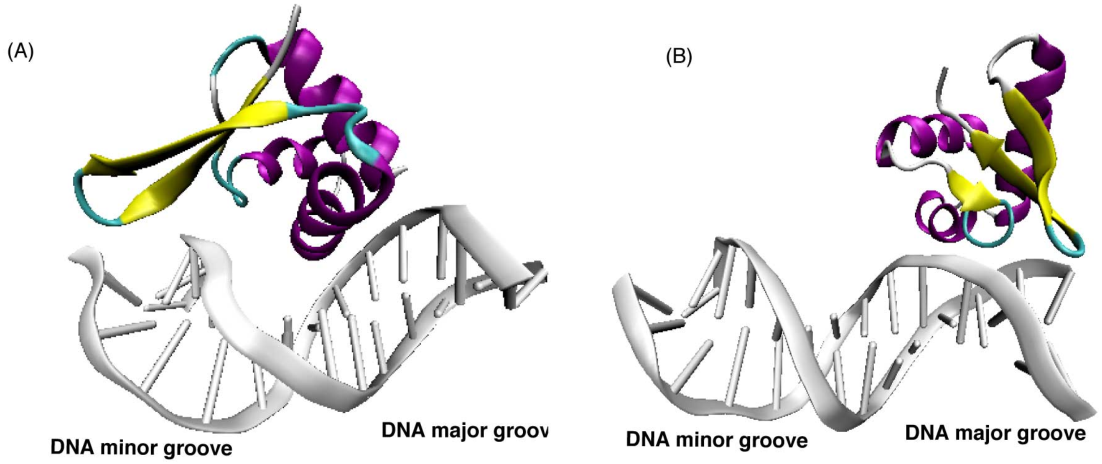


Figure 6S4: The minimum energy configuration of mutant in complex with DNA extracted using FEL (A) W1-mutant at 53820 ps with RMSD 0.31 nm and radius of gyration values 0.22 nm, and Gibb's free energy of 0 kJ/mol, (B) H-mutant at 66916 ps with RMSD and radius of gyration values 1.177 and 1.776 nm respectively and Gibb's free energy of 0 kJ/mol.

Table 6S1: The percent existence of the long-range hydrogen bond of (A) wild type protein, (B) W1-mutant and, (C) H-mutant in the absence of DNA. This specific hydrogen bond was calculated using GROMACS built in tool g_hbond with 0.35 nm maximum bond distance between the donor and the acceptor and 60^{o} maximum angle at the hydrogen atom.

(A) Wild type HOP2

Donor		Acceptor		% Exist.
PHE72	N	LYS61	O	47.251
TYR71	OH	GLU62	OE2	99.732
LYS61	NZ	ALA73	O	16.336
	NZ	ASP74	OD1	16.758
	NZ	ASP74	OD2	18.008
	N	(NH2)75	N	62.099
ILE60	N	NH275	N	34.331
ASP52	N	TYR71	OH	99.26
LYS44	NZ	ALA29	O	32.043
GLN30	N	GLN68	NE2	58.375
ALA29	N	GLN68	NE2	44.861
SER28	OG	LYS69	O	12.066
GLN21	NE2	(NH2)75	N	90.126
TYR19	OH	GLU39	OE2	98.234

(B) W1-mutant

Donor		Acceptor		% Exist.
ASP74	N	ASN24	O	32.833
PHE72	N	LYS61	O	47.009
LYS69	NZ	ASP52	OD2	64.655
LYS69	NZ	GLU62	OE1	64.965
LYS61	NZ	ASP74	N	19.262
	NZ	ASP74	OD1	18.132
	N	(NH2)75	N	48.793
ILE60	N	(NH27)5	N	29.987
LYS59	NZ	(NH2)75	N	97.176
ASP52	N	TYR71	OH	99.7
LYS44	NZ	ALA29	O	39.317
GLN30	N	LYS44	NZ	60.975
SER28	OG	LYS69	O	17.748
TYR27	N	TYR71	N	95.858
GLN21	NE2	(NH2)75	N	94.706
TYR19	OH	ASP31	OD2	11.534
TYR19	OH	ASN35	ND2	97.21

(C) H-mutant

Donor		Acceptor		% Exist.
NH275	N	GLY58	O	20.582
ASP74	N	LYS59	O	35.483
ALA73	N	ARG25	O	83.904
TYR71	OH	GLU62	OE2	99.756
LYS69	NZ	ASP52	OD2	72.117
	NZ	GLU62	OE1	64.473
GLN68	NE2	ALA29	O	11.652
THR64	OG1	LYS69	NZ	30.645
LYS61	NZ	ASP74	OD1	19.582
	N	ALA73	O	83.538
	N	(NH2)75	N	54.079
ILE60	N	(NH2)75	N	27.725
LYS59	NZ	ASP74	OD2	97.078
ASP52	N	TYR71	OH	99.624
ALA30	N	GLN68	NE2	57.623
ALA29	N	GLN68	NE2	71.113
TYR27	N	TYR71	N	96.568
ASN24	ND2	ASP74	OD2	97.368
GLN23	NE2	ASN35	ND2	91.71
GLN21	NE2	(NH2)75	N	98.66
TYR19	OH	ASN35	ND2	94.902

Table 6S2. Eigenvalues from PCA analysis of the HOP2 protein-DNA complexes. The eigenvalue reflects the amplitude of the motion. Characterization of the DNA dynamics for several large amplitude motions are also listed for the wild-type and the W1-mutant.

Wild Type		**W1-mutant**		**H-mutant**
Eigenvalue (nm)	DNA motion	Eigenvalue (nm)	DNA motion	Eigenvalue (nm)
7.1	Rolling	20.0	Twisting	570.9
4.8	Bending	5.0	Rolling	148.5
2.5	Bending	3.6	Twisting	91.5
2.0		3.0		45.8
1.5		2.9		29.9
1.2		2.3		17.5
0.9		1.6		8.9
0.9		1.4		7.4
0.8		1.2		4.6
0.7		1.1		3.6

CHAPTER VII

IMPLEMENTATION OF MD SIMULATIONS TO DESIGN MODELS FOR MEMBRANE PROTEINS AND VERIFICATION BY SOLID-STATE NMR

All MD simulations were performed on up to 600 processors on linux cluster supercomputer using software GROMACS 4.5.5 and GROMOS96 54A7 force field combined with lipid interaction parameters and simple point charge-extended (SPCE) water model (111, 112). The non-bonded van der Waals interactions were estimated using Lennard-Jones potential with cutoff value of 1.2 nm and the bonds were constrained by linear constraint solver (LINCS) algorithm (113). Electrostatic forces and energies were calculated using Particle-Mesh Ewald (PME) summation algorithm with cutoff value of 1.2 nm (114).

Two equilibration phases, constant volume (NVT) and constant pressure (NPT) ensembles, were subsequently carried out, each with 1 fs time steps. In the first phase, the system was coupled to a strong temperature bath using V-rescale coupling (115) with temperature coupling constant of $\tau_T = 0.1$ ps to maintain system temperature at 300 K. In the second phase, Parrinello-Rahman pressure coupling (116) with coupling constant $\tau_P = 5.0$ ps to maintain the pressure semi-isotropically at 1 bar and a weak Nose-Hoover temperature coupling with a coupling constant $\tau_T = 0.5$ ps (117-119) was used to ensure a true NPT ensemble. A total of 15 ns equilibration was followed by a 50 ns production run of molecular dynamics (MD) in 2 fs step size, during which temperature and pressure were maintained using weak coupling methods (Nose-Hoover with $\tau_T = 0.5$ ps and Parrinello-Rahman with $\tau_P = 2$ ps).

For a given structure model, backbone and β-carbon chemical shifts were predicted by shiftX (120). Based on these chemical shifts, ^{13}C-^{13}C 2D spectra were simulated by program peaks2ucsf in the Sparky package (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) with assistance of a custom computer script.

7.1 Membrane attachment and structure models of lipid storage droplet protein 1

Neutral lipid triglycerides, a main reserve for fat and energy, are stored in organelles called lipid droplets, which contain energy density 10 times that of hydrated proteins and carbohydrates (121). Utilization of the stored triglycerides requires enzymatic breakdown (lipolysis) by lipases, while the surface layer of the droplet controls the accessibility of lipases to the stored triglycerides. Among the proteins surrounding the lipid droplet surface, proteins in the PAT family (named after three earliest members) have raised great interest in recent year. These proteins actively regulate the storage and release of triglycerides. One of such proteins found in insects is lipid storage droplet protein 1 (Lsd1), also known as PLIN1. It interacts with lipid droplet to control access of lipase to triglycerides thus regulates the lipids homeostasis. In fact PLIN1 serves as a lypolytic switch, which upon protein kinase A (PKA) mediated phosphorylation, promotes the activation of triglyceride lipolysis (122).

The association of these proteins on the surface of the lipid droplets is critical to their ability to properly regulate both storage and release of the triglycerides in the droplets. Despite the need to understand the interaction between these proteins and the lipid droplets, progress has been hampered by the scarcity of 3D structural information of these proteins. Furthermore, the structure of PLIN1 and its relation to functions remain elusive due difficulties in structural calculations as it is insoluble and difficult to crystallize. Here we report the first structural basis for the lipid droplet attachment of *Drosophila melanogaster* PLIN1 using solid state NMR in combination with molecular dynamics simulation.

7.1.1 Model building

The structure models for any lipid droplet proteins of PAT protein family bound to membrane were unavailable so we were interested to see if any transmembrane domains existed in PLIN1. We used HMMTOP server to predict transmembrane domain and topology but PLIN1 showed no such domain (123). Also, the conserved PAT domain of the PLIN1 is highly soluble and unlikely to participate in lipid interaction. Bioinformatics analysis indicated that four predicted helices might participate in membrane targeting: hydrophobic helices H6 (249-261) and H8 (290-300) as well as amphipathic helices with high hydrophobic moment H7 (265-275) and H9 (301-318). The spin diffusion NMR experiments also performed on the membrane embedded Lsd1 corroborated this hypothesis. We therefore built structural models for the membrane-binding domain, contained in the segment stretching from residue 249 to 318. This segment also included a highly conserved hydrophilic motif ^{282}EPENQARP^{289} that could act as a modulator of lipolysis.

