5-7-2016

Using NMR Spectroscopy to Investigate the Binding of DNA to the Globular Domain of Histone H1.0

Sri Ramya Tata

Using NMR spectroscopy to investigate the binding of DNA to the globular domain of histone H1.0

By

Sri Ramya Tata

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Chemistry
in the Department of Chemistry

Mississippi State, Mississippi

May 2016

Using NMR spectroscopy to investigate the binding of DNA to the globular domain of histone H1.0

By

Sri Ramya Tata

Approved:

_____

Nicholas C. Fitzkee
(Major Professor)

_____

Edwin A. Lewis
(Committee Member)

_____

Joseph P. Emerson
(Committee Member)

_____

Stephen C. Foster
(Graduate Coordinator)

_____

R. Gregory Dunaway
Dean
College of Arts & Sciences

Name: Sri Ramya Tata

Date of Degree: May 7, 2016

Institution: Mississippi State University

Major Field: Chemistry

Major Professors: Nicholas C. Fitzkee

Title of Study: Using NMR spectroscopy to investigate the binding of DNA to the globular domain of histone H1.0

Pages in Study 80

Candidate for Degree of Master of Science

Linker histones (H1) are a family of lysine-rich proteins that bind to or near the point at which DNA enters and exits the nucleosomal core. A number of studies have shown that H1 expression levels are altered in cancer and that variant-specific changes can be observed in different tumor cells. Although several crystal structures are published for core-histone/DNA complex (nucleosome), the location of linker histone and its interactions are poorly understood.

This study attempts to answer several questions regarding the interactions of histone H1 with double stranded partner DNA. Preliminary NMR assignments of this protein have been determined. We also investigated the structural changes in histone H1.0 globular domain induced by DNA binding. During the course of this project it was observed that subtle changes in pH could affect NMR spectral quality. We investigated the pH dependence of the protein stability by performing Circular Dichroism (CD) experiments.

DEDICATION

To My Parents,
Sri. Samba Siva Rao Tata and
Smt. Lakshmi Tata

To my Husband,
Dr. Krishna Kanth Ayyalasomayajula

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 General Introduction

In the eukaryotic cell, DNA exists as a nucleoprotein complex known as chromatin.[1] As far back as 1871, the properties of Chromatin have been intensely studied by scientists. *Chromatin* – a very readable book on structure and function of chromatin by K.E. Van Holde – investigates this macromolecule using a physical chemical approach by explaining different aspects of chromatin such as proteins of the chromatin (histones and non-histones proteins), isolation and separation of histones, and histone variants etc.[2] Structurally, chromatin is complex of macromolecules found in the cells, consisting of DNA, proteins, and RNA. Chromatin has many functions in the cell, several of which are (1) the packaging of DNA into a sufficiently small volume fit in the cell, (2) reinforcing the DNA macromolecule to allow mitosis, (3) preventing DNA damage, and (4) controlling gene expression and DNA replication. The primary protein components of chromatin are histones. DNA wraps around histone proteins forming nucleosomes; the "beads and string" structure.

The basic structural unit of chromatin, the nucleosome, was discovered by Roger Kornberg in 1974.[3] Nucleosome core particles represent the first level of chromatin organization and are composed of two copies of each of histones H2A, H2B, H3 and H4 assembled in an octameric core, approximately 63 Å in diameter. Tightly wrapped around

this core particle, 146-147 base pairs (bp) of DNA form a left-hand, super helical turn. The DNA wraps approximately 1.65 times around the core protein particle, forming a complex of around 100 Å across. All the core histones exist as dimers, and possess a similar fold structure: three alpha helices linked by loops.[4, 5] In addition to these core histones, linker histone H1 and its variants are also associated with the linker region of DNA.[6]

The linker histone H1 is thought to bind the nucleosome at the entry and exit sites of the DNA. A DNA linker of variable length connects nucleosomes at intervals of approximately 20-80 bp, forming an array of core particles at intervals of approximately 30 nm.[7-9] This linker DNA separates each nucleosome from either the succeeding or preceding nucleosomes. Those nucleosomal polymers that are assembled on a single DNA molecule are known as nucleosomal arrays.[10] These nucleosomal arrays undergo a second level of condensation by binding to the fifth histone protein thus forming a chromatin fiber approximately 30 nm in diameter. Self-association of these chromatin fibers makes up the final level of DNA condensation.[11-13] A schematic representation of a nucleosome resolved by X-ray crystallography with a resolution of 2.81 Å is shown in the figure 1.1.[14]

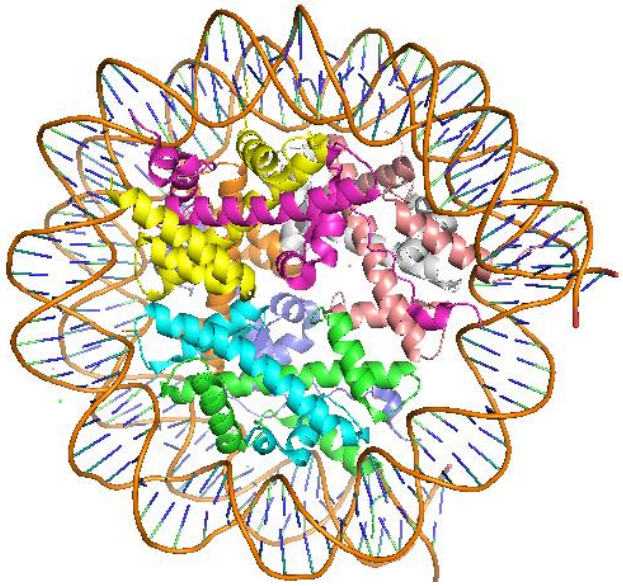

Figure 1.1 Schematic representation of X-ray crystal structure of a nucleosome (PDB ID: 3X1S)

1.2 Linker Histones

Histone H1 is a chromatin protein found in most eukaryotes. Linker histones are a family of lysine rich proteins that bind to or near the point at which DNA enters and exits the nucleosomal core in a roughly 1:1 stoichiometry of linker histone protein to nucleosome and organize an additional 20 base pairs of linker DNA, forming a chromatosome. [15] A Chromatosome model was shown in the Figure 1.2. Histone H1 is one of the five main histone protein families. The linker histone H1 and its various isoforms have been implicated in the formation of higher order structures and gene expression.[16] Histone H5 performs the same function as histone H1 and replaces the H1 in certain cells. In addition, the H1 histones have been shown to be involved in regulation of developmental genes.[17] A common feature of this protein family is a tripartite structure in which a globular domain of 80 amino acids is bridged by two less

structured N & C- terminal tails. Eukaryotic histone H1 constitute a family with many variants, which differ in their affinity in binding with chromatin. It is known that a single eukaryotic cell consists of several variants of histone H1 simultaneously, but whether these different isoforms are functionally important is unclear.[6, 18, 19]

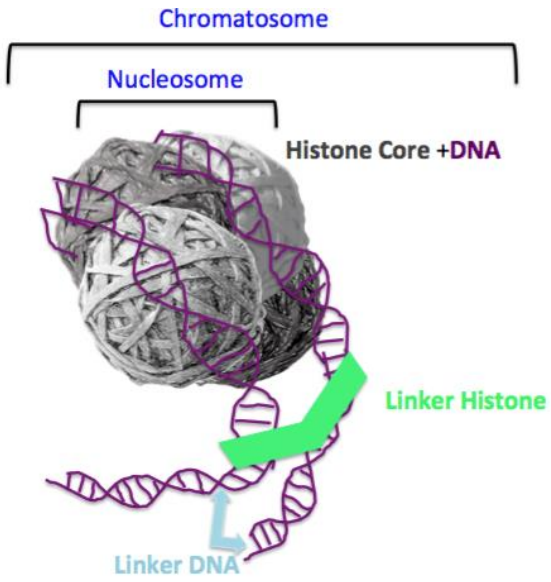

Figure 1.2 Chromatosome model

1.3 Linker Histone Variants

The linker histone family is comprised of many proteins associated with nucleosomes and linker DNA. They may differ in their amino acid composition, molecular mass, sequence, and physiochemical properties, but they share several structural features.[20] For instance, it is known from X-ray crystallography that $H1^0$ is similar in structure with H5.[21]

Eukaryotes have different variants in different tissues. Until now 11 different variants or subtypes have been identified in mammals. Interestingly, distributions of

genes that express these eleven linker histone subtypes are conserved between human and mouse. Initially they were named according to their chromatographic and electrophoretic properties- H1a, H1b, H1c, but eventually cloned and sequenced each of these sub types and introduced alternative nomenclature- H1.0 (or $H1^0$), H1.1-H1.5.[22] The linker histone subtypes are expressed in different cells, H1.1-H1.5, H1.0, and H1.X are expressed in somatic cells, remaining four are expressed in germ cells; H1t, H1T2, and HILS1 in sperm cells and H1oo in oocytes.[22, 23]

Classification of these subtypes is based either on cell-cycle dependent or replication dependent manner, explaining that the expression of these variants is strongly connected to specific phase of cell cycle. The linker histone H1.0 subtype is not involved directly in cell cycle; it is expressed constitutively, suggesting that its function is independent from replication.[24] Somatic subtype H1.X is involved in mitotic progression and is known to localize in nucleolus.[25] Mammalian linker histones variants have been shown to differentially affect gene expression, indicating different functional roles.[26] However, the functional aspects of linker histone subtypes are less understood.[27] To date, many scientists have investigated the interactions of core histone proteins with DNA. For instance, Angelov *et al*. in 2001 studied preferential interaction of core tail domains with linker DNA. This group investigated the primary DNA target for the majority of tails in mono nucleosomes. Experimental results from their work suggested that core histone tails prefer to bind linker DNA instead of intra nucleosome core DNA and they also observed the cross-linking of tail-DNA is not dependent on the ionic strength.[28] Chunyang *et al.* in 2003 investigated intra- and inter nucleosomal protein-DNA interactions of the core histone tail domains by building a di-

nucleosome model system. Their investigations showed that the N-terminal tails of H2A and H2B make inter nucleosomal histone-DNA interactions with in the di-nucleosome, whereas H3 & H4 don't. Furthermore this group observed that binding of di-nucleosome with linker histone increased the association of H3 & H4 with the linker DNA region.[29]

Machha *et al*. hypothesized that DNA binding induces unfolding in the H1 histone globular domain. They investigated the thermos-dynamical parameters for the binding of intact linker histone $H1^0$, globular domain ($H1^0$ –G) and C-terminal ($H1^0$ –C) domain to the DNA. Their observations have shown that the enthalpy change for formation of $H1^0$/DNA or $H1^0$ –C/DNA complex is very unfavorable and this complex formation is driven by large change in entropy.[30] CD results from their work indicated that features at 245 and 280nm were both attenuated upon formation of histone-DNA complex, suggesting loss of structure. Furthermore, the H1.4/DNA structure stabilization was characterized using DSC, which also supported their hypothesis.[31] In contrast, the exact location and function of linker histones with DNA are very less studied. In this study, linker histone subtype H1.0 globular domain has been used to determine the structural aspects using Nuclear Magnetic Resonance (NMR) Spectroscopy.

Linker histones, small basic proteins (~21K.Da) typically have a tripartite structure composed of central winged globular domain approximately 65 amino acid residues, short unstructured N-terminal domain approximately with 35 amino acid residues and highly basic C-terminal domain with approximately 110 amino acid residues. The unstructured N and C-terminal domains constitute half of the protein molecule.[32, 33] Brown, *et al.* studied interaction surface and positioning of the globular domain of linker histone H1.0 on the chromatosome.[34] In analogous study, *George*

Eric, *et al.* had shown that the globular domain of variant H1c interactions with nucleosome with a binding orientation that is distinct than that of H1.0.[35] This explains that differential orientation of globular domain could impart functional specificity to the linker histone isotopes. Thus it is very important to elucidate the structural properties of H1 isotypes, which can dictate the binding characteristics.[36]

1.4 Role of Linker Histones in Cancer

A number of studies have shown that H1 levels are altered in cancer and that variant-specific changes can be observed in different tumor cells. It is certain that DNA changes its confirmation on binding with histone proteins. While the three-dimensional structures of the histone H1 and H5 core domains have been established experimentally, the structure of the full protein in complex with DNA, including the N- and C-terminal tails, is not known.[37] Therefore, it is unclear whether these mutations are functionally important or simply the results of dysregulation of DNA repair mechanisms. It is necessary to understand the structure of full protein in complex with DNA to elucidate the importance of mutations in the protein structure and function as well. Covalent posttranslational modifications can alter protein function. They may affect the protein interactions, stability, and many cellular processes.[38] In particular, H1 Histones are known to involved in the both ovarian and breast cancers.

1.4.1 H1 Histones in Ovarian Cancer

Ovarian cancer has the highest mortality rate among gynecological malignancies and ranks as the 4th most common cancer in women.[39] Epigenetic regulations contribute significantly to ovarian cancer development and progression, and Histone H1,

with its large number of sites involved in posttranslational modification, is likely involved through this mechanism. Evidence shows that alteration in DNA methylation or histone modifications are responsible for silencing of tumor suppressor genes or up-regulating the cancer-promoting genes.[40]

Medrzycki *et al.* in 2012 investigated the potential connection between specific linker histone variants with ovarian cancer through expression profiling of various H1 subtypes and found that H1 subtypes displayed vastly different expression patterns in malignant ovarian cancer. Microarray & immunoblotting experiments suggested that, specifically $H1^0$, H1.1, H1.4 and H1.X are significantly reduced in expression, whereas H1.3 drastically increased expression in ovarian adenocarcinomas suggesting that linker histones can serve as potential biomarkers.[23]

1.4.2 H1 Histones in Breast Cancer

Experimental studies performed by Lu *et al.*, specifically investigated in posttranslational modifications of linker histone variants in human tissues. This group found that in both normal and in breast cancer tissue lysine methylation is the second most frequent posttranslational modification of linker histones, after phosphorylation. The study revealed that, in cultured breast and cervical cancer cells, lysine acetylation occurs frequently, whereas methylation occurred less often than in normal cells. Whereas, linker histones isolated from mouse tissues methylation appeared to be more frequent.[41]

Both Medrzycki *et al.* and Lu *et al.* have discussed lysine methylation as a posttranslational modification observed in cancer cells. Besides methylation, lysine acetylation and serine phosphorylation also influence the changes in function of histone.