Two models of the most probable membrane-binding motif (residues 249 to 318) containing predicted helices 6 to 9 (124) were manually constructed based on hydrophobicity distribution: *trans*-model in Fig. 7.1 (A) and *cis*-model in Fig. 7.2 (A). The amphipathic helices H7 and H9 most likely lie in the membrane-cytosol interface. The loop between H7 and H8 is so placed as to expose the hydrophilic conserved ^{282}EPENQARP^{289}, which could be functionally important. This requires the hydrophobic helices H6 and H8 to be placed parallel to the membrane surface. Had these two helices been aligned perpendicular to the membrane surface, this hydrophilic stretch would have been pulled into the membrane. First, Cα trace was designed with assistance of Visual Molecular Dynamics (VMD) (125) and then all-atom molecule structure was generated by Structural Alphabet based protein Backbone Builder from alpha Carbon trace (SABBAC) 1.3 (26). The two starting models (shown in Fig. 5A and Fig. S4) were energy minimized in vacuum using GROMACS 4.5.5 (126) and then embedded in a bilayer containing 512 DMPG molecules, with the amphipathic helices right above the lipid phosphate groups. The atomic level coordinates and interaction parameters for well-equilibrated DMPG lipid bilayer (127) were downloaded from Lipidbook (27). These models were put in periodic boxes, solvated with pre–equilibrated water and counter–ions were added to obtain electrically neutral system. These systems were then taken through steepest descent energy minimization and found to converge to physically realistic minimum energy value with maximum force less than 100 kJ/mol/nm. Then a total of 15 ns equilibration was performed, followed by a 50 ns production run of molecular dynamics.

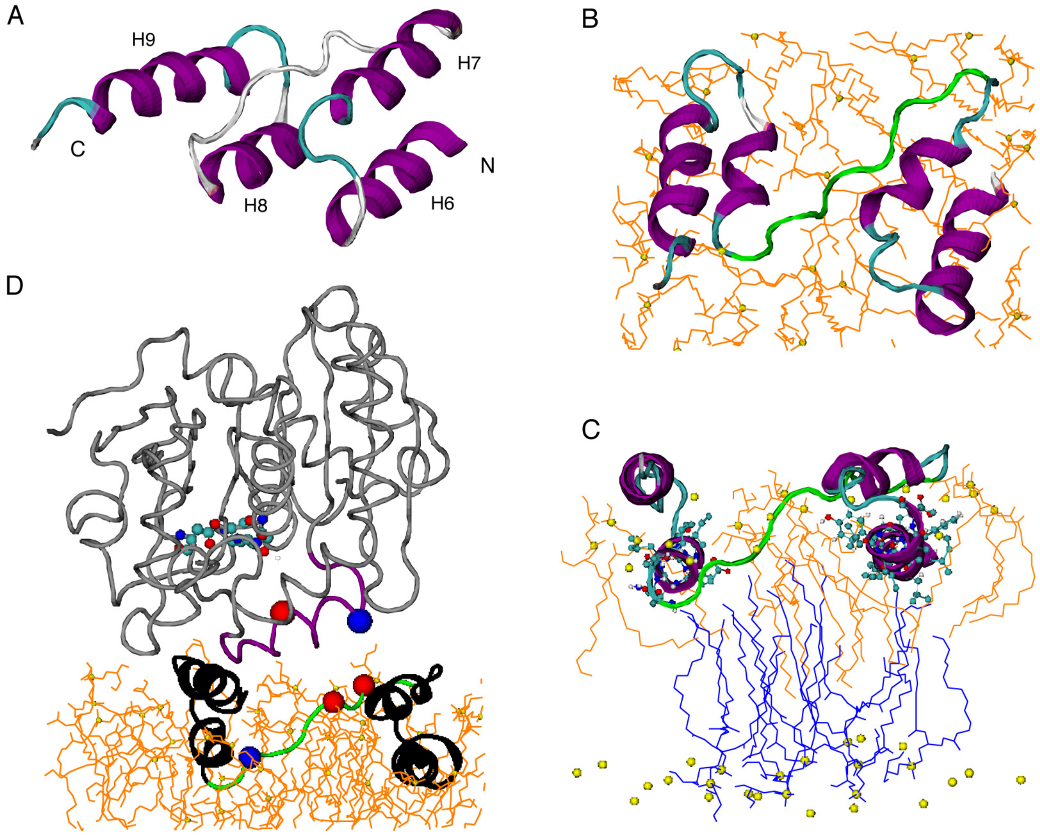


Figure 7.1. The *trans*-model of PLIN1 membrane-anchoring motif. (A) Starting model constructed based on hydrophobicity of predicted helices (hydrophobic H6 and H8, amphipathic H7 and H9) and NMR membrane insertion information. (B) Top view of the model after 50 ns MD simulation in DMPG lipid bilayer. Alpha helices are shown in purple, coils in gray, turns in

cyan, and ^{282}EPENQARP^{289} motif in green. Lipids of the top leaflet are shown in orange, with the phosphorus shown in yellow sphere. (C) Side view of the model showing proximity between helices 6 and 8 with the lipid methyl groups. The lipids from top layer are shown in orange lines while those from bottom layer in blue. (D) Juxtaposition of the triglyceride lipase structure (PDB 4TGL) to the PLIN1 model. PLIN1 is shown in black, with the ^{282}EPENQARP^{289} motif shown in green, alpha carbon of E282 and E284 in red ball, alpha carbon of R288 in blue ball. TGL is shown in gray, with the opened lid (residues 82 to 96) in purple, the exposed catalytic center (S144, D203, H257) in CPK molecular models, alpha carbon of R86 in blue ball, and alpha carbon of D91 in red ball.

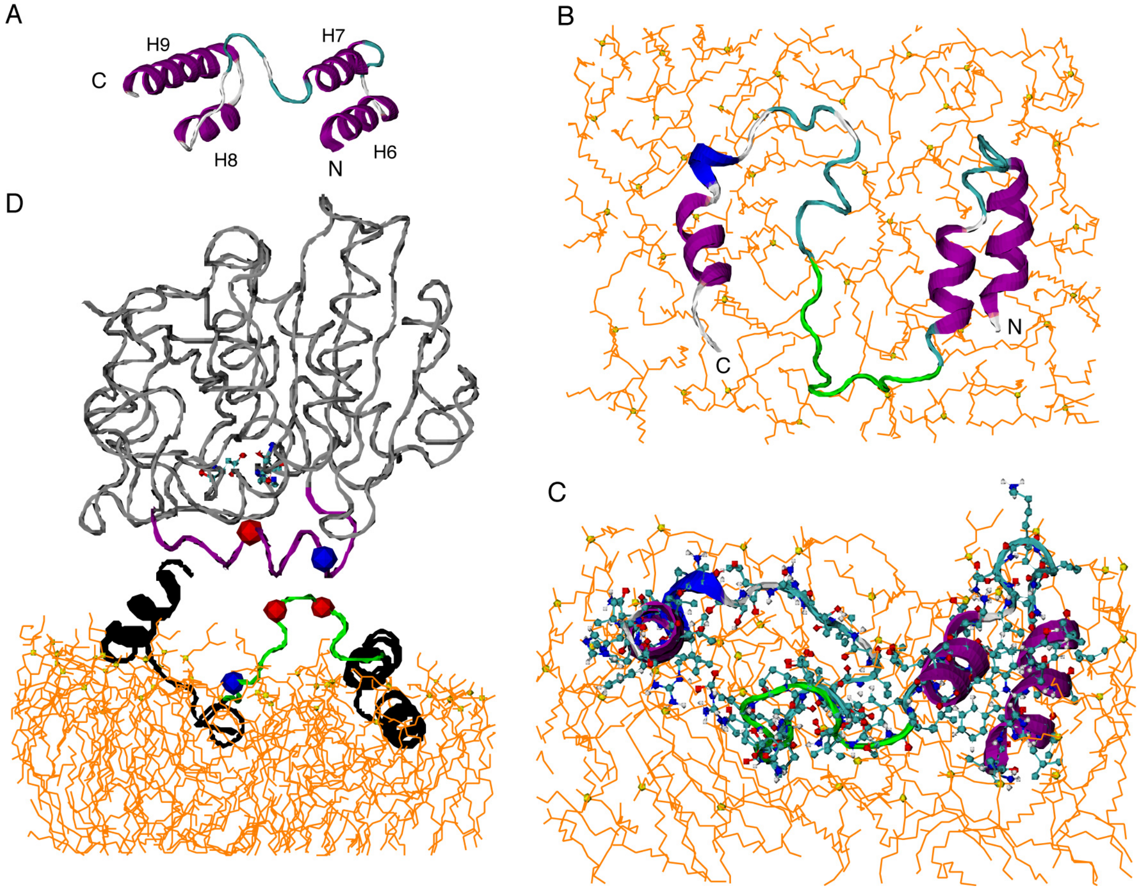


Figure 7.2. The *cis*-model of Lsd1 membrane-anchoring motif. (A) Starting model constructed based on hydrophobicity of predicted helices (hydrophobic H6 and H8, amphipathic H7 and H9) (124) and NMR membrane insertion information. (B) and (C) are two views of the model after 50 ns MD simulation in DMPG lipid bilayer. Alpha helices are shown in purple, 3-10 helices in blue, coils in gray, turns in cyan, and ^{282}EPENQARP^{289} motif in green. Lipids are shown in orange, with the phosphorus shown in yellow sphere. Only one bilayer leaflet is shown since the protein does not penetrate into the other leaflet. (D) Juxtaposition of triglyceride lipase (PDB 4TGL) to Lsd1. Lsd1 is shown in black, with the ^{282}EPENQARP^{289} motif shown in green, alpha carbon of E282 and E284 in red ball, alpha carbon of R288 in blue ball. TGL is shown in gray, with the opened lid (residues 82 to 96) in purple, the exposed catalytic center (S144, D203, H257) in CPK molecular models, alpha carbon of R86 in blue ball, and alpha carbon of D91 in red ball.

The final structure of the *trans*-model is shown in Fig. 7.1 B and 7.1 C. All helices were able to maintain the helical structures, but helices 7 and 8 slightly unwound. In the final structure, H6 and H8 both have close contacts with the terminal methyl groups of the acyl chains from both leaflets, agreeing with the spin diffusion data (Fig. 7.1C). The acyl chains of these lipids from the top leaflet wrapped around helices to accommodate perturbation caused by the protein. Such perturbation also caused the two leaflets to come close locally, allowing the terminal methyl groups from the bottom leaflet contact H6 and H8. Interestingly, lipids from the bottom leaflet were getting very close to the surface in the region enclosed by the helices. This may imply that in real lipid droplet, triglyceride molecules might be brought close to the surface by the same mechanism, ready to leave the droplet in the event of lipolysis.