H1 lysine acetylation *in vivo* was first demonstrated for H1.4K26. Vaquero A. *et al.* in 2004 characterized human SirT1, a HDAC (histone deacetylase) involved in initiating repressive chromatin, interacts with H1.4 and is also able to deacetylate H1.4K26 in vitro.[42] However, the enzyme that acetylates and its biological function are yet to be found.[38] Wisniewski *et al*. observed various methylations, acetylation, and phosphorylation sites of linker histone H1 variants using mass spectroscopy. [43]

Histone phosphorylation mainly includes serine and threonine residues, although Xiao *et al*. identified tyrosine phosphorylation. They demonstrated that WSTF (Williams-Beuren syndrome transcription factor), which has intrinsic tyrosine kinase activity, phosphorylates Tyr 142 of H2A.X, and is also important in DNA damage response.[44] Thus, a few class(s) of modifications or alterations (methylation, acetylation, and phosphorylation) occurred in nucleosome assembly have been implicated in cancers and other human diseases.

1.5 NMR Spectroscopic Studies of Linker Histones

Nuclear Magnetic Resonance (NMR) and X-Ray crystallography are the only two techniques capable of determining the structure of biological macromolecules like proteins and nuclei acids at atomic resolution. In addition, it is possible to study time-dependent phenomena with NMR, such as intramolecular dynamics in macromolecules, reaction kinetics, molecular recognition or protein folding. NMR spectroscopy is the only technique that allows the determination of three-dimensional structures of protein molecules in solution phase. In contrast to most other methods, NMR spectroscopy studies chemical properties by probing the chemical environment of individual nuclei.

Therefore, there is no exaggeration in accepting the truth that NMR spectroscopy is a potential tool in structure determination of biomolecules in the solution phase.

There have been several NMR-based studies of linker histones. In 1993, Cerf C. *et al.* studied a recombinant 75 amino acid polypeptide corresponding to the globular domain of chicken histone H1 (GH1) by ^{1}H homonuclear and ^{1}H -^{15}N heteronuclear 2D NMR spectroscopy. They determined the secondary structure of GH1 by sequence assignment and β-proton resonances and found that GH1 consists of three helical regions and a β-hairpin. It is shown that chicken histone H1 (GH1) structure is similar to H5, and both NMR and X-Ray structures of H5 are determined. Whereas, β-hairpin loop was only reported in X-Ray structure and was not observed in GH5 NMR structure.[21]

Ono K. *et al.* solved the structure of globular domain of linker histone of *Saccharomyces cerevisiae* by NMR and observed a winged helix-turn-helix motif. One of the domains is marginally stable, and suggested that poor spectral quality could be due to the dynamics of the protein. Gel mobility assay performed demonstrated that Hho1p preferentially binds to supercoiled DNA over linear DNA.[6] Ali T. *et al*. in 2004 investigated on globular domain (Hho1p- putative linker histone, unusually, has two domains) of Saccharomyces cerevisiae, which is a homolog of globular domain of histone H1 and determined the structure of the two domains GI and GII by NMR. Their investigations shown that both isolated GI and GII domains of yeast linker histone Hho1p adopt similar winged-helix structure in presence of high concentrations of anions (phosphate, sulfate, perchlorate) and both the domains might be functional. Rather, functions of these domains are not explained.[45]

Despite the numerous efforts during the years, structural and functional roles of linker histones remain ill defined. It is very well known that linker histone globular domain is in involved in stabilizing the entry and exit sites. In addition, it was hypothesized that N- and C- terminal domains of linker histones are highly unstructured *in vivo*, but not *in vitro*. In spite of numerous studies regarding the interplay between structure and function of histone H1, there are many questions that are unanswered.[46] Because of Histone H1 importance in cancers and involvement in epigenetic gene control, there is a pressing need to understand the interactions between histone H1 and DNA. Such an understanding will provide a better insight into the functional aspects of this important protein. Additionally, a detailed knowledge of H1 histone function could provide key clues to how chromatin is organized and restructured in cells.

1.6 Specific Research Objectives

A structural model for the linker histone/DNA complex (nucleosome) could potentially lead a deeper understanding of the functional and regulatory roles of the linker histone proteins. While NMR has been performed on several close homologs of Histone H1, there is reason to believe that differences in sequence may be functionally significant. Sequence differences can lead to differences in posttranslational modification and therefore differences in epigenetic control. Therefore, it is desirable to work with actual human protein sequences instead of homologous proteins.

This study attempts to answer several questions regarding the interactions of histone H1 with double stranded partner DNA. We have used NMR to investigate the globular domain of the H1.0 histone isoform form mice. As a first step, we have determined the preliminary NMR assignments of this protein, which are necessary for all

subsequent NMR work. In addition, we are also interested in studying the structural changes in histone H1.0 globular domain induced by DNA binding. Machha *et. al*. hypothesized that DNA binding induces unfolding in the H1 histone globular domain and his findings were explained in the above paragraphs. Therefore, our second step in this work was to observe this unfolding directly and determine its extent. Finally, during the course of this project it was observed that subtle changes in pH could affect NMR spectral quality. In the third step of this work, we investigated the pH dependence of the protein stability by performing Circular Dichrosim (CD) experiments.

1.7 References

1. Kasinsky, H.E., et al., Origin of H1 linker histones. FASEB J, 2001. 15(1): p. 34-42.

2. Langmore, J.P., Chromatin. Cell. 59(2): p. 243-244.

3. Kornberg, R.D., Chromatin structure: a repeating unit of histones and DNA. Science, 1974. 184(4139): p. 868-71.

4. Mariño-Ramírez, L., et al., Histone structure and nucleosome stability. Expert review of proteomics, 2005. 2(5): p. 719-729.

5. Woodcock, C.L., Chromatin architecture. Curr Opin Struct Biol, 2006. 16(2): p. 213-20

6. Ono, K., et al., The linker histone homolog Hho1p from Saccharomyces cerevisiae represents a winged helix-turn-helix fold as determined by NMR spectroscopy. Nucleic Acids Res, 2003. 31(24): p. 7199-207.

7. Finch, J.T. and A. Klug, Solenoidal model for superstructure in chromatin. Proceedings of the National Academy of Sciences, 1976. 73(6): p. 1897-1901.

8. Thoma, F. and T. Koller, Influence of histone H1 on chromatin structure. Cell. 12(1): p. 101-107.

9. Thoma, F., T. Koller, and A. Klug, Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. The Journal of Cell Biology, 1979. 83(2): p. 403-427.

10. Xu, D., et al., Characterization of Protein Structure and Function at Genome Scale with a Computational Prediction Pipeline, in Genetic Engineering, J. Setlow, Editor. 2003, Springer US. p. 269-293.

11. Hansen, J.C., et al., Intrinsic protein disorder, amino acid composition, and histone terminal domains. J Biol Chem, 2006. 281(4): p. 1853-6.

12. Lu, X. and J.C. Hansen, Identification of specific functional subdomains within the linker histone H10 C-terminal domain. J Biol Chem, 2004. 279(10): p. 8701-7.

13. Lu, X., et al., In vitro chromatin self-association and its relevance to genome architecture. Biochem Cell Biol, 2006. 84(4): p. 411-7.

14. Padavattan, S., et al., Structural and functional analyses of nucleosome complexes with mouse histone variants TH2a and TH2b, involved in reprogramming. Biochem Biophys Res Commun, 2015. 464(3): p. 929-35.

15. Ramaswamy, A., I. Bahar, and I. Ioshikhes, Structural dynamics of nucleosome core particle: comparison with nucleosomes containing histone variants. Proteins, 2005. 58(3): p. 683-96.

16. Vignali, M. and J.L. Workman, Location and function of linker histones. Nature structural biology, 1998. 5(12): p. 1025-1028.

17. Lu, X., et al., Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. Genes & Development, 2009. 23(4): p. 452-465.

18. Mariño-Ramírez, L., et al., The Histone Database: a comprehensive resource for histones and histone fold-containing proteins. Proteins, 2006. 62(4): p. 838-842.

19. Ramakrishnan, V., et al., Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature, 1993. 362(6417): p. 219-23.

20. Thiriet, C. and J.J. Hayes, Assembly into chromatin and subtype-specific transcriptional effects of exogenous linker histones directly introduced into a living Physarum cell. J Cell Sci, 2001. 114(Pt 5): p. 965-73.

21. Cerf, C., et al., Homo- and heteronuclear two-dimensional NMR studies of the globular domain of histone H1: sequential assignment and secondary structure. Biochemistry, 1993. 32(42): p. 11345-51.

22. Walter, L., et al., Chromosome mapping of rat histone genes H1fv, H1d, H1t, Th2a and Th2b. Cytogenet Cell Genet, 1996. 75(2-3): p. 136-9.

23. Medrzycki, M., et al., Profiling of linker histone variants in ovarian cancer. Frontiers in bioscience (Landmark edition), 2012. 17: p. 396-406.

24. Zlatanova, J. and D. Doenecke, Histone H1 zero: a major player in cell differentiation? FASEB J, 1994. 8(15): p. 1260-8.

25. Happel, N., E. Schulze, and D. Doenecke, Characterisation of human histone H1x. Biol Chem, 2005. 386(6): p. 541-51.

26. Alami, R., et al., Mammalian linker-histone subtypes differentially affect gene expression in vivo. Proc Natl Acad Sci U S A, 2003. 100(10): p. 5920-5.

27. Ausio, J., Histone variants--the structure behind the function. Brief Funct Genomic Proteomic, 2006. 5(3): p. 228-43.

28. Angelov, D., et al., Preferential interaction of the core histone tail domains with linker DNA. Proc Natl Acad Sci U S A, 2001. 98(12): p. 6599-604.

29. Zheng, C. and J.J. Hayes, Intra- and Inter-nucleosomal Protein-DNA Interactions of the Core Histone Tail Domains in a Model System. Journal of Biological Chemistry, 2003. 278(26): p. 24217-24224.

30. Machha, V.R., et al., Calorimetric studies of the interactions of linker histone H10 and its carboxyl (H10–C) and globular (H10–G) domains with calf-thymus DNA. Biophysical Chemistry, 2013. 184: p. 22-28.

31. Machha, V.R., et al., Exploring the energetics of histone H1.1 and H1.4 duplex DNA interactions. Biophys Chem, 2014. 185: p. 32-8.

32. Allan, J., et al., The structure of histone H1 and its location in chromatin. Nature, 1980. 288(5792): p. 675-9.

33. Subirana, J.A., Analysis of the charge distribution in the C-terminal region of histone H1 as related to its interaction with DNA. Biopolymers, 1990. 29(10-11): p. 1351-7.

34. Brown, D.T., T. Izard, and T. Misteli, Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. Nat Struct Mol Biol, 2006. 13(3): p. 250-5.

35. George, E.M., et al., Nucleosome Interaction Surface of Linker Histone H1c Is Distinct from That of H1(0). The Journal of Biological Chemistry, 2010. 285(27): p. 20891-20896.

36. Vyas, P. and D.T. Brown, N- and C-terminal Domains Determine Differential Nucleosomal Binding Geometry and Affinity of Linker Histone Isotypes H1(0) and H1c. The Journal of Biological Chemistry, 2012. 287(15): p. 11778-11787.

37. Scaffidi, P., Histone H1 alterations in cancer. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms.

38. Cohen, I., et al., Histone Modifiers in Cancer: Friends or Foes? Genes & Cancer, 2011. 2(6): p. 631-647.

39. Jemal, A., et al., Cancer statistics, 2008. CA Cancer J Clin, 2008. 58(2): p. 71-96.

40. Balch, C., et al., Minireview: Epigenetic Changes in Ovarian Cancer. Endocrinology, 2009. 150(9): p. 4003-4011.

41. Lu, A., et al., Mapping of Lysine Monomethylation of Linker Histones in Human Breast and Its Cancer. Journal of Proteome Research, 2009. 8(9): p. 4207-4215.

42. Vaquero, A., et al., Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. Mol Cell, 2004. 16(1): p. 93-105.

43. Wisniewski, J.R., et al., Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. Mol Cell Proteomics, 2007. 6(1): p.

44. Xiao, A., et al., WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity. Nature, 2009. 457(7225): p. 57-62.

45. Ali, T., et al., Two homologous domains of similar structure but different stability in the yeast linker histone, Hho1p. J Mol Biol, 2004. 338(1): p. 139-48.

46. Roque, A., I. Ponte, and P. Suau, Interplay between histone H1 structure and function. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms.

.

CHAPTER II

MATERIALS AND METHODS

This chapter describes the technical and experimental aspects of the multiple biophysical techniques used in this project, such as Nuclear Magnetic Resonance (NMR), Circular Dichroism (CD), and Mass Spectroscopy (MS); purification techniques like High-Performance Liquid Chromatography (HPLC), and Gel filtration Chromatography (GC) techniques that are commonly used throughout the study. A description of the instruments, and experimental methodologies are included in this chapter.

2.1 Methods

2.1.1 Nuclear Magnetic Resonance Spectroscopy

2.1.1.1 Introduction

For more than fifty years, nuclear magnetic resonance spectroscopy (NMR) has been a prominent technique in determining the structure of chemical compounds. NMR is a physical and chemical phenomenon in which nuclei in magnetic field absorb and re-emit electromagnetic radiation. Physically, the emitted radiation/energy has a specific frequency that depends on the strength of the magnetic field (B), and chemically, the radiation depends on the magnetic properties and chemical environment of the atoms (e.g. ^{15}N, ^{13}C, ^{31}P, etc.). A key feature of NMR is that the resonance frequency (ω) is directly proportional to the applied magnetic field (B_0).[1] On a basic level, the principles of NMR spectroscopy are similar to other forms of spectroscopy. Usually a photon of

light causes a transition from ground state to excited state. In visible spectroscopy, electrons absorb the energy, whereas in NMR spectroscopy the absorbed photon promotes a nuclear spin from ground state to excited state. NMR, therefore, results from the absorption of energy by a nucleus as it changes its spin in a static magnetic field. Only nuclei with nonzero spin numbers are detected by this technique; examples of such nuclei include: ^{13}C, ^{1}H, ^{2}H, ^{19}F, ^{15}N, and ^{31}P.

Protons (^{1}H) are the most studied among all other nuclei and their resonance spectrum is characteristic of the particular analyte being examined. Properties of analytes can be examined via through-bond spin-spin couplings (J) and the through-space nuclear Overhauser effect (NOE). Each spin in the nucleus gives rise to a nuclear magnetic resonance line. The exact resonance frequency depends on the chemical environment of each spin. For instance, the NMR spectrum of a protein will show many peaks with slightly different frequencies. These frequencies of individual nuclei are known as chemical shifts.[2] NMR spectroscopy of biological molecules (proteins), commonly called as protein NMR, is a well-established field of structural biology in which information about structure and dynamics of proteins and their complexes are studied.

2.1.1.2 Brief History

In past 40 years NMR spectroscopy has established as a major technical approach in the field of biology, chemistry and medicine. It became an important phenomenon in understanding the structure and dynamics of macromolecules (protein, nucleic acids, and their complexes). An improvement in NMR instrumentation and experimental techniques enabled many new applications in the field of science. Table below gives an overview of history in the field of NMR spectroscopy. [3]

Table 2.1 NMR History

Year	Scientist	Observations
1938	Isidor Rabi	Nuclear magnetic resonance was 1st described[4]
1946	Bloch, Purcell	Extended the work for liquids (detected nuclear induction signal in water)[5] and solids (detected NMR absorption in paraffin)[6]
1948	Bloembergen, Purcell & Pound	Studied relaxation effects in Nuclear Magnetic Resonance (NMR) absorption.[7]
1951	Parkard & Arnold	Introduced chemical shifts as a result of –OH hydrogen bonding –the solvent effect.[8]
1953	Over Hauser	Indicated the effects of electron spin populations to nuclear spin polarization.[9]
1957	Lauterbur & Holm	1st ^{13}C spectra was obtained.[10]
1962	Anderson & Freeman	Described theoretical basis of NOE (Nuclear Overhauser Effect) and experimentally verified.[11]
1966	Ernst & Anderson	Proposed Fourier transform math to process the Free Induction Decay (FID) from a pulse experiment.
1971	Jeener	First two-dimensional NMR experiment (COSY).[12]
1985	Kurt Wüthrich	Development of three-dimensional structures of biological macromolecules in solution.[13]
1985 to early 2000's	Kurt Wüthrich, Paul Lauterbur & Peter Mansfield	Advances in bio molecular NMR and Magnetic Resonance Imaging (MRI).[13, 14]
2000 to present	Ad Bax, Lewis E. Kay	Made many contributions for protein relaxation/ dynamics experiments[15, 16], Residual Dipolar Coupling (RDC's)[17] [18]& methyl TROSY.[19, 20] CEST experiments[21]

2.1.1.3 Limitations in Macromolecular NMR

Macromolecular NMR has become a fundamental approach in determining the dynamic and structural details of proteins, nucleic acids and their complexes. To date macromolecular NMR spectroscopy has emerged as one of the principle techniques of structural biology. It stands unique in its abilities by accurately measuring and probing many biophysical processes, including protein folding and binding. Despite this,

molecular size (and in particular slow rotational diffusion) significant limits the application of NMR to large molecular systems. Slow tumbling rates result in fast relaxation times, which can severely reduce the sensitivity of the experiment.[22, 23] Molecular weight limitations and their relative experimental techniques are listed in Table 2.2. Recent advances in macro molecular NMR such as methyl TROSY allows to measure the chemical shifts of higher molecular weight proteins as well.[19, 20]

Table 2.2 Molecular Weight Limitations for Chemical Shift Assignments

Mol. Wt.	**Technique**	**Observed spins**	**Dimensionality**
<10 KDa	Homonuclear	^{1}H	2D
10-15 KDa	^{15}N-homonuclear	^{1}H, ^{15}N	3D, 4D
15-30 KDa	Triple Resonance	^{1}H, ^{15}N, ^{13}C	3D, 4D
30-60 KDa	Triple Resonance/deuterated	^{1}H, ^{15}N, ^{13}C	3D, 4D
60-100 KDa	Triple Resonance/deuterated/TROSY	^{1}H, ^{15}N, ^{13}C	3D, 4D

2.1.1.4 Instrumentation

2.1.1.4.1 Overview

The NMR spectrometer consists of a superconducting magnet, probe, radio frequency (rf) transmitter, receiver, and processing computer. A stable static magnetic field produced by the magnet induces magnetization in the sample. The transmitter emits an oscillating electromagnetic field of various strengths (B_1) to interact with nuclei. The signal, or free induction decay (FID), is generated by the nuclei, detected by the probe coil, amplified by a preamplifier, and detected by a receiver. Finally, the computer processes the digitalized FID. [24]

2.1.1.4.2 Magnet

The magnet is an essential part of the spectrometer. A magnet is made of a cylindrical coil of superconducting wires. These wires carry an electric current, which creates the magnetic field, and the superconducting properties allow them to maintain the field for long periods of time (often years) without an applied external voltage. However, to maintain superconductivity, the solenoids are continually immersed in liquid helium in a large dewar. To further insulate the instrument, liquid nitrogen compartment (77K) surround the liquid helium dewar. This liquid nitrogen jacket is over the magnet and acts as a refrigerant to block/prevent the heat from reaching the liquid helium jacket. A reflective shield made of aluminum foil, acts as an additional layer of protection. Homogeneity and stability in the magnetic field is essential for the magnet, so all of the coil windings and insulating dewars must be manufactured to exacting standards. As field strength increases, sensitivity also increases, but the manufacturing process and requirements on homogeneity become more demanding. [25]

2.1.1.4.3 Transmitter

The transmitter is composed of frequency synthesizer, RF signal generator, a transmitter controller and an RF amplifier. It mainly functions in generating the pulse. The frequency synthesizer is used as a RF source, a stable source in generating the signal. The output of the synthesizer is gated to create a pulse of RF energy. The transmitter controller controls the length and timing of the pulse. The RF amplifier boots the signal to high power levels (approximately 100W or more) thus increasing the intensity in the magnetic field simultaneously shortening the 90° pulse length. [26]

2.1.1.4.4 Probe

A cylindrical metal tube, which is inserted center of the bore of magnet, is known as the probe. A small coil is assembled at the top, so that sample is dropped in the coil from the top of the magnet. This small coil is also useful to excite and detect the signal and acts as the key source for pulse magnetization. This forms a part of tuned circuit, consisting of a coil and several capacitors, which modulate the RF tuning (impedance) in the coil. Having appropriate tuning is very essential to maximize the power transfer between probe, transmitter and the receiver. Any change in the concentrations of the ions or solvent requires tuning and matching of the probe for high sensitivity. [27]

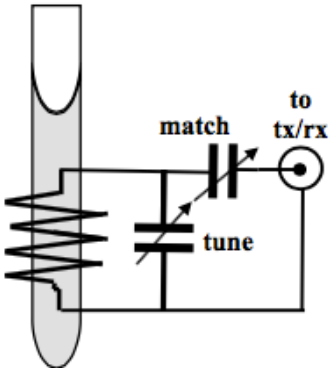

Figure 2.1 Diagrammatic representations of key parts of NMR probe [27]

2.1.1.4.5 Receiver

The receiver performs detection and amplification of the signal to a suitable level for digitalization. Measuring the amplitude/voltage of the signal, phase modulation and demodulating the NMR signal from carrier frequency can be defined as detection. The preamplifier located near the probe enhances the signal that comes from the probe in order to lower the loss of signal received before it reaches the receiver. These amplified

RF signals are tuned to small range near the carrier frequency for detection. RF frequency amplification process may also include the amplification of noise that is generated along with the signal, and is necessary to tune the signal outside the bandwidth before it reaches the receiver. [25,26]

2.1.1.5 Protein Structure Determination

NMR spectroscopy can be used for resonance-assignment approaches and structural determination methods for any reasonable size proteins over a wide range of time scales.[28] Assigning resonance lines to specific atoms of a protein in a spectrum is an essential step prior to the structure determination. Traditional assignment of a protein involves four steps. (1) Resonances associated with spin of one residue (often called spin-system) should be collected. These resonances may include both main chain (H_N, N_H, Hα, Cα, CO) and side chain atoms. (2) Amino acid types should be identified based on the spin-systems. (3) Adjacent residues in the protein are used for pairing the spin-systems and this pairwise association connecting as many spin-systems as possible leads to a segment of polypeptide chain. (4) Finally, adjacent polypeptide segments are matched to actual primary sequence of the protein itself. [1]

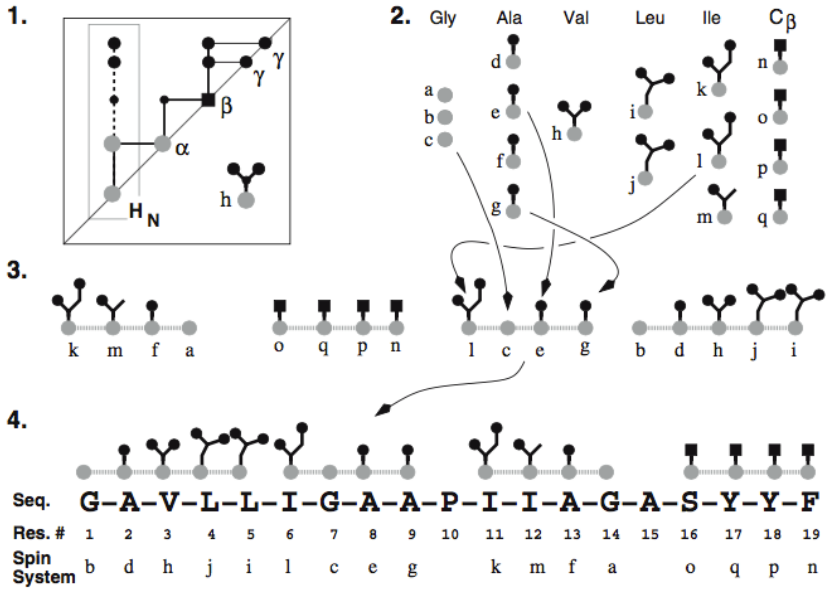

Figure 2.2 Overview of Resonance Assignment Strategy

Grey dots represent the main chain atoms, large black dots represent methyl groups, and the square black box represents β-carbons. The spin systems are given arbitrary labels, e.g. a-q.[1]

2.1.1.6 Assignment Strategies

Data sets that are recorded through multi-dimensional NMR for a known/unknown protein can vary greatly. The difference may be observed because of the nature and size of the protein or it can also be dependent on the level of expertise, data collection strategy or by instrument-specific limitations.[29-31] There are many automated programs invented for the backbone assignments. Some of the programs include, AUTOASSIGN, GARANT/DYANA etc.[32, 33] While automated assignment can work well for small, well-resolved proteins, human intervention is often needed to correct mistakes, and large/disordered proteins still require manual assignment. This manual approach is described below. All the experiments in this study were performed on a Bruker AVANCE III 600 MHz spectrometer, equipped with multinuclear bio molecular

cryogenic (QCI) probe for observation of ^{1}H while decoupling ^{13}C, ^{15}N, and/or ^{31}P. Spectra were processed and analyzed with nmrPipe tools, nmrDraw[34], and Sparky.

In the initial stages of an investigation by NMR spectroscopy, each resonance in the NMR spectrum must be associated with a specific nucleus in the molecule. Each resonance must be assigned to a spin in a particular amino acid residue in the protein sequence. An isotopically labeled protein is especially important in protein NMR, and in multi-dimensional NMR in particular. When a protein is labeled with ^{13}C and ^{15}N it is possible to record the triple resonance experiments that transfer magnetization over the peptide bond, and thus connect different spin systems through bonds. This is typically performed using the following experiments –**HSQC, HNCA, HN(CO)CA, HN(CA)CO, HNCO, HNCACB, CBCA(CO)NH. TOCSY** resolves an addition carbon dimension and is very helpful for side chain assignments. **NOESY** can provide the information about the additional peaks corresponding to atoms that are close in space but that do not belong to sequential residues.[35] The magnetization transfer pathways of the above-mentioned experiments are explained below in detail.

2.1.1.6.1 Heteronuclear Single Quantum Coherence Spectroscopy (HSQC)

The ^{15}N HSQC experiment is the most frequently recorded experiment in protein NMR. This is a two-dimensional NMR experiment that allows one to obtain the two-dimensional chemical shift spectrum between ^{1}H and X heteronuclei. Here the "X" can be either ^{15}N or ^{13}C commonly. This technique relays on proton-detection and is highly sensitive. This is a very standard experiment that shows all the H-N correlations. Initially magnetization is transferred from the hydrogen to the attached nitrogen through J-coupling. The chemical shift is evolved on the nitrogen and the magnetization is then

transferred on to the hydrogen for detection. This is the basic spectrum used for analysis, and it represents the "fingerprint" of the protein. From it one can assess whether other experiments are likely to work and is also very helpful to determine whether deuteration of the protein is necessary for further experimentation.