In Fig. 7.1 D, the crystal structure of a fungus triglyceride lipase (128) is juxtaposed to the PLIN1 model. In this structure, the lid (^{82}SSSIRNWIADLTFVP^{96}, with residues 83 to 84 and 91 to 95 as hinges and charged residues underlined) of the lipase is propped open by diethyl *p*-

nitrophenyl phosphate to expose a patch of hydrophobic area of 800 Å^2 and the catalytic center residues S144, D203, H257. It is interesting that the footage of the lipase matches with PLIN1 *trans*-model, and that the lipase lid has a similar orientation to the ^{282}EPENQARP^{289} (charged residues underlined) motif of PLIN1. More importantly, two possible electrostatic interactions, one between the two glutamates (^{282}E and ^{284}E) on PLIN1 and ^{86}R on the lipase and the other between ^{288}R on PLIN1 and ^{91}D on lipase, may be responsible to open the lipase lid in place of diethyl *p*-nitrophenyl phosphate. There is no structure of any Drosophila triglyceride lipase, however, sequence alignments show that several Drosophila fat body triglyceride lipases share conserved amino acids with the lid segment of the fungus lipase (Fig. 7.3). The charged residues ^{86}R and ^{91}D of the fungus lipase have their counterparts in the Drosophila lipases (Fig. 7.3). Specifically, RLRNFTND of two isoforms of CG8552, which has been considered to be activated by PLIN1 in insects (124, 129), could facilitate electrostatic interaction for three pairs of amino acids with ^{282}EPENQARP^{289} of PLIN1.

```
4TGL       69 GDSEKTIYIVFRGSSSIRNWIAD
CG8552B/C 260 GIDEKLKSITLESIPRLRNFTND
              *  **   *        **   *

4TGL       81 GSSSIRNWIADLTFVPVSYP
CG11055   303 GSSIKVNRLIELPAEPLKLP
              ***   *    *   *   *

4TGL       72 EKTIYIVFRGSSSIRNWIAD
CG8823    365 EKWAHLDFIWGTEARKYVYD
              **     *      *    *
```

Figure 7.3. Sequence alignments of fungus *Rhizomucor miehei* triglyceride lipase chain A (PDB ID: 4TGL) (128) with *Drosophila melanogaster* triglyceride lipases: two isoforms of CG8552 (NP_001188714.1 and NP_001188715.1) (130), CG11055 (NP_611463.1) (130), and CG8823 (NP_477331.1) (131). The alignments were performed using program SIM (132). ^{86}R and ^{91}D of the fungus lipase and corresponding charged residues in other proteins are underlined.

The interaction between the two proteins may also cause reorganization of the PLIN1 structure and further perturbation to the local phospholipid molecules. Energy released from the electrostatic interactions could convert to mechanical energy, causing the lipase to push against H7 and H9 of PLIN1 (Fig. 5D). This could results in an increased area for the region enclosed by the helices, creating a passage for the stored triglyceride molecules, which were already very close to the surface, to diffuse toward the catalytic center. The interaction between PLIN1 and triglyceride lipase might probably be modulated by phosphorylation and Ca^{2+}. Phosphorylation of PLIN1 promotes lipase activity in hydrolyzing triglycerides stored inside the lipid droplet (122, 124) and Ca^{2+} is also an activator of lipolysis (133). These facts suggest that phosphorylation and binding of Ca^{2+} could promote conformational changes affecting certain protein regions. These changes would affect the interaction of the protein with lipid, affecting the accessibility of the lipases to the triglyceride molecules, or the interaction of PLIN1 with lipases and/or other proteins required in the activation process. Thus, future studies of the structures of PLIN1 with and without Ca^{2+} may identify structural changes that would explain the role of certain protein regions in the function of PLIN1.

The starting and final structures of the *cis*-model are shown in Fig. 7.2. Helix 8 is completely unwound and helix 9 also significantly changed its structure. Neither the *trans*-model nor *cis*-model penetrates into the bottom layer of DMPG bilayer, in agreement with the proposed function of attaching the protein to the phospholipids monolayer covering the lipid droplet.

7.1.2 Model verification by NMR data

Backbone and β-carbon chemical shifts were predicted from structures by program shiftX (120). Based on these chemical shifts, ^{13}C-^{13}C 2D spectra were simulated and compared with experimental spectra (Fig. 7.4). The simulated spectrum of *trans*-model agrees very well with experiments. The Cβ-CO peaks of T262 and T291 (lower left corner of Fig. 7.4A) do not have corresponding peaks in the experimental spectrum acquired with 10 ms DARR mixing, which is chosen for establishing short-range (1- and 2-bond) resonance correlations. Matching peaks are found for the experimental spectrum acquired with 50 ms DARR for long-range correlations (Fig. 7.5B). These two threonine residues possibly undergo unfavorable dynamics, resulting in weaker dipolar coupling between Cβ and CO. On the contrary, more unmatched peaks are found for the *cis*-model (Fig. 7.4B and Fig. S7.5B). The T291 and T 262 Cβ-CO peak (lower left corner of Fig. 7.4 B) do not have matching peak even for the spectrum acquired with 50 ms DARR mixing (Fig. S5B). A280, T300, and S303 are also mismatched. Therefore, the *cis*-model does not agree with NMR data.

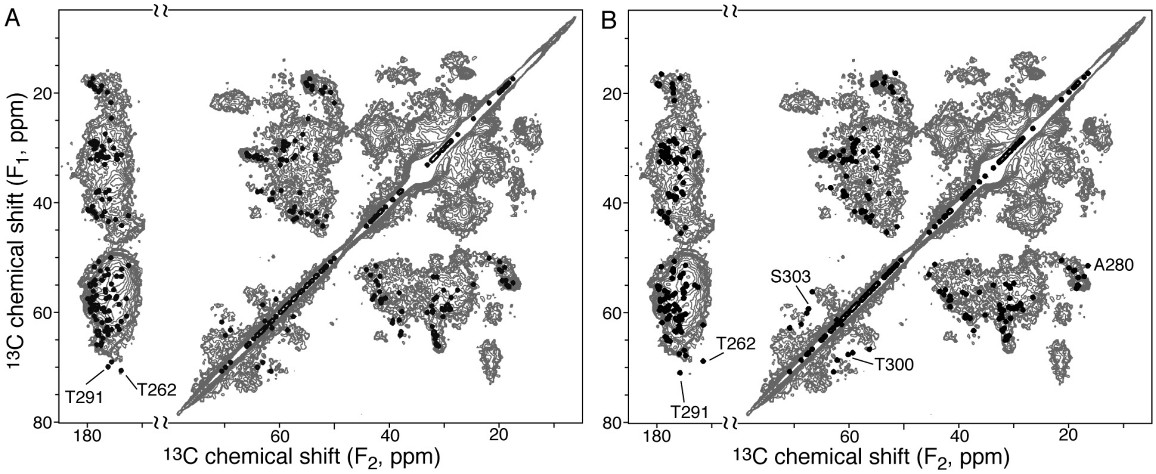


Figure 7.4. Verification of structure models by ^{13}C-^{13}C 2D data. The experimental spectrum (gray) was obtained with 10 ms DARR mixing (134). Simulated spectra (black) for the *trans*-model (A) and *cis*-model (B) only consist of Cα-Cβ, Cα-CO, and Cβ-CO correlations, without other side chain carbons.

Although the *trans*-model agreed well with NMR data, the real structure may deviate from it. First, other regions of the protein might significantly affect the actual structure of the four helices. Second, it is uncertain whether PLIN1 is oligomerized in the active form, and oligomerization has not been taken into consideration in the current model.

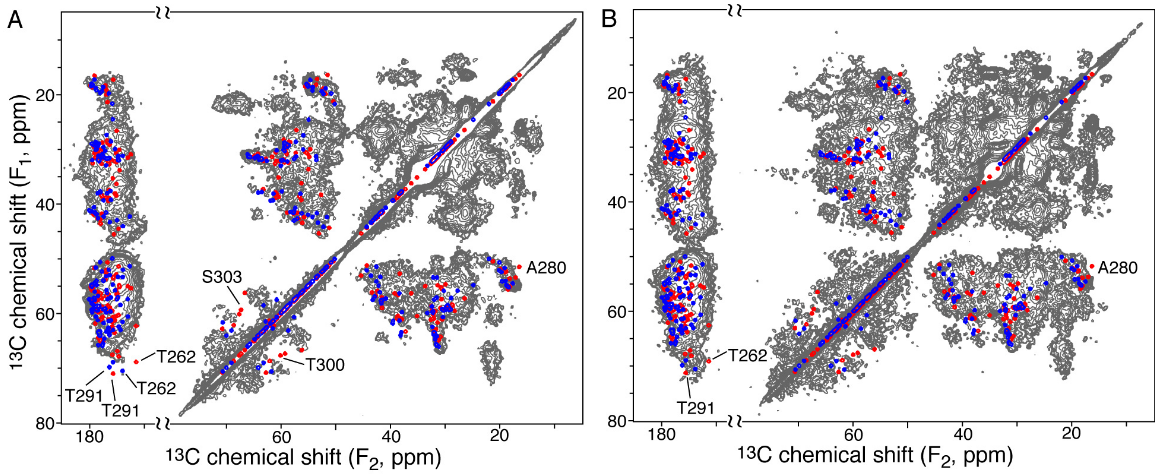


Figure 7.5. Verification of structure models by ^{13}C-^{13}C 2D data with 10 ms (A) and 50 ms (B) DARR mixing. Simulated spectra are shown in blue and red for the *trans*-model and *cis*-model, respectively. The simulated spectra only consist of Cα-Cβ, Cα-CO, and Cβ-CO correlations, without other side chain carbons. Outlying resonances are labeled.

7.2.3 Conclusion

NMR spin diffusion experiments were consistent with the predicted membrane attachment motif of PLIN1, and they indicated that some regions of PLIN1 have a deep contact with the phospholipid acyl chains near the bilayer center. For a native lipid storage droplet that is covered by a lipid monolayer, PLIN1 could penetrate to the interface of the monolayer and the triglyceride core. Two structure models for the membrane attachment motif were generated based on hydrophobicity analysis and NMR membrane insertion depth information, followed by optimization in lipid environment. Both models consist of four membrane interacting elements that are roughly parallel to the membrane surface. Two amphipathic elements stay on the membrane surface, and two hydrophobic elements are buried deeper. Simulated NMR spectra for the *trans*-model agreed with experimental spectra. Juxtaposition of the triglyceride lipase structure to the PLIN1 *trans*-model suggests a possible interaction of the conserved sequence

(EPENQARP), which is on a long loop between lipid binding elements, with the lipase. The long loop could bind to the lipase lid domain by electrostatic interactions and open the lid to expose the catalytic center. Interaction with the lipase could also cause reorganization of the membrane attachment elements of PLIN1, leading to an increased area for the region enclosed by the membrane attachment motif. A passage may be created by this process for the stored triglyceride molecules, which are already very close to the surface due to perturbation by PLN1 attachment, to diffuse toward the catalytic center. Therefore, this structural model could help design future experiments to elucidate the role of PLIN1 in lipolysis.

7.2 Structural study of membrane anchored syntaxin and implication for target-SNARE

Fusion of cellular vesicles with target membranes causes protein or chemical content to be released from the cell, forming the basis of hormone secretion and neurotransmitter release (135). The soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) proteins play a central role in synaptic vesicle exocytosis (135). Energy released when these proteins form a tightly bundled complex drives membrane fusion (136). The best-studied SNARE proteins are those involved in synaptic vesicle exocytosis in neurons: vesicle-SNARE synaptobrevin and target-SNAREs syntaxin and SNAP25. Synaptobrevin and syntaxin are integral membrane proteins, anchored to membranes by their C-terminal transmembrane domains; SNAP25 is tethered to the plasma membrane via several cysteine-linked palmitoyl chains. It has been generally accepted that SNARE complexes form minimal fusion machinery (136), with activities further modulated by interactions with other proteins such as complexin (137) and synaptotagmin (138).

The juxtamembrane linkers of syntaxin (residues 256-265) and synaptobrevin (residues 85-94) have been assumed to determine the required mechanical-energy transfer from the zippering complex to the apposing membranes (139). The secondary structure and bending stiffness of these linkers is still controversial. Whether the syntaxin SNARE motif is secured on membrane surface has important impact on its interaction with SNAP25 when forming the target-SNARE complex. Therefore, detailed characterization of the structure and topology of membrane-anchored syntaxin alone and after forming target-SNARE complex is critical in understanding the mechanism of intercellular membrane fusion and finding possible reaction intermediates. To elucidate structural features of syntaxin in the context of membrane-anchored environment, we use solid-state NMR spectroscopy and MD simulation to study the structure of syntaxin 1A (183-288) reconstituted into phospholipid vesicles. The results suggest that the linker region exists in both helical and loop conformations depending on lipid composition, and that the N-terminal half of the SNARE motif (including the ionic layer residue Gln226) stays away from the lipid membrane. Control samples either prepared without using reducing agent or consisting of syntaxin and SNAP25 complexes at 2:1 ratio show helical-only linker conformation.

7.2.1 Structural model building and verification

We designed and investigated four structural models, which were then subsequently verified with NMR spectra. Model A (Fig. 7.6A), the extended helical structure of the syntaxin was directly extracted from the crystal structure of SNARE complex (PDB ID: 3HD7) (140). For this model, the structure was directly used to generate simulated spectra. Model B (Fig. 7.6B) was designed by first enforcing an L shape of protein Cα trace that was bent in the 261-263 region and then generating the all atom structure using Structural Alphabet based protein Backbone Builder from Alpha Carbon trace (SABBAC) 1.3 (141). This all-atom molecule was first energy minimized in vacuum and then the transmembrane domain was inserted into neutral POPC bilayer using Visual Molecular Dynamics (VMD) (142). Model C (Fig. 7.6C) was built by inserting the transmembrane domain of Model A into neutral POPC bilayer (27). Model D (Fig. 7.6D) was designed by inserting the syntaxin transmembrane domain of Model A in a lipid bilayer containing POPC and about 20% of negatively charged POPS. The POPS structure and interaction parameters were obtained from Tieleman group (143). Models B, C, and D were subjected to optimizations using a series of MD simulations and structures were analyzed after 50 ns of production run, as described in Materials and Methods (144).

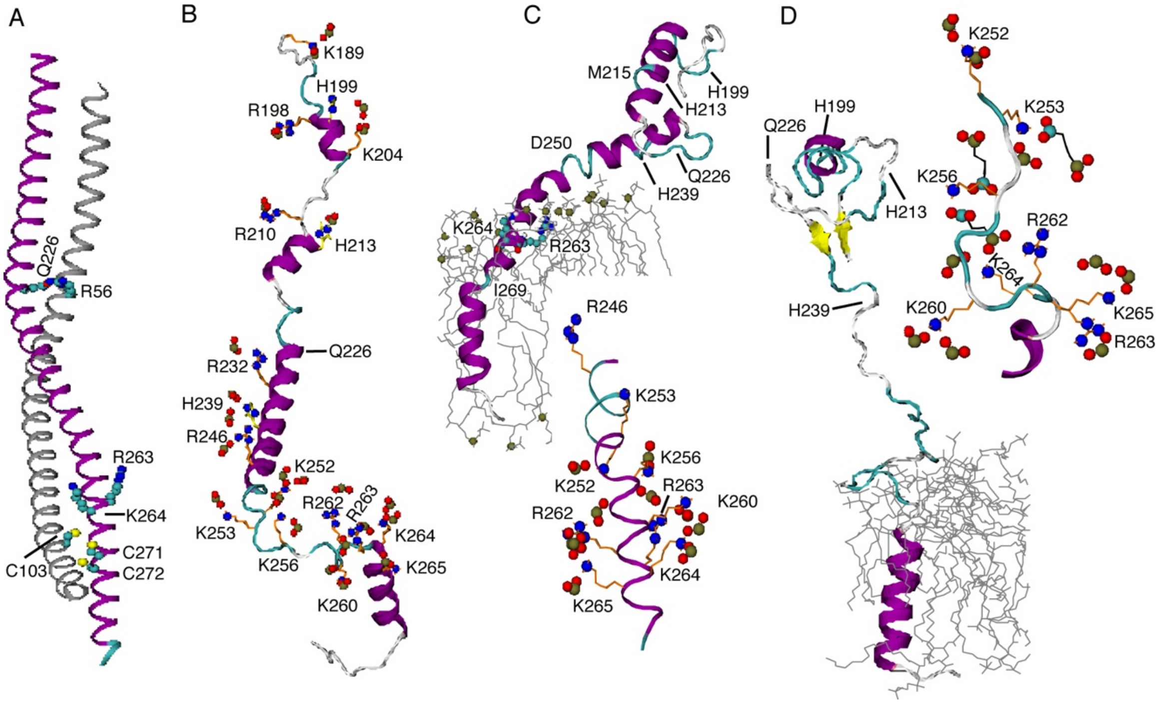


Figure 7.6 Structure models of syntaxin. (A) Model A: extended helical structure of syntaxin with synaptobrevin shown in gray and SNAP25 omitted (PDB ID: 3HD7) (140). (B) Model B: forced L shape formed by the cytoplasmic and transmembrane domain. For clarity, lipids are not shown except for the phosphate groups (phosphorus in tan, oxygen in red for the two oxygen atoms in PO_4 group that are not bonded to carbon) of the lipids interacting with the protein side chains. Sidechains of arginines and lysines are shown in orange sticks. Histidine sidechains are shown in yellow sticks. Sidechain nitrogen atoms are shown in blue balls. (C) Model C: syntaxin in neutral POPC bilayer. Fragment 246-269 is shown in the inset. (D) Model D: syntaxin in negatively charged lipid bilayer (POPC and ~20%POPS). Fragment 252-270 is shown in the inset. The carboxylic carbons groups of POPS are shown with carbons in cyan balls, oxygens in

red. The protein backbone is shown in cartoon with helices in purple, turns in cyan, and coils in silver.

Model A has upright orientation with regard to the lipid bilayer (Fig. 7.6A). Several amino acids mentioned earlier are also shown in the figure: the cysteines in the transmembrane domain and Q226 that forms the zero layer in SNARE complex (145). ^{13}C-^{13}C 2D spectra were simulated based on chemical shift values predicted from the structure model and compared with experimental spectra (Fig. 7.7). Simulated peaks of model A fall within experimental spectral intensities of reconstituted syntaxin with presence of reducing agent (Fig. 7.7). However, peaks clustered together for this predominantly helical model were unable to account for many other experimental peaks.

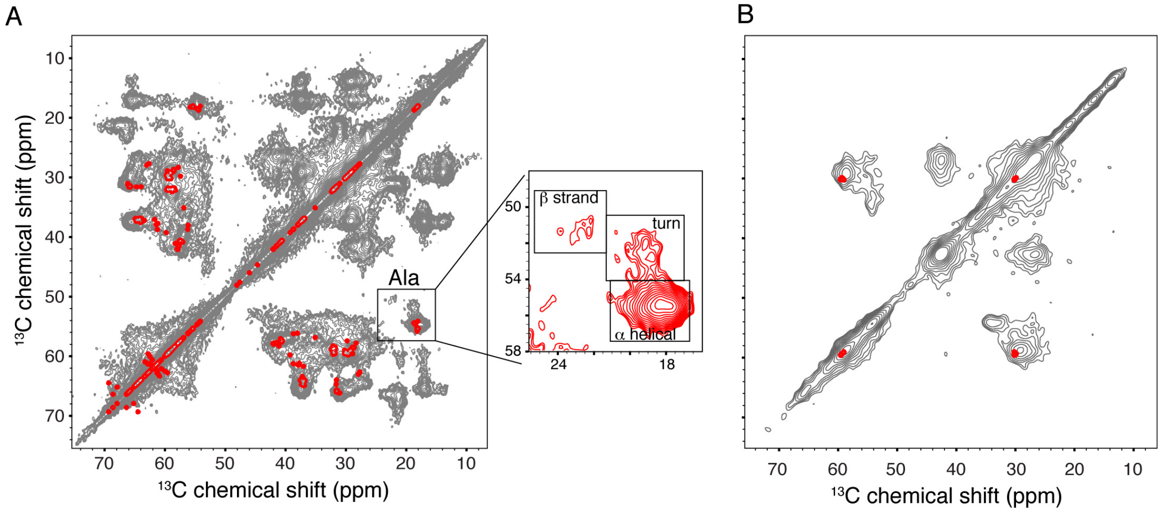


Figure 7.7. Verification of the extended helix structure model A by comparison with NMR spectra. The ^{13}C-^{13}C 2D were acquired with 10 ms DARR mixing for uniform ^{13}C,^{15}N-syntaxin (A) and ^{13}C-R, ^{15}N-K-syntaxin (B). Expansion showing the Ala Cα-Cβ peaks and related secondary structure conformations.