HSQC provides the correlation between the nitrogen and the amide proton, and each amide yields a peak in the spectrum. Signal from the H^N protons in the protein backbone are recorded. Since there is only one backbone H^N proton per amino acid, each HSQC signal represents one single amino acid. Not only the backbone H^N proton signals but also the signal from the NH_2 groups of the side chains of asparagine (Asn) and glutamine (Gln) and aromatic H^N protons of tryptophan (Trp) and histidine (His) can be seen in the spectrum. [36] Below figure shows the magnetization flow in HSQC experiment. [37]

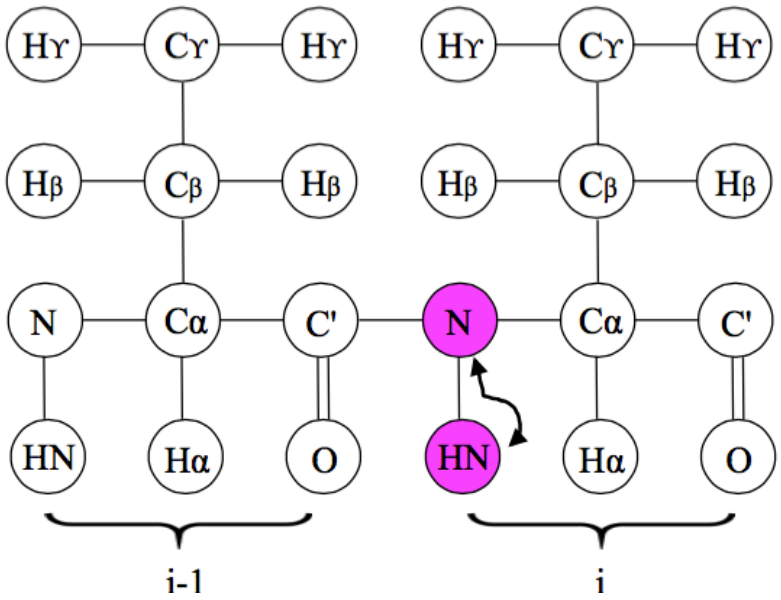

Figure 2.3 Schematic representation of magnetization in HSQC experiment

2.1.1.6.2 HNCA

This is an out-and back triple resonance experiment, where the magnetization starts on the H^N proton and is transferred on to ^{15}N and to $^{13}C\alpha$ then back to ^{15}N and then to $^{1}H^N$ for acquisition.[38] The chemical shift is evolved on $^{1}H^N$, $^{15}N^H$ and $^{13}C\alpha$ resulting in a three-dimensional spectrum. In this experiment, the amide nitrogen is coupled with both Cα of the current (own) residue (C_i) and Cα of the preceding residue (C_{i-1}), resulting in generation of peaks for both the Cα's in the spectrum. Thus the name is attained from its magnetization transfer pathway. The intensity of the peaks for the directly coupled Cα (C_i) residue is usually stronger than the peaks for the Cα of preceding residue (C_{i-1}), which makes assignment in most cases straightforward. This experiment typically requires double-labeled protein (^{15}N, ^{13}C) and is very useful in backbone assignment in combination with other triple resonance experiments.[39-41] Below figure shows the magnetization flow in HNCA experiment. [42]

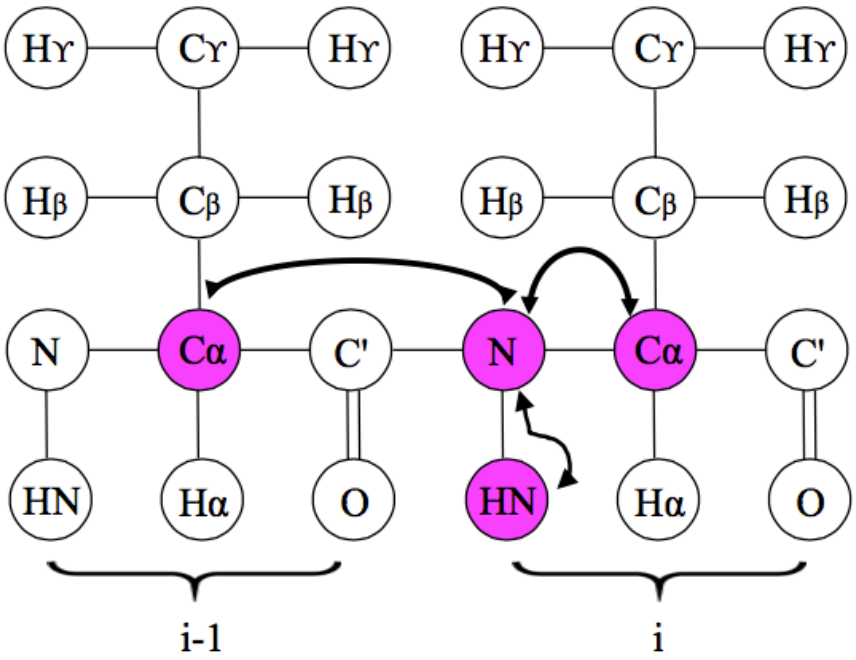

Figure 2.4 Schematic representation of magnetization flow in HNCA

2.1.1.6.3 HN(CO)CA

Similar to the HNCA experiment, HN(CO)CA also records a signal for $^{13}C\alpha$. In HNCA experiment signals for both C_i and C_{i-1} ($^{13}C\alpha$) are seen, whereas the HN(CO)CA generates signal only for C_{i-1} ($^{13}C\alpha$). Magnetization flows from ^{1}H to ^{15}N, and then to ^{13}CO, from there it transfers to $^{13}C\alpha$ and transfers back to ^{13}CO and ^{1}H for detection. This experiment is very sensitive for detecting the $^{13}C\alpha$ of the preceding residue C_{i-1}. Chemical shifts are evolved for ^{1}H, ^{15}N and $^{13}C\alpha$ (C_{i-1}). Along with the combination of other triple resonance experiments this is very helpful for backbone assignments, as it can reduce the uncertainty arising from crowded spin systems in the HNCA.[41, 43] Below figure shows the magnetization flow for HN(CO)CA experiment. [44]

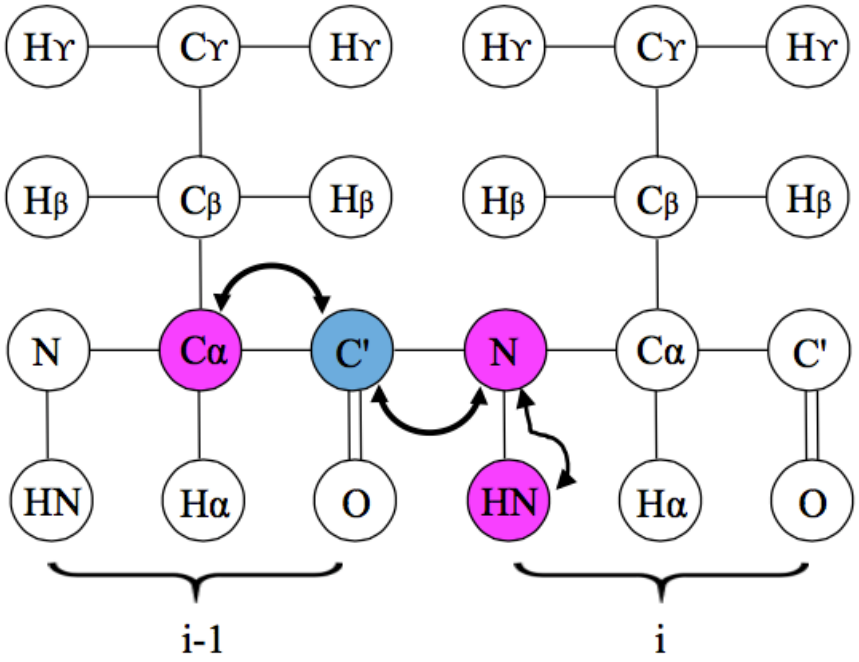

Figure 2.5 Schematic representation of magnetization flow in HN(CO)CA

No chemical shift is recorded on the blue C' nucleus

2.1.1.6.4 HNCO

The three-dimensional HNCO is the most sensitive triple resonance experiment, although it does not provide information on spin-system connectivity. This spectrum correlates the backbone ^{15}N-^{1}H pair of one residue with the carbonyl ^{13}CO resonance of the preceding residue. This experiment requires double-labeled protein as well (^{15}N, ^{13}C). This experiment involves amide (NH) proteins, and is recorded in H_2O. In this experiment the magnetization is passed from ^{1}H to ^{15}N and then selectively to carbonyl ^{13}C through the ^{15}N-^{13}CO J-coupling (-15 Hz). Magnetization is the passed back from ^{15}N to ^{1}H for detection. Thus the name is obtained from its magnetization pathway. This experiment can evolve the signal for all three nuclei. Addition to these signals from the three nuclei, side chain correlations of asparagine (Asn) and glutamine (Gln) side chains

are also visible. This experiment is mainly used for recording the signal for CO, which is needed for complete backbone assignments.[39, 41] Below figure shows the magnetization flow in HNCO experiment. [45]

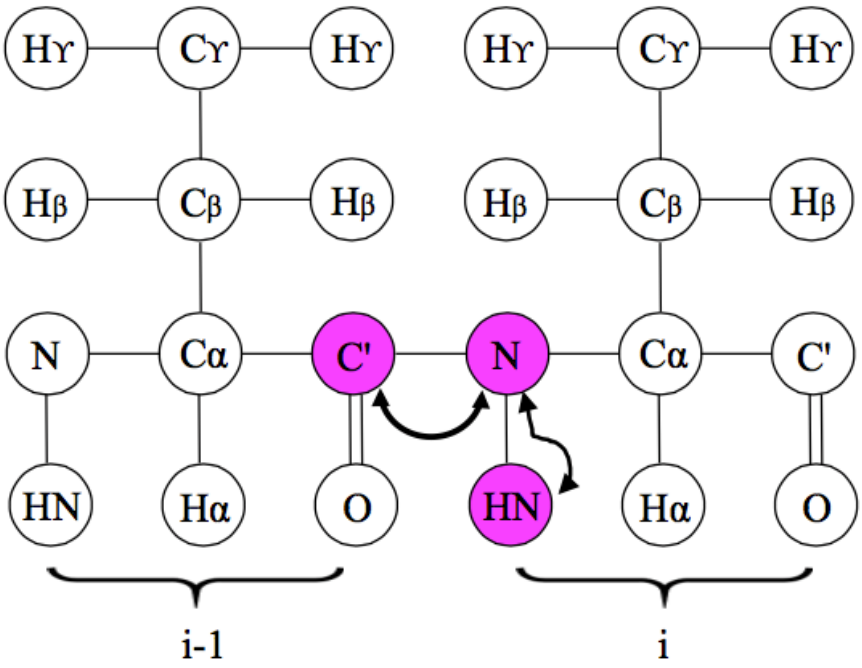

Figure 2.6 Schematic representation of magnetization in HNCO

2.1.1.6.5 HN(CA)CO

HN(CA)CO is also a three-dimensional experiment, which is designed to correlate intra- and inter-residue backbone connections. Here the magnetization is transferred from ^{1}H to ^{15}N then to N-Cα J-coupling to $^{13}Cα$, from which transfer occurs to ^{13}CO through $^{13}Cα$-^{13}CO J-coupling. From there it is transferred back in the same way for detection. It is to be noted that chemical shifts are evolved for only ^{1}H, ^{15}N and ^{13}CO, and not for $^{13}Cα$. Thus, the name is derived from its magnetization pathway. In this

experiment the amide nitrogen (NH) is coupled with both the $^{13}C\alpha$ of the current residue (C_i) and the preceding residue (C_{i-1}) $^{13}C\alpha$, thus transferring the magnetization to ^{13}CO of the current residue (C_i) and preceding residue (C_{i-1}) respectively. This transfer results in generating the signal for both $^{13}C'$ carbon where the signal from the own residue (C_i) has high intensity/stronger signal compared with the preceding residue (C_{i-1}). An overlap of spectra, obtained from both HNCO and HN(CA)CO makes it easy to differentiate between CO_i and CO_{i-1} for each NH group.[46] Because of the multiple transfer steps involved, this spectrum is the least sensitive of the spectra discussed so far and is seldom used in practice. Below figure shows the magnetization flow in HN(CA)CO experiment. [47]

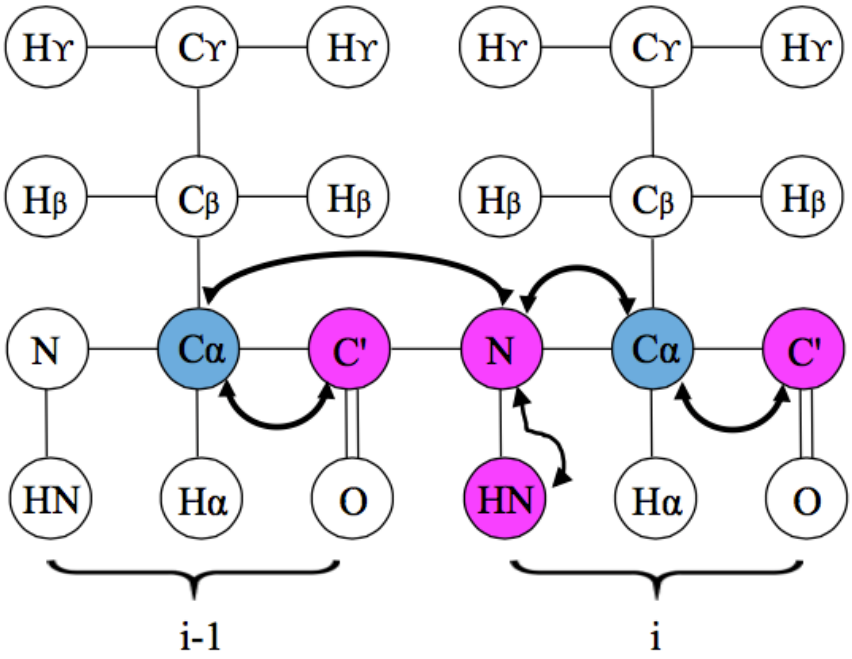

Figure 2.7 Schematic representation of magnetization flow in HN(CA)CO

2.1.1.6.6 CBCANH

HNCACB experiment is also named based on the magnetization pathway. The name itself explains that signals/chemical shift is evolved for ^{1}H, ^{15}N, $^{13}C\alpha$, and $^{13}C\beta$. This experiment correlates the proton and nitrogen for Cα and Cβ for both preceding residue C_{i-1} and the amide residue C_i. Magnetization starts on $^{1}H\alpha$ to $^{13}C\alpha$ and from $^{1}H\beta$ to $^{13}C\beta$ respectively, then from $^{13}C\beta$ to $^{13}C\alpha$, from here to $^{15}N^{H}$ to $^{1}H^{N}$ for detection. A chemical shift evolves for $^{13}C\alpha$ and $^{13}C\beta$ for both own residue C_i and preceding residue C_{i-1} resulting in four peaks foe single residue. This is less sensitive experiment, and might be less suitable for large proteins. In general, a 180° phase difference is observed for the signals obtained for $^{13}C\alpha$ and $^{13}C\beta$, meaning that $^{13}C\alpha$ peaks should have the opposite sign of $^{13}C\beta$ peaks.[41] Below, the figure shoes the magnetization pathway for the HNCACB experiment.[48]