The cytosolic domain of model B in neutral lipids (Fig. 7.6B) remained on the bilayer surface during the MD simulation. Most amino acids of the linker (256-265) were in the turn conformation except for two helical residues K264 and K265. Spectra simulated from this model

(Fig. 7.8) exhibited very good agreement with experimental ^{13}C-^{13}C 2D spectra with uniform ^{13}C and ^{13}C-Arg labeled samples, respectively. Specifically the ^{263}R Cα-Cβ peak agreed with the experimental signal corresponding to the turn conformation (Fig. 7.8B). Most of the basic groups of the arginines and lysines residues in the linker region and in the rest of the cytosolic domain were in close contact with lipid phosphate groups.

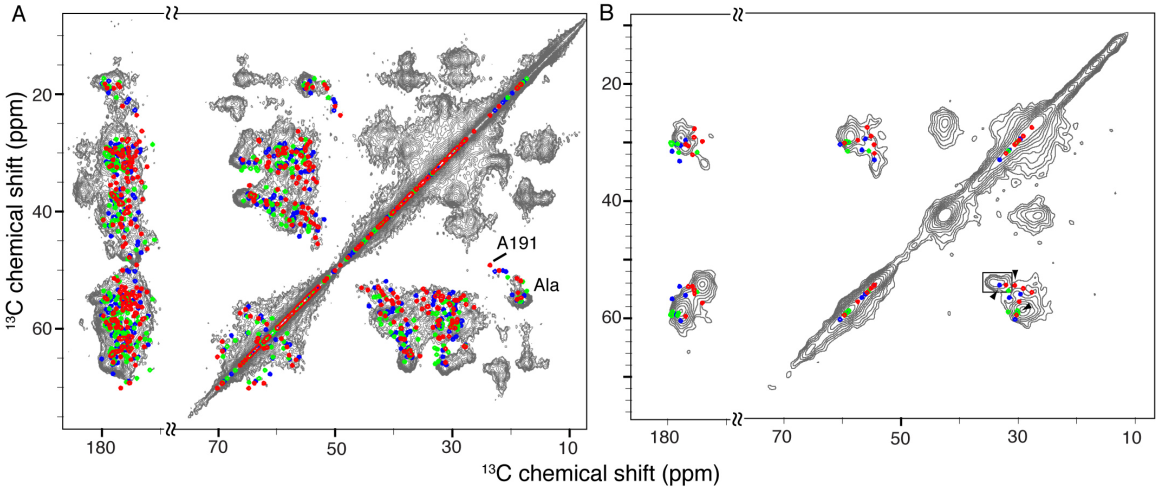


Figure 7.8 Verification of structure models (B in blue, C in green, D in red) by comparison with NMR spectra. The ^{13}C-^{13}C 2D were acquired with 10 ms DARR mixing for uniform ^{13}C,^{15}N-syntaxin (A) and ^{13}C-R, ^{15}N-K-syntaxin (B). Samples were prepared with 85:15 POPC:POPS lipids and a reducing agent. ^{263}R peaks in Fig. 7B are marked by arrowheads. Signals corresponding to turn conformation arginines are marked by a rectangle. Spectrum of the ^{13}C,^{15}N-syntaxin is also shown without overlaying simulated peaks in Fig. S3 of the Supplementary Data.

The linker region (residues 256-265) of model C in neutral lipids remained in helical conformation during the MD simulation. The linker bent away from the normal of lipid membrane surface and the transmembrane helix broke at I269. As shown in the charge analysis of region 246-269 in the inset of Fig. 6C, the side chains of K256, R262, K260, K264 and K265 had close contact with lipid phosphorous head groups, indicating electrostatic interactions between the linker region and POPC lipid membrane. Most of the simulated peaks fell on experimental intensities (Fig. 7.8). However, the peak of ^{263}R, which is in helical conformation in model C, was far away from the turn conformation signals.

The linker of model D in acidic lipids unfolded during MD simulation (Fig. 7.6D). In contrast to model C, the linker of model D is very similar to model B. As shown in the inset of Fig. 7.6D, the basic side chains of K253 and K256 bonded to the negatively charged carboxyl groups of POPS. The interaction between these groups could have provided additional force to tightly pull the linker, which is rich in basic amino acids, to the membrane. It seems the stronger lipid protein interaction was responsible for the unfolding of the helix conformation. Simulated peaks agreed well with experimental spectra (Fig. 7.8). The R263 was in random coil conformation and its simulated peak was quite close to the observed peak that corresponds to turn conformation (Fig. 7.8B).

In model B, the imidazole nitrogens of H239 and H213 were 2.7 and 2.8 Å away from lipid phosphate oxygens, forming strong hydrogen bonds, while H199 had no nearby phosphate groups. Therefore, model B is inconsistent with high sidechain mobility observed in ^{15}N variable temperature experiments by NMR. However in models C and D, all three histidines stayed away from the bilayer surface, agreeing with the sidechain dynamics results. For models C and D, the segment (183-250) that remained in the solution exhibited very different structures, very likely due to misfolding without the stabilization by Habc of full-length syntaxin (146), another syntaxin or other SNARE motifs.

In summary, models A, C, and D could coexist in the sample. Model A has a helical linker and it represents the conformation of syntaxin in homo dimers or in complex with SNAP25. Model C has stiffer helical linker and it needs neutral lipids microenvironment, which could exist in the sample containing POPC and POPS if there are clusters of POPC. Model D captures the flexible linker feature of model B without the need to enforce bent L shape. Rather the linker is “naturally” bent toward the bilayer due to the stronger interaction with acidic lipids. Previous studies showed that the polybasic motif of the linker region has a high affinity to acidic lipids (147) and that mutation of the these basic residues to alanines could lead to membrane fusion defect (148). Moreover, acidic lipids were shown to disperse cholesterol promoted syntaxin clusters through electrostatic interactions with basic linker region of syntaxin (149).

The solution structure of syntaxin in dodecylphosphocholine (DPC) micelle has unfolded linker formed by residues 253-261 (150). Specifically, all arginines, including R263, were in helical conformation. Therefore, the micelle bound structure does not account for the arginine NMR peaks corresponding to turn conformation observed in the lipid bilayer bound syntaxin. In the micelle bound structure, helix break occurred at Q226. In our models (B-D), Q226 is also exposed, either inside or at the edge of a coil region, and is ready to form the zero layer with Q53 and Q174 in SNAP25 and R56 in synaptobrevin (145).

7.2.2 Conclusions

In this study, we employ NMR techniques and MD simulations to investigate the structure of membrane-bound syntaxin before and after forming complexes with SNAP25. The linker region of syntaxin is one of the potential effectors for membrane fusion (147). Different conformations of the linker region may represent the different forms of syntaxin *in vivo*. The linker adopts random coil conformation when dispersed by acidic lipids and exists in monomer. It adopts helical conformation in complexes on clusters. Our MD simulation results confirm these interactions by showing that basic residues in linker region are in close proximity to the head groups of acidic lipid molecules. The conformational change of the linker region may cause the orientation change of the target-SNARE complex and promote its binding to synaptobrevin. Moreover, this dynamic transition from a flexible random coil structure to a stiffer helical structure is likely related to the fusion process of SNARE complex.

REFERENCES

1. Alberts B JA, Lewis J, et al. (2002) *Molecular Biology of the Cell* (Garland Science, New York) 4th Ed.
2. Pain RH (1996) *Mechanisms of Protein Folding (Frontiers in Molecular Biology)* (Oxford University Press) 1st Ed.
3. Brar GA*, et al.* (2012) High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science* 335(6068):552-557.
4. Petronczki M, Siomos MF, & Nasmyth K (2003) Un menage a quatre: The molecular biology of chromosome segregation in meiosis. *Cell* 112(4):423-440.
5. McKee BD (2004) Homologous pairing and chromosome dynamics in meiosis and mitosis. *Bba-Gene Struct Expr* 1677(1-3):165-180.
6. Tsai JH & McKee BD (2011) Homologous pairing and the role of pairing centers in meiosis. *J Cell Sci* 124(12):1955-1963.
7. Petukhova GV*, et al.* (2005) The Hop2 and Mnd1 proteins act in concert with Rad51 and Dmc1 in meiotic recombination. *Nat Struct Mol Biol* 12(5):449-453.
8. Bannister LA*, et al.* (2007) A dominant, recombination-defective allele of Dmc1 causing male-specific sterility. *PLoS Biol* 5(5):e105.
9. Keeney S, Giroux CN, & Kleckner N (1997) Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* 88(3):375-384.
10. Kurumizaka H*, et al.* (2001) Homologous-pairing activity of the human DNA-repair proteins Xrcc3.Rad51C. *Proc Natl Acad Sci U S A* 98(10):5538-5543.
11. Moore JK & Haber JE (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Mol Cell Biol* 16(5):2164-2173.
12. Moktan H*, et al.* (2014) Solution Structure and DNA-binding Properties of the Winged Helix Domain of the Meiotic Recombination HOP2 Protein. *J Biol Chem*.
13. Zhao W*, et al.* (2014) Mechanistic insights into the role of Hop2-Mnd1 in meiotic homologous DNA pairing. *Nucleic Acids Res* 42(2):906-917.
14. Leu JY, Chua PR, & Roeder GS (1998) The meiosis-specific hop2 protein of S-cerevisiae ensures synapsis between homologous chromosomes. *Cell* 94(3):375-386.
15. Pezza RJ, Petukhova GV, Ghirlando R, & Camerini-Otero RD (2006) Molecular activities of meiosis-specific proteins Hop2, Mnd1, and the Hop2-Mnd1 complex. *Journal of Biological Chemistry* 281(27):18426-18434.
16. Pezza RJ, Voloshin ON, Vanevski F, & Camerini-Otero RD (2007) Hop2/Mnd1 acts on two critical steps in Dmc1-promoted homologous pairing. *Gene Dev* 21(14):1758-1766.
17. Pezza RJ*, et al.* (2014) The dual role of HOP2 in mammalian meiotic homologous recombination. *Nucleic Acids Res* 42(4):2346-2357.
18. Purcell EM, Torrey HC, & Pound RV (1946) Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys Rev* 69(1-2):37-38.
19. Bloch F (1946) Nuclear Induction. *Phys Rev* 70(7-8):460-474

20. Williamson MP, Havel TF, & Wuthrich K (1985) Solution Conformation of Proteinase Inhibitor-Iia from Bull Seminal Plasma by H-1 Nuclear Magnetic-Resonance and Distance Geometry. *J Mol Biol* 182(2):295-315.
21. Teng Q (2014) *Structural Biology–Practical NMR Applications* (Springer Science).
22. Bloembergen N, Purcell EM, & Pound RV (1948) Relaxation Effects in Nuclear Magnetic Resonance Absorption. *Phys Rev* 73(7):679-712.
23. Van Der Spoel D*, et al.* (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26(16):1701-1718.
24. Brunne RM & Vangunsteren WF (1993) Dynamic Properties of Bovine Pancreatic Trypsin-Inhibitor from a Molecular-Dynamics Simulation at 5000-Atm. *Febs Lett* 323(3):215-217.
25. Humphrey W, Dalke A, & Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph Model* 14(1):33-38.
26. Maupetit J, Gautier R, & Tuffery P (2006) SABBAC: online structural alphabet-based protein backbone reconstruction from alpha-carbon trace. *Nucleic Acids Research* 34:W147-W151.
27. Domanski J, Stansfeld P, Sansom MP, & Beckstein O (2010) Lipidbook: A Public Repository for Force-Field Parameters Used in Membrane Simulations. *J. Membrane Biol.* 236(3):255-258.
28. Kandt C, Ash WL, & Tieleman DP (2007) Setting up and running molecular dynamics simulations of membrane proteins. *Methods* 41(4):475-488.
29. Faraldo-Gomez JD, Smith GR, & Sansom MS (2002) Setting up and optimization of membrane protein simulations. *Eur Biophys J* 31(3):217-227.
30. Bascom-Slack CA, Ross L.O., and Dawson, D. S. ed (1997) *Chiasmata, crossovers, and meiotic chromosome segregation.*), pp 253-284.
31. Zickler D & Kleckner N (1998) The leptotene-zygotene transition of meiosis. *Annu Rev Genet* 32:619-697.
32. Roeder GS (1997) Meiotic chromosomes: it takes two to tango. *Genes Dev* 11(20):2600-2621.
33. Neale MJ & Keeney S (2006) Clarifying the mechanics of DNA strand exchange in meiotic recombination. *Nature* 442(7099):153-158.
34. Chen YK*, et al.* (2004) Heterodimeric complexes of Hop2 and Mnd1 function with Dmc1 to promote meiotic homolog juxtaposition and strand assimilation. *Proc Natl Acad Sci U S A* 101(29):10572-10577.
35. Petukhova GV*, et al.* (2005) The Hop2 and Mnd1 proteins act in concert with Rad51 and Dmc1 in meiotic recombination. *Nature structural & molecular biology* 12(5):449-453.
36. Pezza RJ, Petukhova GV, Ghirlando R, & Camerini-Otero RD (2006) Molecular activities of meiosis-specific proteins Hop2, Mnd1, and the Hop2-Mnd1 complex. *The Journal of biological chemistry* 281(27):18426-18434.
37. Chi P, San Filippo J, Sehorn MG, Petukhova GV, & Sung P (2007) Bipartite stimulatory action of the Hop2-Mnd1 complex on the Rad51 recombinase. *Genes Dev* 21(14):1747-1757.
38. Zhao W*, et al.* (2013) Mechanistic insights into the role of Hop2-Mnd1 in meiotic homologous DNA pairing. *Nucleic Acids Res*.
39. Leu JY, Chua PR, & Roeder GS (1998) The meiosis-specific Hop2 protein of S. cerevisiae ensures synapsis between homologous chromosomes. *Cell* 94(3):375-386.
40. Petukhova GV, Romanienko PJ, & Camerini-Otero RD (2003) The Hop2 protein has a direct role in promoting interhomolog interactions during mouse meiosis. *Developmental cell* 5(6):927-936.

41. Pezza RJ*, et al.* (2013) The dual role of HOP2 in mammalian meiotic homologous recombination. *Nucleic Acids Res*.
42. Gajiwala KS & Burley SK (2000) Winged helix proteins. *Current opinion in structural biology* 10(1):110-116.
43. Delaglio F*, et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* 6(3):277-293.
44. Bahrami A, Assadi AH, Markley JL, & Eghbalnia HR (2009) Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. *PLoS computational biology* 5(3):e1000307.
45. Berjanskii MV & Wishart DS (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* 127(43):14970-14971.
46. Shen Y, Delaglio F, Cornilescu G, & Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44(4):213-223.
47. Shen Y*, et al.* (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. USA* 105(12):4685-4690.
48. Shen Y, Vernon R, Baker D, & Bax A (2009) De novo protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR* 43(2):63-78.
49. Raman S*, et al.* (2009) Accurate Automated Protein NMR Structure Determination Using Unassigned NOESY Data. *J. Am. Chem. Soc.* 132(1):202-207.
50. Tang Y*, et al.* (2010) Fully automated high-quality NMR structure determination of small 2H-enriched proteins. *J. Struct. Funct. Genom.* 11(4):223-232.
51. Huang YJ, Powers R, & Montelione GT (2005) Protein NMR Recall, Precision, and F-measure Scores (RPF Scores):,Äö√Ñ√∂,àö√ë,àö,àÇ,Äö√†√∂,àö√´,Äö√†√∂¬¨¢ Structure Quality Assessment Measures Based on Information Retrieval Statistics. *J. Am. Chem. Soc.* 127(6):1665-1674.
52. Laskowski RA, MacArthur MW, Moss DS, & Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26(2):283-291.
53. Hooft RW, Vriend G, Sander C, & Abola EE (1996) Errors in protein structures. *Nature* 381(6580):272.
54. Boudet J*, et al.* (2007) Conformational and thermodynamic changes of the repressor/DNA operator complex upon monomerization shed new light on regulation mechanisms of bacterial resistance against beta-lactam antibiotics. *Nucleic Acids Res* 35(13):4384-4395.
55. Ko J*, et al.* (2011) The FALC-Loop web server for protein loop modeling. *Nucleic Acids Res* 39(Web Server issue):W210-214.
56. Gibson DG*, et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* 6(5):343-345.
57. Shen Y, Delaglio F, Cornilescu G, & Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44(4):213-223.
58. Spronk CAEM, Nabuurs SB, Krieger E, Vriend G, & Vuister GW (2004) Validation of protein structures derived by NMR spectroscopy. *Prog Nucl Mag Res Sp* 45(3-4):315-337.
59. Humphrey W, Dalke A, & Schulten K (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* 14(1):33-38, 27-38.
60. Melckebeke HV*, et al.* (2003) Solution structural study of BlaI: implications for the repression of genes involved in beta-lactam antibiotic resistance. *J Mol Biol* 333(4):711-720.

61. Tsai KL*, et al.* (2007) Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification. *Nucleic Acids Res* 35(20):6984-6994.
62. Kitano K, Kim SY, & Hakoshima T (2010) Structural basis for DNA strand separation by the unconventional winged-helix domain of RecQ helicase WRN. *Structure* 18(2):177-187.
63. Zhang AP, Pigli YZ, & Rice PA (2010) Structure of the LexA-DNA complex and implications for SOS box measurement. *Nature* 466(7308):883-886.
64. Jin C, Marsden I, Chen X, & Liao X (1999) Dynamic DNA contacts observed in the NMR structure of winged helix protein-DNA complex. *J Mol Biol* 289(4):683-690.
65. Blanco AG, Sola M, Gomis-Ruth FX, & Coll M (2002) Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure* 10(5):701-713.
66. Garnett JA, Marincs F, Baumberg S, Stockley PG, & Phillips SE (2008) Structure and function of the arginine repressor-operator complex from Bacillus subtilis. *J Mol Biol* 379(2):284-298.
67. Boura E, Rezabkova L, Brynda J, Obsilova V, & Obsil T (2010) Structure of the human FOXO4-DBD-DNA complex at 1.9 A resolution reveals new details of FOXO binding to the DNA. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 12):1351-1357.
68. Gajiwala KS*, et al.* (2000) Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* 403(6772):916-921.
69. Nicholls A, Sharp KA, & Honig B (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11(4):281-296.
70. Costa RH, Grayson DR, & Darnell JE, Jr. (1989) Multiple hepatocyte-enriched nuclear factors function in the regulation of transthyretin and alpha 1-antitrypsin genes. *Mol Cell Biol* 9(4):1415-1425.
71. Finnin MS*, et al.* (1997) The activation domain of the MotA transcription factor from bacteriophage T4. *The EMBO journal* 16(8):1992-2003.
72. Moktan H. DDH (2017) Wing 1 of protein HOP2 is as important as helix 3 in DNA binding by MD simulation. *J Biomol Struct Dyn.*
73. Kang HA*, et al.* (2015) Crystal structure of Hop2-Mnd1 and mechanistic insights into its role in meiotic recombination. *Nucleic Acids Research* 43(7):3841-3856.
74. Chan YL, Brown MS, Qin DM, Handa N, & Bishop DK (2014) The Third Exon of the Budding Yeast Meiotic Recombination Gene HOP2 Is Required for Calcium-dependent and Recombinase Dmc1-specific Stimulation of Homologous Strand Assimilation. *Journal of Biological Chemistry* 289(26):18076-18086.
75. Gajiwala KS & Burley SK (2000) Winged helix proteins. *Curr Opin Struc Biol* 10(1):110-116.
76. Yin DX & Mackerell AD (1998) Combined ab initio/empirical approach for optimization of Lennard–Jones parameters. *J. Comput. Chem.* 19:334–348.
77. Zheng GH, Lu XJ, & Olson WK (2009) Web 3DNA-a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Research* 37:W240-W246.
78. Pierce BG*, et al.* (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* 30(12):1771-1773.
79. Safo MK*, et al.* (2005) Crystal structures of the BlaI repressor from Staphylococcus aureus and its complex with DNA: Insights into transcriptional regulation of the bla and mec operons. *J Bacteriol* 187(5):1833-1844.
80. Jorgensen WL & Madura JD (1983) Quantum and Statistical Studies of Liquids .25. Solvation and Conformation of Methanol in Water. *J Am Chem Soc* 105(6):1407-1413.