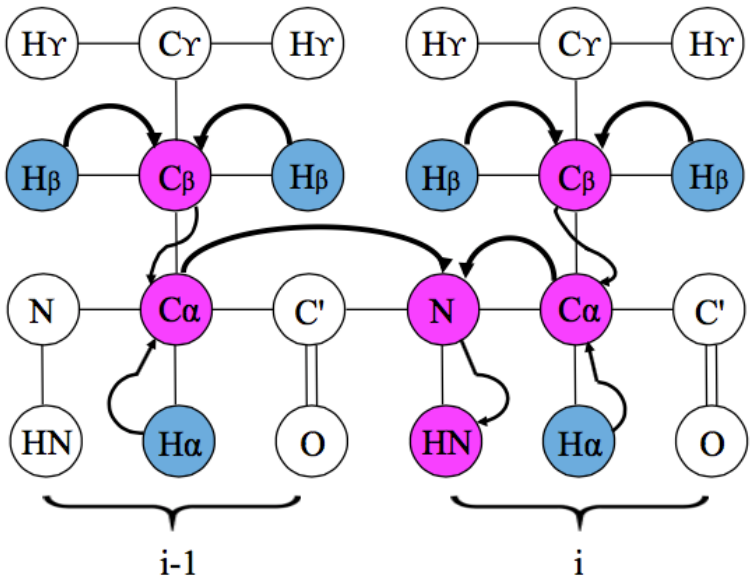

Figure 2.8 Schematic representation of magnetization flow in HN(CA)CB

2.1.1.6.7 CBCA(CO)NH

Similar to HN(CO)CA, CBCA(CO)NH also generates the signal for preceding residue (C_{i-1}). In the HN(CO)CA experiment, a signal for the $^{13}C\alpha$ of the preceding residue is visible, whereas in the CBCA(CO)NH experiment, both $^{13}C\alpha$ and $^{13}C\beta$ of the preceding/C_{i-1} residue is observed. The magnetization transfer pathway for the CBCA(CO)NH starts from $^{1}H\alpha$ and $^{1}H\beta$ to $^{13}C\alpha$ and $^{13}C\beta$ of the C_{i-1} residue respectively, then steps from $^{13}C\beta$ to $^{13}C\alpha$ and to $^{15}N^{H}$ and $^{1}H^{N}$ via ^{13}CO for detection. This may be less suitable for large proteins, but it can be useful in resolving overlap in the CBCANH. This experiment gives two peaks $^{13}C\alpha$ and $^{13}C\beta$ of the preceding residue, which is very helpful in conforming the preceding residue during backbone assignments of a protein.[41] The figure below shows the magnetization transfer pathway of the CBCA(CO)NH experiment.[49]

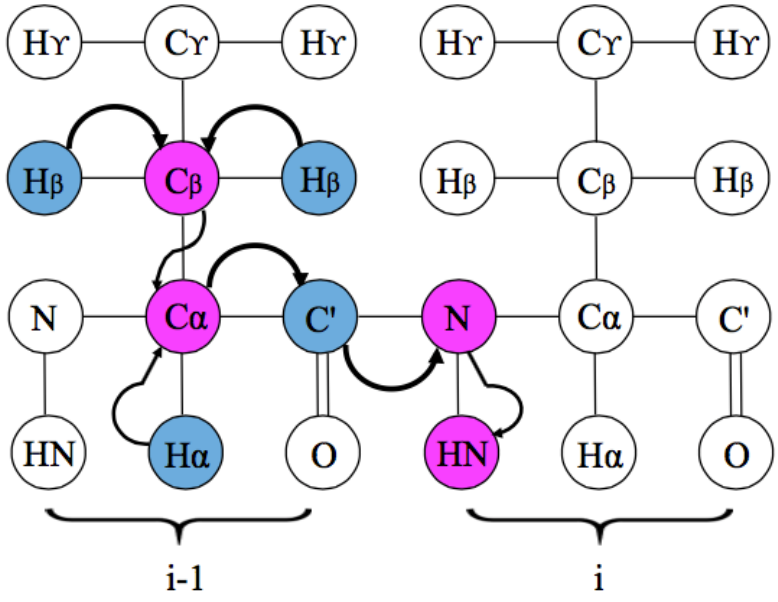

Figure 2.9 Schematic representation of magnetization flow in CBCA(CO)NH

2.1.2 Circular Dichroism

Circular Dichroism (CD) measures the difference in absorbance of right-and left handed circularly polarized light. This is a form of light absorption spectroscopy.[50] CD spectroscopy is widely used in studying large proteins and had many applications in bio molecular areas. Some of the applications include- characterizing secondary structure of a protein (folded/unfolded)[51], conformational changes/stability in a protein when subjected to pH changes, to determine the conformational changes in protein-protein, protein-DNA, DNA-DNA or DNA-ligand interactions, comparison of structural changes in mutants.[52] Schematic representation of incident beam, left and right polarized components, and an example CD spectrum is shown in the figure 2.10. An Olis DSM 20 spectropolarimeter was used to study pH changes in the histone globular domain.

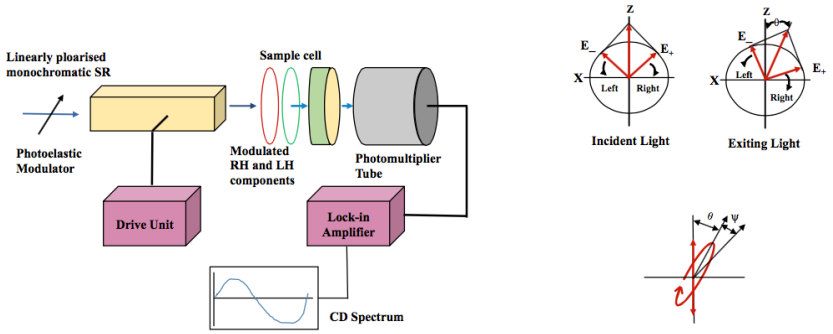

Figure 2.10 Block diagram of Circular Dichroism (CD) [53]

2.2 Materials

2.2.1 H1.0 Globular Domain

2.2.1.1 Expression

$H1^0$ globular domain is a 9KDa protein with 87 amino acid residues. Generally, our expression and purification protocol was based on methods described previously.[54, 55] $H1^0$ globular domain is expressed using a bacterial strain of *E. coli* (Rosetta2 (DE3) *pLysS*) transformed with a pET-11d expression vector. A single colony of cells is used to inoculate 1L YT medium with 100μg/ml ampicillin and allowed to grow at 37°C in an incubator until it reaches the mid-exponential phase (A_{600} was 0.6 to 0.8). At this time isopropyl thio-β-D-galactoside (IPTG) was added to a concentration of 1mM, and the cells were allowed to incubate for 4-6hrs, which resulted in maximal accumulation of all the protein.

2.2.1.2 Preparation

Cells are harvested after using a Beckman centrifuge at a speed of 18,000g for 25mins. The pellet is collected and stored in 50ml of lysis buffer (50mM Tris-pH8.0, 50mM EDTA, 50mM NaCl, 0.5% Triton X-100) at -80°C overnight (if needed). Cells are thawed on ice and 1mg/ml of lysozyme (i.e., 50mg/50ml of lysis buffer) is added and allowed to rock on ice for 1hr and centrifuged at maximum speed for 40min. Now the supernatant is collected and re-suspended in 5% $HClO_4$ ($1/10^{th}$ of lysate volume) which solubilized H1 protein and its proteolytic fragments. Then it is allowed to rock on ice, stirring for 1hr and centrifuged again at 18,000g for 30 min to remove the insoluble proteins and DNA. The supernatant was collected and re-suspended with $1/6^{th}$ volume of NH_4OH. The clarified lysate is concentrated down to 10ml using amicon concentrators at

a speed of 4,000 rpm for 30min to remove excess solvents and applied to a grace Vydac C18 reverse-phase column for purification.[54, 56]

2.2.1.3 Reverse-phased HPLC

Reverse-phased HPLC Vydac C18 column (CAT # 218TP510) was used in the purification process. Column is washed and equilibrated with buffer A (0.1% TFA (trifluroacetic acid) and 99.9% H_2O). The lysate was applied to the column and the protein is eluted using a gradient of 0% buffer B (95% acetonitrile and 0.1% TFA) for 5min, 0-40% B in 20min, and 40-70% B in 30min. A 5ml loop was used and 3ml of the protein was loaded on the column each time to reduce the risk of overloading the protein on the column.[56] The flow rate was 1ml/min and the protein is eluted at about 30min (figure 2.11). [54]

2.2.1.4 Ion-exchange chromatography

After HPLC, the sample is injected on a cation ion-exchange column attached to an AKTA Purifier FPLC system. The column is washed and equilibrated with buffer A (20mM NaPi & 50mM NaCl-pH7.0) and eluted using a gradient of 0-100% buffer B (20mM NaPi & 1M NaCl-pH7.0) over 60minutes, time may varies depending on the volume injected. The flow rate is 5ml/min and pressure limit is 0.5Mpa. Protein is at eluted about 340-380ml (figure 2.12). [56]

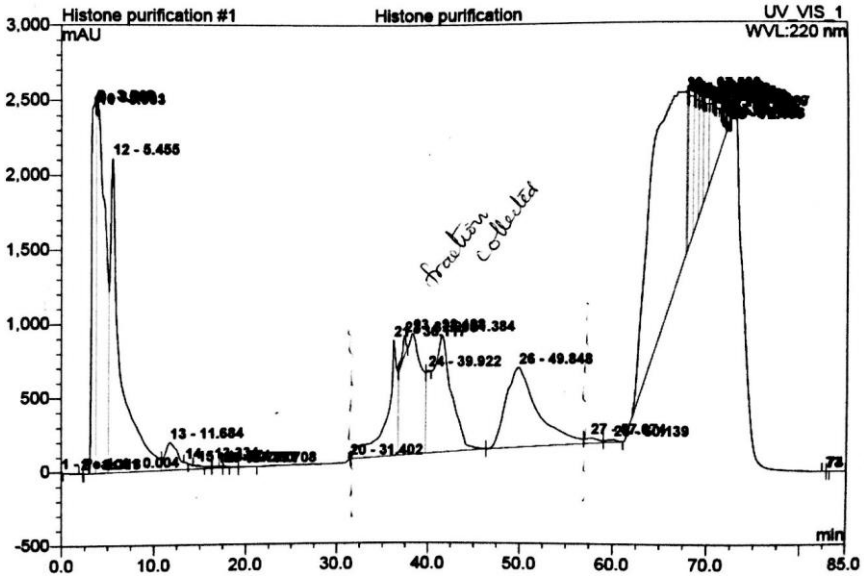

Figure 2.11 Chromatogram of H1.0g lysate purified on reverse-phased HPLC

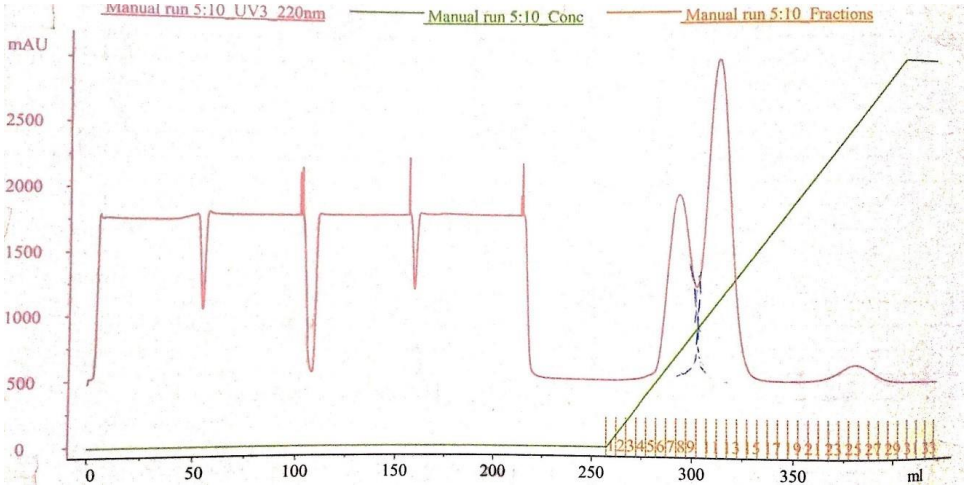

Figure 2.12 Chromatogram of Post HPLC sample

Sample is purified on ion-exchange column

2.2.1.5 Modified Preparation

In this work, several modifications are made to the original prep explained above. These changes are made in order to prevent the protein loss during several injections on

HPLC, and to remove the impurity observed in NMR spectrum for the original prep. Harvested cell pellet is stored at -80°C, this can be a potential stopping point before proceeding to the next step of purification. On the following day the pellet was re-suspended in 50ml of lysis buffer & 1mg/ml lysozyme and allowed to rock on the ice, stirring for 1hr. This step proceeds as the original prep. Then, rather than concentrating the lysate, it is dialyzed in 4L of water (stirring) in ion exchange buffer A (20mM NaPi, 50mM NaCl). This dialyzed lysate is concentrated to 10ml using amicon concentrators and loaded on to the ion exchange column for purification. Protein is eluted at 15% buffer B (20mM NaPi & 1M NaCl) instead of 100% buffer B gradient (figure 2.13). Fractions are pooled from 150-325ml and these fractions are injected on a Superdex 75 gel filtration column and eluted with 20mM NaPi & 50mM NaCl, pH 7.0(figure 2.14 shows the pooled fractions).

After purification described by Machha, et al.,[55] a set of low intensity but sharp peaks was identified in the NMR ^{15}N HSQC spectra. Mass spectroscopy (ESI LC-MS), SDS polyacrylamide gel electrophoresis (SDS-PAGE), and further NMR spectral characterization confirmed that the peaks corresponded to a protein impurity in the original sample preparation. The presence of this impurity was confirmed by discussion with the original authors. In our research, we removed the impurity by purifying the protein further using the modified prep described above. This purified sample was again verified with ESI LC-MS to compare with the old sample and we found that the impurity was removed after the ion exchange purification step (Figure 2.15).

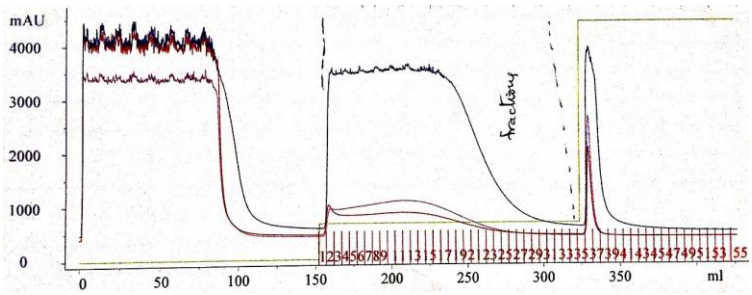

Figure 2.13 Chromatogram of H1.0g lysate purified on ion exchange column

Purified with 15% gradient of buffer B. Pooled fractions are marked.