81. Bussi G, Donadio D, & Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126(1).
82. Parrinello M & Rahman A (1981) Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J Appl Phys* 52(12):7182-7190.
83. Hoover WG (1985) Canonical Dynamics - Equilibrium Phase-Space Distributions. *Phys Rev A* 31(3):1695-1697.
84. Nose S (1984) A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol Phys* 52(2):255-268.
85. Nose S & Klein ML (1983) Constant Pressure Molecular-Dynamics for Molecular-Systems. *Mol Phys* 50(5):1055-1076.
86. Hockney RW, Goel SP, & Eastwood JW (1974) Quiet High-Resolution Computer Models of a Plasma. *J Comput Phys* 14(2):148-158.
87. Darden T, York D, & Pedersen L (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J Chem Phys* 98(12):10089-10092.
88. Hess B, Bekker H, Berendsen HJC, & Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* 18(12):1463-1472.
89. Ma L, Sundlass NK, Raines RT, & Cui Q (2011) Disruption and Formation of Surface Salt Bridges Are Coupled to DNA Binding by the Integration Host Factor: A Computational Analysis. *Biochemistry-Us* 50(2):266-275.
90. Sittel F, Jain A, & Stock G (2014) Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates. *J Chem Phys* 141(1).
91. Hou TJ, Wang JM, Li YY, & Wang W (2011) Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J Chem Inf Model* 51(1):69-82.
92. Kollman PA*, et al.* (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts Chem Res* 33(12):889-897.
93. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, & Case DA (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J Am Chem Soc* 120(37):9401-9409.
94. Homeyer N & Gohlke H (2012) Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol Inform* 31(2):114-122.
95. Kumari R, Kumar R, Lynn A, & Consort OSDD (2014) g_mmpbsa-A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J Chem Inf Model* 54(7):1951-1962.
96. Schutz CN & Warshel A (2001) What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins-Structure Function and Bioinformatics* 44(4):400-417.
97. Ramos RM & Moreira IS (2013) Computational Alanine Scanning Mutagenesis-An Improved Methodological Approach for Protein-DNA Complexes. *J Chem Theory Comput* 9(9):4243-4256.
98. Li L, Li C, Zhang Z, & Alexov E (2013) On the 'Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J Chem Theory Comput* 9(4):2126-2136.
99. Cuervo A*, et al.* (2014) Direct measurement of the dielectric polarization properties of DNA. *P Natl Acad Sci USA* 111(35):E3624-E3630.
100. Gronenborn AM*, et al.* (1991) A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein-G. *Science* 253(5020):657-661.
101. Passner JM, Ryoo HD, Shen LY, Mann RS, & Aggarwal AK (1999) Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature* 397(6721):714-719.

102. Coulocheri SA, Pigis DG, Papavassiliou KA, & Papavassiliou AG (2007) Hydrogen bonds in protein-DNA complexes: Where geometry meets plasticity. *Biochimie* 89(11):1291-1303.
103. Luscombe NM, Laskowski RA, & Thornton JM (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Research* 29(13):2860-2874.
104. Anderson KM*, et al.* (2013) Direct Observation of the Ion-Pair Dynamics at a Protein-DNA Interface by NMR Spectroscopy. *J Am Chem Soc* 135(9):3613-3619.
105. Esadze A, Li DW, Wang T, Bruschweiler R, & Iwahara J (2011) Dynamics of lysine side-chain amino groups in a protein studied by heteronuclear 1H-15N NMR spectroscopy. *J Am Chem Soc* 133(4):909-919.
106. Xu D, Tsai CJ, & Nussinov R (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng* 10(9):999-1012.
107. Ha JH, Spolar RS, & Record MT (1989) Role of the Hydrophobic Effect in Stability of Site-Specific Protein-DNA Complexes. *J Mol Biol* 209(4):801-816.
108. Meyer EE, Rosenberg KJ, & Israelachvili J (2006) Recent progress in understanding hydrophobic interactions. *P Natl Acad Sci USA* 103(43):15739-15746.
109. Israelachvili J & Pashley R (1982) The Hydrophobic Interaction Is Long-Range, Decaying Exponentially with Distance. *Nature* 300(5890):341-342.
110. Maisuradze GG, Liwo A, & Scheraga HA (2010) Relation between Free Energy Landscapes of Proteins and Dynamics. *J Chem Theory Comput* 6(2):583-595.
111. Pronk S*, et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845-854.
112. Schmid N*, et al.* (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J Biophy* 40(7):843-856.
113. Hess B, Bekker H, Berendsen HJC, & Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18(12):1463-1472.
114. Darden T, York D, & Pedersen L (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* 98(12):10089-10092.
115. Bussi G, Donadio D, & Parrinello M (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126(1): 014101:014101-014107.
116. Parrinello M & Rahman A (1981) Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J. Appl. Phys.* 52(12):7182-7190.
117. Nose S (1984) A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* 52(2):255-268.
118. Nose S & Klein ML (1983) Constant Pressure Molecular-Dynamics for Molecular-Systems. *Mol. Phys.* 50(5):1055-1076.
119. Hoover WG (1985) Canonical Dynamics - Equilibrium Phase-Space Distributions. *Phys. Rev. A* 31(3):1695-1697.
120. Neal S, Nip AM, Zhang H, & Wishart DS (2003) Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. *J. Biomol. NMR* 26:215-240.
121. Wolins NE, Brasaemle DL, & Bickel PE (2006) A proposed model of fat packaging by exchangeable lipid droplet proteins. *Febs Lett* 580(23):5484-5491.
122. Patel RT, Soulages JL, Hariharasundaram B, & Arrese EL (2005) Activation of the lipid droplet controls the rate of lipolysis of triglycerides in the insect fat body. *J. Biol. Chem.* 280:22624-22631.
123. Tusnady GE & Simon I (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17(9):849-850.
124. Arrese EL*, et al.* (2008) Function and structure of lipid storage droplet protein 1 studied in lipoprotein complexes. *Arch. Biochem. Biophys.* 473(1):42-47.

125. Humphrey W, Dalke A, & Schulten K (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* 14(1):33-38, 27-38.

126. Pronk S*, et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845-854.

127. Piggot TJ, Holdbrook DA, & Khalid S (2011) Electroporation of the E. coli and S. Aureus Membranes: Molecular Dynamics Simulations of Complex Bacterial Membranes. *J. Phys. Chem. B* 115(45):13381-13388.

128. Derewenda U, Brzozowski AM, Lawson DM, & Derewenda ZS (1992) Catalysis at the interface: the anatomy of a conformational change in a triglyceride lipase. *Biochemistry-Us* 31(5):1532-1541.

129. Bickel PE, Tansey JT, & Welte MA (2009) PAT proteins, an ancient family of lipid droplet proteins that regulate cellular lipid stores. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* 1791(6):419-440.

130. Hoskins RA*, et al.* (2007) Sequence Finishing and Mapping of Drosophila melanogaster Heterochromatin. *Science* 316(5831):1625-1628.

131. Zinke I, Kirchner C, Chao LC, Tetzlaff MT, & Pankratz MJ (1999) Suppression of food intake and growth by amino acids in Drosophila: the role of pumpless, a fat body expressed gene with homology to vertebrate glycine cleavage system. *Development* 126(23):5275-5284.

132. Huang X & Miller W (1991) A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* 12(3):337-357.

133. Arrese EL, Flowers MT, Gazard JL, & Wells MA (1999) Calcium and cAMP are second messengers in the adipokinetic hormone-induced lipolysis of triacylglycerols in Manduca sexta fat body. *J. Lipid Res.* 40(3):556-564.

134. Takegoshi K, Nakamura S, & Terao T (2001) ^{13}C-^{1}H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett.* 344(5-6):631-637.

135. Brunger AT (2001) Structure of Proteins Involved in Synaptic Vesicle Fusion in Neurons. *Annual Review of Biophysics and Biomolecular Structure* 30(1):157-171.

136. Weber T*, et al.* (1998) SNAREpins: Minimal Machinery for Membrane Fusion. *Cell* 92(6):759-772.

137. Chen X*, et al.* (2002) Three-dimensional structure of the complexin/SNARE complex. *Neuron* 33(3):397-409.

138. Ernst JA & Brunger AT (2003) High Resolution Structure, Stability, and Synaptotagmin Binding of a Truncated Neuronal SNARE Complex. *J. Biol. Chem.* 278(10):8630-8636.

139. Knecht V & Grubmuller H (2003) Mechanical coupling via the membrane fusion SNARE protein syntaxin 1A: a molecular dynamics study. *Biophys. J.* 84:1527-1547.

140. Stein A, Weber G, Wahl MC, & Jahn R (2009) Helical extension of the neuronal SNARE complex into the membrane. *Nature* 460(7254):525-528.

141. Maupetit J, Gautier R, & Tuffery P (2006) SABBAC: online Structural Alphabet-based protein BackBone reconstruction from Alpha-Carbon trace. *Nucleic Acids Res.* 34(Web Server issue):W147-151.

142. Humphrey W, Dalke A, & Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14(1):33-38.

143. Mukhopadhyay P, Monticelli L, & Tieleman DP (2004) Molecular dynamics simulation of a palmitoyl-oleoyl phosphatidylserine bilayer with Na+ counterions and NaCl. *Biophysical journal* 86(3):1601-1609.

144. Pronk S*, et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845-854.

145. Sutton RB, Fasshauer D, Jahn R, & Brunger AT (1998) Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. *Nature* 395(6700):347-353.

146. Fernandez I*, et al.* (1998) Three-dimensional structure of an evolutionarily conserved N-terminal domain of syntaxin 1A. *Cell* 94(6):841-849.
147. Murray DH & Tamm LK (2009) Clustering of Syntaxin-1A in Model Membranes Is Modulated by Phosphatidylinositol 4,5-Bisphosphate and Cholesterol. *Biochem.* 48(21):4617-4625.
148. Lam AD, Tryoen-Toth P, Tsai B, Vitale N, & Stuenkel EL (2008) SNARE-catalyzed fusion events are regulated by Syntaxin1A-lipid interactions. *Mol. Biol. Cell* 19(2):485-497.
149. Murray DH & Tamm LK (2011) Molecular Mechanism of Cholesterol- and Polyphosphoinositide-Mediated Syntaxin Clustering. *Biochemistry-Us* 50(42):9014-9022.
150. Liang B, Kiessling V, & Tamm LK (2013) Prefusion structure of syntaxin-1A suggests pathway for folding into neuronal trans-SNARE complex fusion intermediate. *P. Natl. Acad. Sci. USA* 110(48):19384-19389.