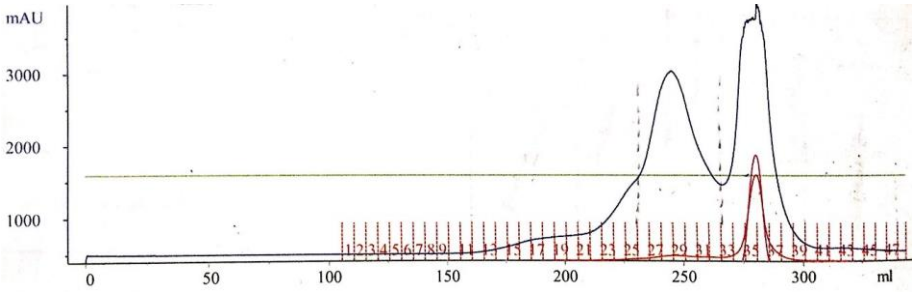

Figure 2.14 Gel filtration chromatogram of H1.0g

Pooled fractions are marked

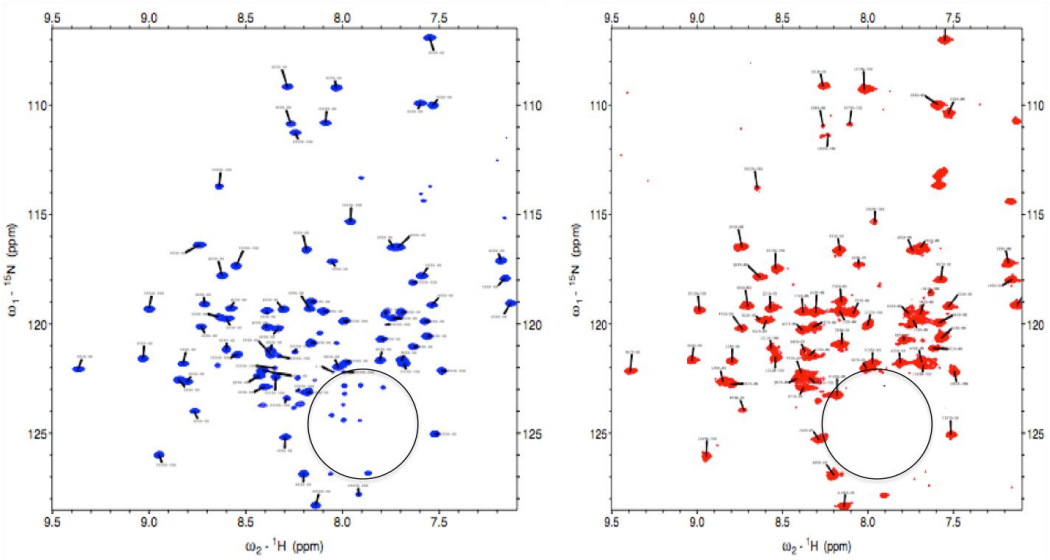

Figure 2.15 NMR spectra for purified H1.0g

original prep (A) and modified prep (B) removed impurities are circled

2.2.2 NMR Sample Preparation

Protein NMR is performed on aqueous samples of purified protein. A typical sample consists of 300-600 µl with a protein concentration of 0.1-3mM. For this study, we used 500 µl of 0.1mM sample in various buffers, described in detail in future chapters. A set of TROSY-HSQC experiments were performed to determine the optimum conditions such as temperature and pH for collecting the data sets for assignments. By comparing these NMR spectra, it is observed that 140µM protein at pH 6.8 and temperature 293K was optimal for acquisition. HSQC spectra collected at different pH & temperature conditions are shown in Chapter III.

2.3 References

1. Rule, G.S., and Hitchens, T. Kelvin. , *Fundamentals of protein NMR spectroscopy*. 2006, New York: Springer.

2. Van, H.K.E., W.C Johnson, and Pui S. Ho, *Principles of Physical Biochemistry*. 2006, Upper Saddle River, NJ: Pearson/Prentice Hall . print.

3. Becker, E.D., *A BRIEF HISTORY OF NUCLEAR MAGNETIC RESONANCE.* Analytical Chemistry, 1993. **65**(6): p. 295A-302A.

4. Rabi, I.I., et al., *A New Method of Measuring Nuclear Magnetic Moment.* Physical Review, 1938. **53**(4): p. 318-318.

5. Bloch, F., W.W. Hansen, and M. Packard, *The Nuclear Induction Experiment.* Physical Review, 1946. **70**(7-8): p. 474-485.

6. Purcell, E.M., H.C. Torrey, and R.V. Pound, *Resonance Absorption by Nuclear Magnetic Moments in a Solid.* Physical Review, 1946. **69**(1-2): p. 37-38.

7. Bloembergen, N., E.M. Purcell, and R.V. Pound, *Relaxation Effects in Nuclear Magnetic Resonance Absorption.* Physical Review, 1948. **73**(7): p. 679-712.

8. Arnold, J.T., S.S. Dharmatti, and M.E. Packard, *Chemical Effects on Nuclear Induction Signals from Organic Compounds.* The Journal of Chemical Physics, 1951. **19**(4): p. 507-507.

9. Overhauser, A.W., *Polarization of Nuclei in Metals.* Physical Review, 1953. **92**(2): p. 411-415.

10. Stothers, J.B., *Carbon-13 nuclear magnetic resonance spectroscopy.* Quarterly Reviews, Chemical Society, 1965. **19**(2): p. 144-167.

11. Anderson, W.A. and R. Freeman, *Influence of a Second Radiofrequency Field on High-Resolution Nuclear Magnetic Resonance Spectra.* The Journal of Chemical Physics, 1962. **37**(1): p. 85-103.

12. Jeener, J., et al., *Investigation of exchange processes by two-dimensional NMR spectroscopy.* The Journal of Chemical Physics, 1979. **71**(11): p. 4546-4553.

13. Palmer, A.G. and D.J. Patel, *Kurt Wüthrich and NMR of Biological Macromolecules.* Structure. **10**(12): p. 1603-1604.

14. Mansfield, P. and P.K. Grannell, *NMR 'diffraction' in solids?* Journal of Physics C: Solid State Physics, 1973. **6**(22): p. L422.

15. Kay, L.E., D.A. Torchia, and A. Bax, *Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease.* Biochemistry, 1989. **28**(23): p. 8972-9.

16. Tjandra, N. and A. Bax, *Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium.* Science, 1997. **278**(5340): p. 1111-1114.

17. Tjandra, N., S. Grzesiek, and A. Bax, *Magnetic Field Dependence of Nitrogen−Proton J Splittings in 15N-Enriched Human Ubiquitin Resulting from Relaxation Interference and Residual Dipolar Coupling.* Journal of the American Chemical Society, 1996. **118**(26): p. 6264-6272.

18. Maltsev, A.S., et al., *Improved Cross Validation of a Static Ubiquitin Structure Derived from High Precision Residual Dipolar Couplings Measured in a Drug-Based Liquid Crystalline Phase.* Journal of the American Chemical Society, 2014. **136**(10): p. 3752-3755.

19. Ollerenshaw, J.E., V. Tugarinov, and L.E. Kay, *Methyl TROSY: explanation and experimental verification.* Magnetic Resonance in Chemistry, 2003. **41**(10): p. 843-852.

20. Rosenzweig, R. and L.E. Kay, *Bringing dynamic molecular machines into focus by methyl-TROSY NMR.* Annu Rev Biochem, 2014. **83**: p. 291-315.

21. Long, D., A. Sekhar, and L.E. Kay, *Triple resonance-based 13Cα and 13Cβ CEST experiments for studies of ms timescale dynamics in proteins.* Journal of Biomolecular NMR, 2014. **60**(4): p. 203-208.

22. Yu, H., *Extending the size limit of protein nuclear magnetic resonance.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(2): p. 332-334.

23. Kay, L.E., *Protein dynamics from NMR.* Nat Struct Biol, 1998. **5 Suppl**: p. 513-7.

24. Cavanagh, J., et al., *CHAPTER 3 - EXPERIMENTAL ASPECTS OF NMR SPECTROSCOPY*, in *Protein NMR Spectroscopy (Second Edition)*, J. Cavanagh, et al., Editors. 2007, Academic Press: Burlington. p. 114-270.

25. Teng, Q., *Structural Biology- Practical NMR Applications*. 2nd edition ed. 2013.

26. Teng, Q., *Instrumentation*, in *Structural Biology*. 2013, Springer US. p. 65-101.

27. keeler, J., *Understanding NMR Spectroscopy, 2nd edition*. 2002. 526.

28. Wand, A.J. and S.W. Englander, *Protein complexes studied by NMR spectroscopy.* Current opinion in biotechnology, 1996. **7**(4): p. 403-408.

29. Billeter, M., G. Wagner, and K. Wüthrich, *Solution NMR structure determination of proteins revisited.* Journal of biomolecular NMR, 2008. **42**(3): p. 155-158.

30. Montelione, G.T., et al., *Protein NMR spectroscopy in structural genomics.* Nat Struct Biol, 2000. **7 Suppl**: p. 982-5.

31. Yee, A., et al., *An NMR approach to structural proteomics.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(4): p. 1825-1830.

32. Zimmerman, D.E., et al., *Automated analysis of protein NMR assignments using methods from artificial intelligence1.* Journal of Molecular Biology, 1997. **269**(4): p. 592-610.

33. Moseley, H.N.B. and G.T. Montelione, *Automated analysis of NMR assignments and structures for proteins.* Current Opinion in Structural Biology, 1999. **9**(5): p. 635-642.

34. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.

35. Englander, S.W. and A.J. Wand, *Main-chain-directed strategy for the assignment of proton NMR spectra of proteins.* Biochemistry, 1987. **26**(19): p. 5953-5958.

36. John Cavanagh, W.J.F., Arthur G. Palmer III, Mark Rance and Nicholas J. Skelton *Protein NMR Spectroscopy*. November 2006, Academic Press.

37. *HSQC*. Available from: http://www.protein-nmr.org.uk/pictures/experiment_types/hsqc.png.

38. Ikura, M., L.E. Kay, and A. Bax, *A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin.* Biochemistry, 1990. **29**(19): p. 4659-67.

39. Kay, L.E., et al., *Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins.* Journal of Magnetic Resonance (1969), 1990. **89**(3): p. 496-514.

40. Farmer, B.T., 2nd, et al., *A refocused and optimized HNCA: increased sensitivity and resolution in large macromolecules.* J Biomol NMR, 1992. **2**(2): p. 195-202.

41. Grzesiek, S. and A. Bax, *Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein.* Journal of Magnetic Resonance (1969), 1992. **96**(2): p. 432-440.

42. *HNCA*. Available from: http://www.protein-nmr.org.uk/pictures/experiment_types/hnca.png.

43. Bax, A. and M. Ikura, *An efficient 3D NMR technique for correlating the proton and15N backbone amide resonances with the α-carbon of the preceding residue in uniformly15N/13C enriched proteins.* Journal of Biomolecular NMR. **1**(1): p. 99-104.

44. *HN(CO)CA*. Available from: http://www.protein-nmr.org.uk/pictures/experiment_types/hncoca.png.

45. *HNCO*. Available from: http://www.protein-nmr.org.uk/pictures/experiment_types/hnco.png.

46. Clubb, R.T., V. Thanabal, and G. Wagner, *A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C' chemical shifts in 15N · 13C-labelled proteins.* Journal of Magnetic Resonance (1969), 1992. **97**(1): p. 213-217.

47. *HN(CA)CO*. Available from: http://www.protein-nmr.org.uk/pictures/experiment_types/hncaco.png.

48. *HNCACB*. Available from: http://www.protein-nmr.org.uk/pictures/experiment_types/hncacb.png.

49. *CBCA(CO)NH.*

50. Martin, S.R. and M.J. Schilstra, *Circular dichroism and its application to the study of biomolecules.* Methods Cell Biol, 2008. **84**: p. 263-93.

51. Van Holde, K.E., W. Curtis Johnson, and P. Shing Ho. , *"Chapter 10 Linear and Circular Dichroism." Principles of Physical Biochemistry.* 2006, New Jersey: Pearson Education.

52. Greenfield, N.J., *Applications of circular dichroism in protein and peptide analysis.* TrAC Trends in Analytical Chemistry, 1999. **18**(4): p. 236-244.

53. Kelly, S.M., T.J. Jess, and N.C. Price, *How to study proteins by circular dichroism.* Biochim Biophys Acta, 2005. **1751**(2): p. 119-39.

54. Wellman, S.E., et al., *Purification of mouse H1 histones expressed in Escherichia coli.* Biotechnol Appl Biochem, 1997. **26 (Pt 2)**: p. 117-23.

55. Machha, V.R., et al., *Calorimetric studies of the interactions of linker histone H10 and its carboxyl (H10–C) and globular (H10–G) domains with calf-thymus DNA.* Biophysical Chemistry, 2013. **184**: p. 22-28.

56. Machha, V.R., Thermodynamic characterization of linker histone binding interactions with ds-DNA, in Department of chemistry2014, Mississippi State University. p. 95.

CHAPTER III

RESULTS

3.1 Introduction

Several close homologs of Histone H1 have been studied using NMR, and there is reason to believe that differences in sequence may be functionally significant[1, 2]. Notable variations can lead to differences in posttranslational modification and therefore differences in epigenetic control[3-5]. Moreover, the sequences themselves may result in differing thermodynamic stability and conformational dynamics. Therefore, it is desirable to work with actual human protein sequences instead of homologous proteins.

This study attempts to answer a few questions regarding the interactions of histone H1 with double stranded partner DNA. We have used NMR to investigate the globular domain of the H1.0 histone isoform form mice. As a first step, we have determined the preliminary NMR assignments of this protein, which are necessary for all subsequent NMR work. In addition, we have also performed several NMR experiments to study the structural changes in histone H1.0 globular domain induced by DNA binding.

Machha *et al*. hypothesized that DNA binding induces unfolding in the H1 histone globular domain, and these findings were explained in chapter I. Therefore, our second step in this work was to observe this unfolding directly and determine its extent. While no large perturbations in the H1.0 chemical shifts were observed in the presence of DNA, a decrease in intensity for several residues was identified.

Finally, during the course of this project it was observed that subtle changes in pH could affect the protein structure, as monitored by NMR spectral quality. Changes that are observed in spectra resulting from pH changes are clearly explained below. In the third part of this work, we investigated the pH dependence of the protein stability by performing Circular Dichroism (CD) experiments.

The experiments described here are explained in detail in chapter II (Materials and Methods). This chapter focuses on the details of assignments, which include residue identification strategy, number of residues identified, sequence details, example strip plot, HSQC assigned spectrum, HSQC spectra for histone/DNA complex, and CD studies for pH dependence. All the NMR experiments in this study were performed on a Bruker AVANCE III 600 MHz spectrometer, equipped with multinuclear bio molecular cryogenic (QCI) probe for observation of ^{1}H while decoupling ^{13}C, ^{15}N, and/or ^{31}P. Spectra were processed and analyzed using NMR Pipe[6] and SPARKY. All the CD experiments were performed using an Olis DSM 20 spectropolarimeter (Bogart, GA, USA).

3.2 Sequence Information

In this project, we are working with globular domain of histone H1.0 from *Mus musculus*. It is a 9 KDa protein with 7 negatively charged (Asp + Glu) and 16 positively charged (Arg + Lys) amino acids. H1.0 globular domain consists of 87 amino acids with a theoretical pI of 10.12 and the sequence used in this work was:

MATDHPKYSDMIVAAIQAEKNRAGSSRQSIQKYIKSHYKVGENADSQIKL SIKRLVTTGVLKQTKGVGASGSFRLAKGDEPKRSVAF

3.3 Back Bone Sequential Assignments

Peaks of certain chemical shifts and the intensities are the experimental observables generated by NMR. It is necessary to assign/determine the chemical shifts of all carbon, nitrogen and proton atom in order to link the 2D- & 3D experimental data and to perform quantitative analysis on subsequent experiments.[7] Once the NMR spectra for 2D- and 3D experiments are recorded, spectra were processed using the UNIX-based software NMR Pipe[6] In the next step, peaks are assigned using SPARKY, a tool for visualization, analysis, and assignment of NMR spectra.

Assignments for a protein require data from both 2D & 3D experiments. In a two-dimensional ^{15}N-^{1}H HSQC spectrum, all the peaks correspond to the chemical shifts of nitrogen and a proton for every NH (or NH_2) group present in the protein. Thus, the HSQC spectrum represents a fingerprint for the protein, uniquely identifying the conformation based on the chemical environment of each amide group. The HNCA spectrum correlates these NH chemical shifts with the Cα chemical shifts of the current residue (*i*) and the previous residue (*i-1*) in a three-dimensional spectrum. Similarly, as described in chapter II, the HN(CO)CA spectrum correlates only the previous (*i-1*) residue with the associated NH chemical shifts, the HNCACB correlates both Cα and Cβ chemical shifts, and the HNCO correlates NH with the previous residue's (*i-1*) carbonyl carbon chemical shift. Thus, the combination of these spectra can be used to obtain the complete sequential backbone assignment for a protein like H1.0 (excluding proline residues, which lack an NH group).

In our work we have followed the traditional strategies for the assignment of H1.0g histone. First published in 1999, SPARKY remains one of the most widely used

visualization tools for the analysis of NMR data (https://www.cgl.ucsf.edu/home/sparky/). SPARKY software streamlines peak picking and spin system identification, which in generating rapid assignment of small to mid-sized (1-20 KDa) proteins. Important features of sparky that facilitate assignment include:

1. Automated and semi-automated peak picking. Peak picking streamlines the search for new spin systems. Peaks in Sparky can be picked manually with a mouse-click, or automatically using a region-based "click and drag" approach.
2. Facile generation of complex strip plots. SPARKY enables the rapid generation of strip plots, allowing one to quickly confirm whether spin systems can be linked. Cα and Cβ chemical shifts can be quickly compared to neighboring residues, and incorrect strips can be adjusted quickly. Particular strip coordinates are selected based on the chemical shift values of individual peak in a peak list obtained by peak picking procedure.

Spectrum synchronization. SPARKY also allows synchronizing the spectra such that selection of one peak in HSQC spectrum allows to see the corresponding peaks in HNCA, HN(CO)CA, HNCACB, and HNCO.

The HNCO is a most sensitive experiment and also helps in reducing the ambiguities in crowded regions of the HSQC spectrum. Few standard/characteristic peak locations in a HSQC spectrum help in identifying/assuming the residue type as a first level of spectral understanding, some of which includes, - peaks at the right bottom corner of the spectrum generally give strong peaks and typically represent C-terminal residues. Asn and Gln side chain NH_2 group peaks are generally seen at the right top corner (up field in N and H chemical shift) of the spectrum and logically result in a pair

of peaks, each having the same nitrogen chemical shift but with different hydrogen chemical shifts. Amino acids like glycine, alanine, threonine, serine, have their unique chemical shifts and can be identified easily in HNCACB spectrum. Alanine is the only amino acid with a methyl Cβ chemical shift, typically seen at 19ppm. Glycine doesn't have Cβ chemical shift, therefore a single peak is seen in the range of ~45ppm. Threonine and Serine have characteristic Cβ chemical shifts at 64 and 70ppm respectively. The remaining amino acid Cβ chemical shifts are generally observed in the range of ~29-41ppm. List of average chemical shift of individual spin-systems are taken from protein NMR spectroscopy by Cavanagh.[7]

In this assignment, the primary amino acid sequence of H1.0 globular domain is compared with the strip plots from each spin system. Assignment proceeds by comparing strips to the typical chemical shift values given above, then linking those spin systems based on residue type and $i \rightarrow i\text{-}1$ chemical shifts. While the likelihood of an individual residue's identity may be ambiguous (e.g. a Cβ chemical shift of 38 ppm could correspond to several different residues), as spin systems are linked, the identities become less ambiguous until a final assignment is attained (typically 3-4 spin systems is the minimum require to unambiguously assign a series of links to a sequence of residues). For example, the long connected peptide chain built by using the assignment strategy outlined above matches with G-E-N-A-D-? -? -? -? -L-S-I. This sequence is found only one time in the protein, so the peptide is placed in the corresponding stretch of the protein. The researcher must be confident about the links between D and L (i.e., Cα(i) → Cα(i-1)), but if this is the case, those residues can be assigned as S-Q-I-K even if the Cβ chemical shifts for those systems are ambiguous. Similarly, after building connections for

all most all the residues, they are compared with the protein sequence and thus the assignments are finalized. So far, 90% of the residues have been assigned using the strategy outlined above. Several of the assignments were initially ambiguous and have been identified with the help of other three-dimensional spectra (e.g. TOCSY). Missing assignments are presumably caused either by amide proton solvent exchange or conformational change (signal broadening). The spectra below illustrate representative HSQC spectra for unassigned/assigned H1.0 protein at pH 6.8.

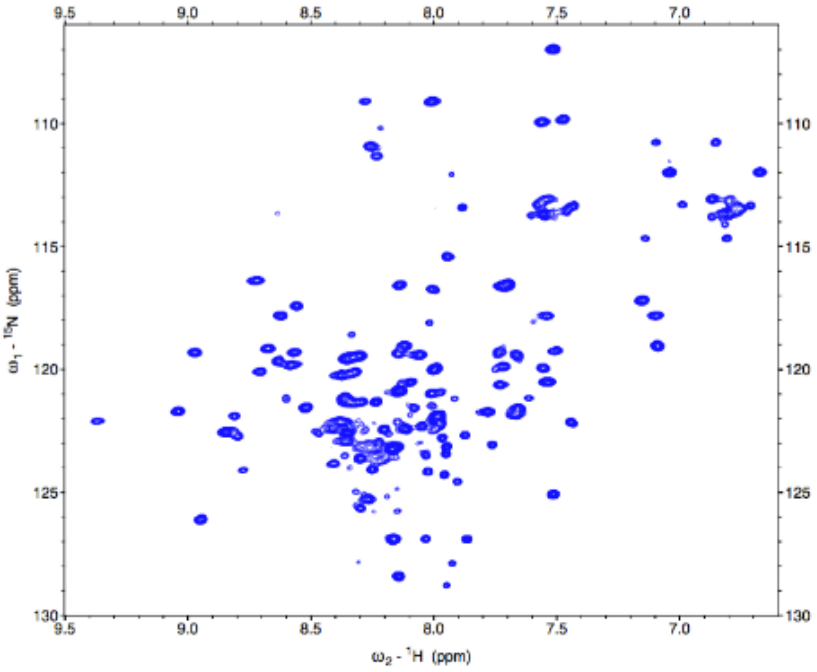

Figure 3.1 Unassigned HSQC Spectrum

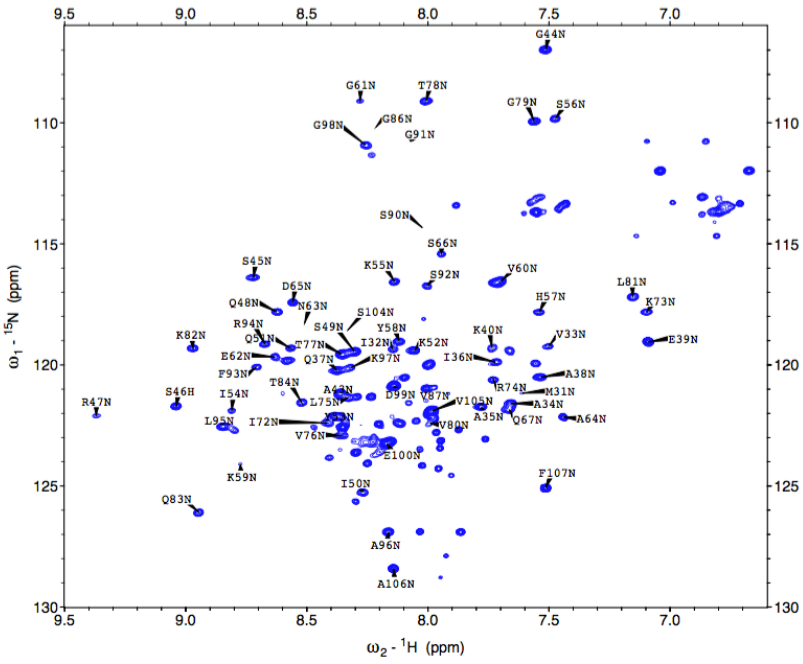

Figure 3.2 HSQC Assigned spectrum

3.4 Strip Plot

SPARKY, a NMR data-visualizing tool includes various options; strip plot is one of them with a standard extension file strips.py. A strip plot is a convenient way of representing 3D spectra in narrow 2D portions. It also allows the display of one or many spectra at the same time.[8] Determining the position of each strip is made by two parameters, i.e., the frequencies of carbon atom and its attached hydrogen atom. Thus, a comparison of multiple 3D spectra becomes handy at one place and is very helpful in searching the peaks with chemical shifts that matches with selected peak.

For instance, the HNCA strip of each residue of a protein can be viewed with the amide proton chemical shifts on x-axis and alpha-carbon chemical shifts on y-axis and amide nitrogen on z-axis (depending on the plane that is appropriate). The HNCA

spectrum correlates amide proton and nitrogen with intra and preceding Cα atoms. Thus, each strip consists of a peak for Cα of a given residue (*i*) and that of the preceding residue (*i-1*). With multiple strips of HNCA at one place it becomes straightforward to identify the peaks that match with selected nitrogen ppm. In order to find the connectivity between peaks, either $C\alpha_i$ or $C\alpha_{i-1}$ peak should be selected, for example, if $C\alpha_{i-1}$ peak is selected it should match with $C\alpha_i$ peak of its next residue or vice versa. Walking through the strips can proceed forward or backward i.e., $C\alpha_i$ to $C\alpha_{i-1}$ (backward) or $C\alpha_{i-1}$ to $C\alpha_i$ (forward). Therefore, connecting the strips and matching/walking along the protein sequence results in identifying the particular residue type.

Below is an example strip plot for sequential assignment of Cα chemical shifts for a short stretch of sequence in H1.0. The ^{15}N ppm is labeled on the top of the respective strips. The Y-axis shows the carbon ppm and x-axis shows the amide proton ppm. Red peaks represent the positively phased peaks and green peaks are negatively phased peaks. This difference in phase results from aliasing of glycine chemical shifts; these residues have chemical shifts outside of the Nyquist cutoff frequency of the digitally sampled spectrum, and therefore they appear as negative peaks at an aliased frequency instead. SPARKY includes functionality to keep track of aliased peaks by adding or subtracting a sweep width as appropriate. Dotted lines are manually drawn using adobe illustrator to shows the connectivity between the peaks. Black circles represent the $C\alpha_i$ peaks and blue circles represent the $C\alpha_{i-1}$ peaks. Each strip in the below figure (strip plot) has two peaks corresponding to a given residue and its preceding residue, and the connectivity is shown with dotted lines. With the help of these strip plots 90% of the residues are assigned for H1.0 globular domain in our work.

There are many practical difficulties in sequential assignment process. Some of them include overlap of peaks; weak peaks, or peaks that are missing altogether. It is not always true that a strip plot can be very clean with no overlap of peaks. Occasionally two peaks can have close proximity in either carbon or proton dimension and can cause uncertainty in assignment. One of the major problems is missing peaks, which can occur for several reasons, including conformational exchange or hydrogen exchange. In such cases, it can be difficult to assign these residues. As explained above, a strip plot can be made with strips from the same spectrum or from multiple spectra. Thus, while a given peak may not appear on one spectrum, it is possible that peak will appear in another spectrum. For example, the duration of the HNCA pulse program is slightly shorter than the HN(CO)CA program, and therefore hydrogen exchange may be less pronounced in an HNCA experiment. During strip plot generation, if it becomes difficult to proceed further by, it is often fruitful to continue work on a different region of protein sequence. In so doing, one can eliminate possible conflicts and overlapping peaks and thus simplify the completion of the first set of strips. Selecting the first residue (strip) is very important, as it can be used to simplify assignment. For example, by starting with a characteristic chemical shift of ~45ppm, one immediately knows that the given residue is a glycine, and one can begin looking for glycine's in the unassigned sequence that match the observed chemical shift patterns. If a strip plot is made with HNCACB spectrum, then a peak with Cβ chemical shifts of ~19ppm for alanine, ~70ppm for threonine, ~65ppm for serine can be chosen as first strip and can be proceed from there. Thus correlating the connected strips with the sequence of the protein helps in assigning the residue type. Multiple strategies are applied while working with strip throughout the process such as plotting

strips from different spectra, correlating the chemical shifts, initial guessing of residue type and identifying the connections etc. Figure 3.3 shows an example strip plot for sequential assignment of C-alpha chemical shifts in a HNCA spectrum.