APPENDICES

A1. MD simulation of N-terminal HOP2 and DNA complex protocol

This protocol is used to simulate the HOP2-DNA complex by using software GROMACS

Method:

1. System generation
   a. Download the pdb file for N-terminal HOP2 from Protein Data Bank (ID: 2MH2)
   b. Add acetyl-group and amide-group to N-terminal and C-terminal of the protein respectively using VMD.
   c. Run energy minimization in vacuum. This will remove any steric–clashes that may be present in crude system.
   d. Generate dsDNA structure in B–form using 3DNA webserver. Use the DNA sequence ^{1}TAGTAGTATGAATTTGCCGCAAGATCTGAT^{30}.
   e. Equilibrate both separately in water for 20ns.
2. Docking
   a. Use Z–dock webserver to dock protein in DNA. Provide the interacting residues (Gly11, Gly14, Phe33, Ala45, Ala46, Lys49, Asp52, Gln68, Tyr71) identified from the NMR chemical shift perturbation study for accurate docking calculations.
   b. Delete any DNA bases very far from protein sites to minimize the system size. This will speed up the simulation.

      Deleted DNA bases: 1–5:56–60 and 20–30:31–4

Adjust the DNA caps accordingly.

c. In the DNA pdb file, deleted atom “P” from the beginning of the strands and changed atom name ‘C5M’ to ‘C7’ for forcefield consistency.

protein_init.pdb is the structure of the protein only from the protein–DNA complex.
trunc_DNA.pdb is the initial structure for DNA only.

The GROMACS commands were used to carry out the following.

1. Topology generation

a.
```
pdb2gmx -f protein_init.pdb -o protein.pdb -ter -water spce
```

b.
```
pdb2gmx -f truncDNA.pdb -o truncatedDNA.pdb -p topolDNA.top
```

Total charge of the DNA system is -26e.

Charge on the protein is 3e.

So the total charge of the system will be -23e.

You have to edit the topology file topol.top to include the topology of DNA into it. Just copy appropriate sections of topolDNA.top into the topol.top in order.

2. Now concatenate the protein and DNA pdb files in a single file.

```
cat protein_init.pdb truncatedDNA.pdb > system.pdb
```

3. Create a box for the system. Put the whole system at the center of the box Leaving at least 1.5nm distance from the surface of the complex.

```
editconf -f system.pdb -o boxed.gro -bt triclinic -c -d 1.5
```

4. Add water:

```
genbox -cs spc216.gro -cp boxed.gro -p topol.top -o system_solv.gro
```

5. To add ions in system to make the overall concentration of ions to 0.12M.

```
grompp -f ions.mdp -p topol.top -c system_solv.gro -o ions.tpr
genion -s ions.tpr -o system_solv_ions.gro -p topol.top -pname NA -nname CL
-conc 0.120
```

6. Carry out Energy Minimization:

```
grompp -f minim.mdp -c system_solv_ions.gro -p topol.top -o em.tpr
mdrun -v -deffnm em
```

7. Generate system index file:

```
make_ndx -f em.gro -o index.ndx
```

Select indices: DNA_Protein and SOL_NA (These groups can be selected for different output profile)

8. Carry out equilibration run (These should be run in cluster computer)

NVT equilibration:

```
grompp -f nvt.mdp -c em.gro -p topol.top -n index.ndx -o nvt.tpr
mdrun -v -deffnm nvt
```

NPT equilibration:

```
grompp -f npt.mdp -c nvt.gro -p topol.top -n index.ndx -o npt.tpr
mdrun -v -deffnm npt
```

9. Final run: in cluster:

```
grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -n index.ndx -o final.tpr
mdrun -v -deffnm final
```

A2. Protocol for LSD1 protein model building and simulation:

The overall model building and simulation using VMD and GROMACS is summarized below.

1. The Cα trace for most probable membrane–binding motif (residues 249–318) was designed in VMD. This is done in VMD by taking this motif from the predicted model structure and making changes to the position of secondary structure elements. Only Cα coordinates were saved. The positions of Cα were changed accordingly in VMD to fit the hypothesis.
2. The all-atom molecule structure was generated by Structural Alphabet based protein Backbone Builder from alpha Carbon trace (SABBAC) 1.3.
3. Two models trans– and cis–models were generated and saved as model1.pdb and model2.pdb.
4. These two structures were energy minimized in vacuum using GROMACS 4.5.5 and resulting energy–minimized structures were saved as heal_model1.pdb and heal_model2.pdb. These protein structures were taken as initial protein structures.
5. Well–equilibrated DMPG bilayer with 128 lipid molecules 'dmpg_bilayer.pdb' and interaction parameters 'dmpg.itp' were downloaded from LIPIDBOOK.
6. The protein models were placed on the bilayer so that the hydrophobic helices reside inside the bilayer however the amphipathic helices stay on the surface of the bilayer. VMD was used for this as following:
    a. Load both protein and bilayer structures in VMD.
    b. Select protein and use 'Mouse → Move → Molecule' from VMD main menu. Or press 8 in keyboard. Just click on the protein molecule and hold it to move to desired position.

c. For MAC (UNIX) or LINUX systems, use combination of ‘shift’ and ‘option’ keys while holding the molecule with mouse to rotate the protein molecule for precise positioning in the bilayer.

d. Once the position of the protein is satisfactory, save the coordinates of the protein as ‘aligned_model1.pdb’.

e. While both molecules are still open in VMD and aligned in proper positions, check lipid molecules that interfere directly with the protein. Look for lipids that have tails forked through protein helix. Note the residue number of such lipids and delete them from the original dmpg_bilayer.pdb file. On doing this, we can avoid major steric–clashes. Check all possible lipid molecules that may pose any clashes with protein.

f. Generate the protein topology:

```
pdb2gmx –f  aligned_model1.pdb –o protein_init.pdb –p topol.top
```

Choose appropriate force field and terminals.

Note the total charge on the protein system. Its ‘-1e’, and its necessary to know the total charge as we need to add equal number of opposite charges to neutralize the system.

The ‘topol.top’ file is the topology file for the protein. We need to add ‘dmpg.itp’ file in the appropriate section of this file to include the interaction parameters for lipids.

g. Concatenate the protein and bilayer coordinate files:

```
cat protein_init.pdb dmpg_bilayer_somedeleted.pdb > system.pdb
```

h. Create a box for the system. Put the whole system at the center of the box Leaving at least 1.5nm distance from the surface of the complex.

```
editconf -f system.pdb -o boxed.gro -bt triclinic -c -d 1.5
```

Copy ‘vdwradii.dat’ file from the forcefield folder to the local folder. Change the radii for ‘C’ from 0.15 to 0.375. This will prevent water molecules getting inside the bilayer from the gaps created by deleting some lipid molecules.

i. Add water:

```
genbox -cs spc216.gro -cp boxed.gro -p topol.top -o system_solv.gro
```

After adding water, make sure to delete vdwradii.dat file or change the radii for ‘C’ to its original value.

j. Add ions in system to neutralize it:

```
→grompp -f ions.mdp -p topol.top -c system_solv.gro -o ions.tpr
→genion -s ions.tpr -o system_solv_ions.gro -p topol.top -pname NA -nname CL –nn (number) –np (number)
```

Replace (number) with the exact number of charge in the system. For example if the total charge of the system is -40e then add –np 40. nn→ number negative; np → number positive.

SELECT the index ‘SOL’ when prompted. This will replace random water molecules from the system with appropriate charged ions.

k. Carry out EM: ‘minim.mdp’ is MD parameter file.

```
grompp -f minim.mdp -c system_solv_ions.gro -p topol.top -o em.tpr
mdrun -v -deffnm em
```

l. Generate system index file:

```
make_ndx -f em.gro -o index.ndx
```

select indices: Protein_DMPG and SOL_NA (or SOL_CL) (These groups can be selected for different output profile)

m. Carry out equilibration run: (Run this in cluster)

NVT equilibration:

```
grompp -f nvt.mdp -c em.gro -p topol.top -n index.ndx -o nvt.tpr
mdrun -v -deffnm nvt
```

NPT equilibration:

```
grompp -f npt.mdp -c nvt.gro -p topol.top -n index.ndx -o npt.tpr
mdrun -v -deffnm npt
```

n. Final run: in cluster:

```
grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -n index.ndx -o final.tpr
mdrun -v -deffnm final
```

The final structure file ‘final.gro’ is considered as the simulated structure and analyzed for this work. The trajectory files final.xtc and final.trr can be analyzed for various other properties.

The md parameter ‘*.mdp’ files used here are copied from internet and modified to suit the need.

This protocol in general works for any membrane protein system. For different lipids, just use different interaction parameter files.

The final protein structure is used to generate the chemical shifts ‘predicted chemical shifts’ using SHIFTX webserver. The *.bmrb’ file that we get from the SHIFTX output is further processed to get CC2D spectrum. This predicted spectrum is then compared with the experimental CC2D NMR spectra.

A3. Protocol for generating equilibrated mixed lipid bilayer (POPC+POPS).

a. There are several methods to generate the mixed lipid bilayer. The MemBuilderII software can effectively be used to generate bilayer structure and topology file consistent with the force field of interest. Visit the site http://bioinf.modares.ac.ir/software/mb2/builder.php.

b. Membuilder may not have all the options for all possible lipids. In that case use the following procedure:

   - Generate a bilayer of appropriate size bilayer containing only one type of lipid (POPC). Or download from some other sources.
   - Download the structure of a single lipid (POPS) and interaction parameters for that lipid. It's available at LIPIDBOOK.
   - Using VMD align one POPS molecule with one random POPC molecule and save POPS molecule as 'pops1.pdb'. Find the residue number of POPC molecule and delete it from the original POPC bilayer file.
   - Concatenate pops1.pdb and popc_bilayer.pdb into 'system1.pdb'. Note 'system1.pdb' now has one of the POPC replaced with POPS. Energy minimize 'system1.pdb' in vacuum.
   - Use similar procedure and replace POPC molecules by POPS molecules until desired ratio is met.

The structure models of Syntaxin1A were simulated in a mixed lipid bilayer created by using the protocol 3b. Part of protocol 2 was used for complete MD simulation.

VITA

Hem Bahadur Moktan Tamang

Candidate for the Degree of

Doctor of Philosophy

Thesis: STRUCTURE AND DYNAMICS STUDIES OF MEMBRANE AND NON-MEMBRANE PROTEINS USING NMR AND MD SIMULATIONS

Major Field: General Physics

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in General Physics at Oklahoma State University, Stillwater, Oklahoma in December 2017.

Completed the requirements for the Master's of Science in Physics (English) at Tribhuvan University, Lalitpur, Nepal in 2009.

Experience:

Graduate Research/Teaching Assistant in Department of Physics, Oklahoma State University, August 2011 - December 2017