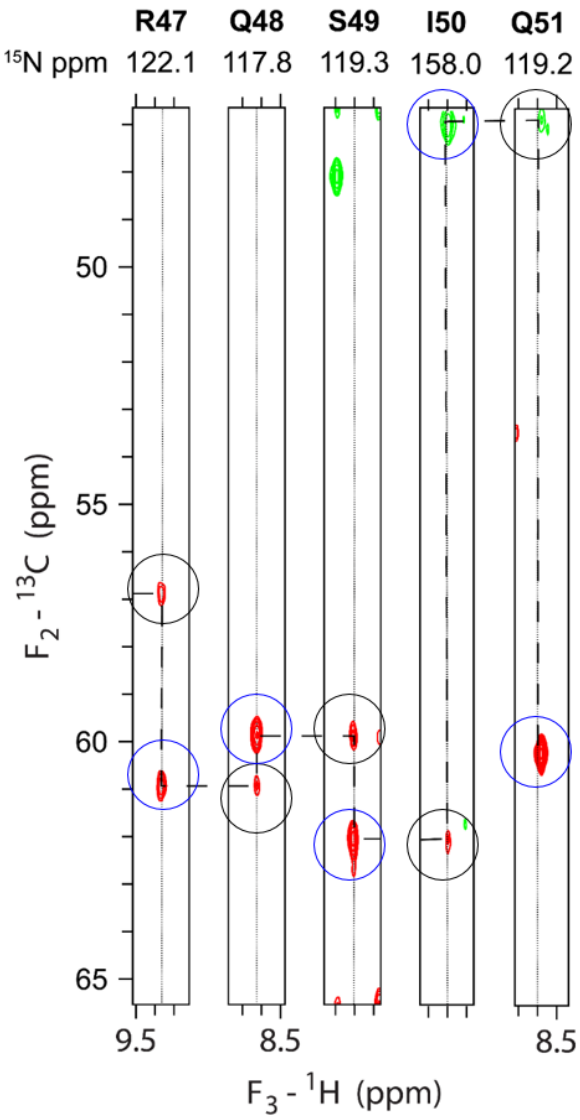

Figure 3.3 Example Strip Plot

Example Strip Plot for Sequential Assignment of C-alpha chemical shifts

3.5 Chemical shifts of Histone H1.0 Globular Domain at pH 6.8

Table 3.1 Chemical shift values for H1.0 globular domain at pH 6.8

Residue	N (ppm)	HN (ppm)	CA (ppm)	CB (ppm)	CO (ppm)
M31	121.125	7.617	59.828	33.642	177.045
I32	119.390	8.147	65.469	37.847	177.026
V33	119.245	7.506	67.674	31.615	177.577
A34	121.604	7.662	55.202	18.131	180.202
A35	121.739	7.783	54.538	20.120	179.920
I36	119.894	7.717	64.735	38.531	180.034
Q37	120.257	8.384	58.445	28.511	177.592
A38	120.536	7.539	53.635	19.128	178.792
E39	119.041	7.094	57.327	29.883	176.470
A43	121.193	8.365	52.259	19.041	177.677
G44	106.984	7.518	44.136	#N/A	172.982
S45	116.408	8.721	58.546	66.211	173.978
S46	121.713	9.042	56.798	65.526	175.056
R47	122.109	9.365	60.95	29.765	177.961
Q48	117.824	8.623	59.816	27.857	178.701
S49	119.453	8.305	62.122	61.709	177.654
I50	125.263	8.271	65.879	38.538	177.678
Q51	119.300	8.567	60.108	29.387	177.456
K52	119.409	8.059	59.597	32.398	179.775

Table 3.1 (Continued)

Y53	122.108	8.389	62.378	38.478	179.567
I54	121.886	8.812	66.903	37.771	177.355
K55	116.583	8.140	59.537	32.139	177.512
S56	109.818	7.476	59.127	64.212	175.246
H57	117.834	7.544	57.914	29.855	172.899
Y58	119.049	8.121	56.329	42.167	174.486
K59	124.068	8.773	56.051	30.612	176.001
V60	116.579	7.704	59.840	33.587	176.469
G61	109.137	8.285	44.217	#N/A	175.124
E62	119.645	8.632	58.228	29.767	177.195
N63	118.270	8.510	52.708	37.392	175.729
A64	122.136	7.444	55.630	19.236	178.772
D65	117.414	8.560	58.679	39.543	179.148
S66	115.433	7.940	61.195	62.347	176.970
Q67	121.867	7.668	58.316	27.780	176.857
I72	122.402	8.413	66.779	38.424	176.819
K73	117.827	7.101	59.945	32.097	179.571
R74	120.637	7.733	59.220	#N/A	179.010
L75	121.380	8.331	57.403	41.630	179.822
V76	122.951	8.358	65.538	68.701	180.455
T77	119.583	8.358	67.051	69.806	176.051

Table 3.1 (Continued)

T78	109.103	8.007	62.501	69.736	176.570
G79	109.963	7.565	45.856	#N/A	175.152
V80	122.156	7.992	65.248	32.821	177.218
L81	117.225	7.154	51.979	38.71	175.583
K82	119.334	8.973	54.278	35.196	174.926
Q83	126.090	8.948	54.547	30.425	175.890
T84	121.564	8.524	61.305	69.732	175.890
G86	110.246	8.220	45.311	#N/A	177.196
V87	121.475	8.010	56.43	30.826	174.014
S90	114.37	8.021	58.531	64.162	177.716
G91	110.738	8.068	45.217	#N/A	174.564
S92	116.746	8.009	57.708	63.884	176.944
F93	120.092	8.707	57.664	43.025	172.616
R94	119.130	8.674	53.385	34.169	174.283
L95	122.548	8.847	55.321	41.097	175.290
A96	126.889	8.166	53.594	19.265	177.778
K97	120.103	8.321	56.079	33.173	177.794
G98	110.950	8.259	45.380	#N/A	176.727
D99	120.868	8.141	53.928	41.050	173.736
E100	123.253	8.170	54.614	29.628	176.000
S104	118.590	8.338	58.192	63.691	176.208

Table 3.1 (Continued)

V105	121.909	7.983	61.793	32.772	174.244
A106	128.417	8.144	52.286	19.516	175.370
F107	125.078	7.515	59.000	40.204	176.177

The residue column represents the single letter code of amino acids in the sequence and their respective residue numbers. Columns N, HN, CA, CB & CO represent the amide, amide proton, and alpha-carbon, beta-carbon and carbonyl carbon chemical shifts in ppm, respectively.

3.6 pH Study

While optimizing NMR spectral quality we noticed a pH-Dependent structural transition for the H1.0 globular domain. We observed a substantial change in the amide proton chemical shifts as the pH was lowered. NMR HSQC spectra collected at different pH conditions are shown in below figures 3.4-3.8. We have also observed the pH change that takes place during protein extraction step as explained in chapter II materials section. This suggests that there may be a cooperative folding transition between lower pH and higher pH. Given our observations, it is likely that this transition occurs just below pH 6.8 (optimal conditions observed from NMR spectrum).

In principle, NMR spectra can be collected at various buffer, temperature or pH conditions. Typically, buffer concentrations range from 20-50mM. A wide range of buffers can be used in the NMR sample, depending on the type of molecule that is measured. Buffers with more protons may interfere with the proton NMR spectra, whereas in a spectrum like isotopically labeled ^{15}N-HSQC, it is not a problem as the protons not attached to the labeled heteronuclei are filtered out. Although high concentrations of buffer can be used, it is always a good practice to minimizing the

concentration of the buffer to avoid non-specific interactions (10-20mM). Similarly, different pH values with a major caveat can be used to collect a NMR spectrum. Backbone and side chain protons can exchange with solvent protons, and the rate of exchange increases logarithmically at above pH ~2.6. When this exchange becomes sufficiently fast, signals from the chemically liable protons will combine with the solvent, and the signal becomes invisible. Most of the times NMR spectra are not collected above pH 7.5 in orders to avoid poor spectral quality.[9]

In our work we have used 50mM phosphate buffer. Figure 3.4-3.8 shows ^{15}N-HSQC NMR spectra collected at different pH values. It is clear observed that change in pH effect the spectrum. Figure 3.4 is collected at pH 7.0 and decent numbers of peaks are visible in the spectrum. It is always necessary to count the signals in a spectrum with number of actual residues in the sequence. Any missing peaks (in number), or the presence of additional peaks can be identified, and this can help to identify additional dynamic processes in the particular protein under investigation. In the course of this study we first collected a ^{15}N-HSQC spectrum for H1.0 globular domain at neutral pH (fig 3.4). We observed less number of peaks when compared with our actual number of residues in the sequence, and our further analysis revealed that certain regions in the protein are undergoing μs –ms conformational exchange. Determining the structure of the protein by X-crystallography is generally performed in single confirmation, whereas protein structure determined through NMR spectroscopy allows studying a protein in solution phase with multiple confirmations, thus allowing understanding the dynamics of the protein. In the process of optimizing we have performed several experiments for the H1.0 globular domain sample at different pH and temperature ranges. Typically, spectra for

both completely folded and unfolded proteins show sharp peaks, whereas NMR spectra for proteins that are partially folded often shows very poor spectral quality as shown in Figure 3.5 (pH 5.9, 297.9K). Surprisingly these, partially-folded proteins can retain secondary structure as indicated in a far–UV CD spectrum, suggesting a highly dynamic, and perhaps molten-globule like state. Large numbers of new peaks are observed in the central region of the spectrum (fig 3.6). This suggests that protein might be unfolding (conformational change) or the protein is degrading. In our case protein degradation is unlikely, as conformational exchange is observed even in a freshly-prepared sample. Thus it is likely that an intermediate exchange is taking place.

Intermediate exchange/chemical exchange is a phenomenon of NMR that refers to a process in which a given nucleus exchanges between two or more chemical environments. Chemical exchange can be either intra-molecular (folding/unfolding of proteins, tautomerization etc.) or inter-molecular (ligand binding to macromolecules, protonation/deprotonation, enzyme catalyzed reactions etc.). Chemical exchange occurs when the microsecond (μs) to millisecond (ms) motions cause a change in isotropic chemical shifts. Experiments for further investigations on intermediate exchange are listed in chapter IV Future Recommendations.

After investigating the pH 6.5 spectra, we further increased the pH value to 6.8 and collected the NMR spectra to observe any changes in spectral quality at slightly higher pH. Surprisingly, we have seen a great decrease in the peaks at the central region of the spectrum (fig 3.7). Spectral appearance can also be affected by change in temperature. Higher temperature can increase the tumbling of the molecule rapidly and line widths can be narrow down. All the spectra (fig 3.4 -3.6) are collected at 297.9K.

The rate of conformational exchange can be influenced by temperature. Several proteins have lower stability at higher temperatures, for such proteins the amide protons rapidly exchange with solvent (typically water) resulting in decreased intensity of signal. So, it is always a good practice for screening the sample/spectral quality at multiple temperatures. The NMR spectrum for pH 6.8 is collected at two different temperatures, fig 3.7 shows the signal collected at pH 6.8 & 287K and fig 3.8 shows the signal measured at pH 6.8 & 293K. Although a change in the signal intensity is observed at the central part of the spectrum when changing the sample pH from 6.5 to 6.8, decreasing temperature results in the appearance of several new peaks in the region between 115-125 nitrogen ppm and 7.0-8.5 proton ppm. On the other hand, better resolution is seen is an ^{15}N-HSQC spectrum at pH 6.8 and 293K. Therefore, it is clearly observed that change in temperature can also affect the exchange process and thus the quality of the NMR spectrum. From all our observations we found that 50mM phosphate buffer, pH 6.8 and temperature 293K are optimal for acquisition.

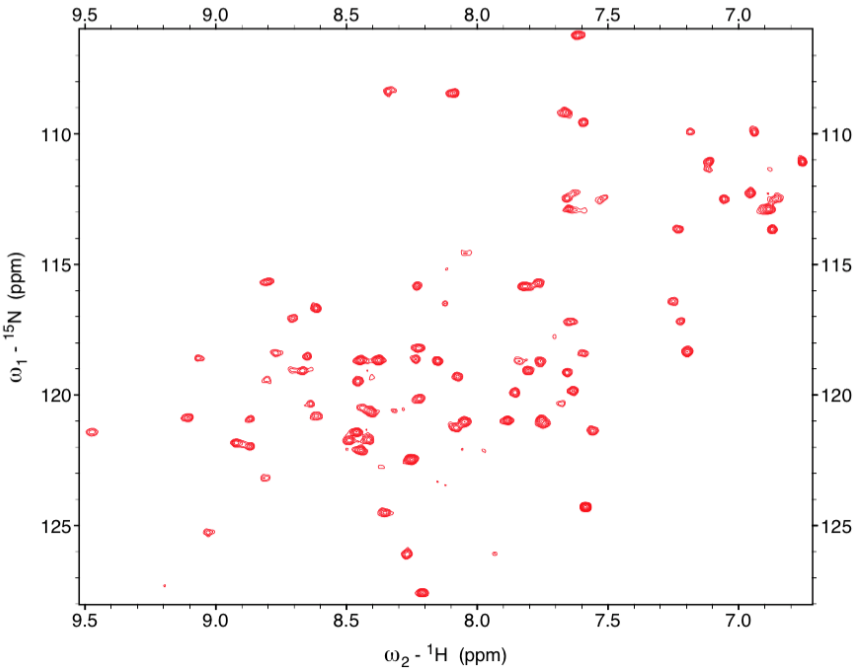

Figure 3.4 HSQC spectrum for purified H1.0g at pH 7.0 and temperature 297.9K

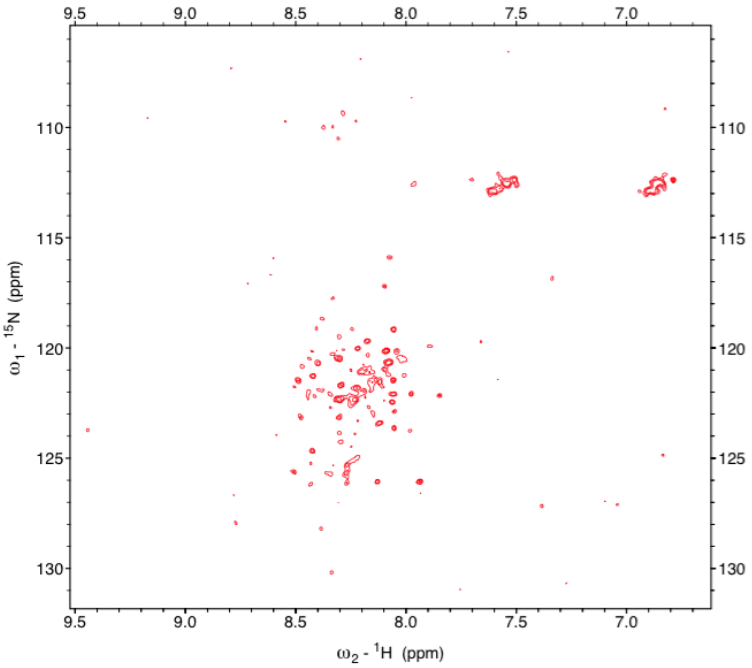

Figure 3.5 HSQC spectrum for purified H1.0g at pH 5.9 and temperature 297.9K

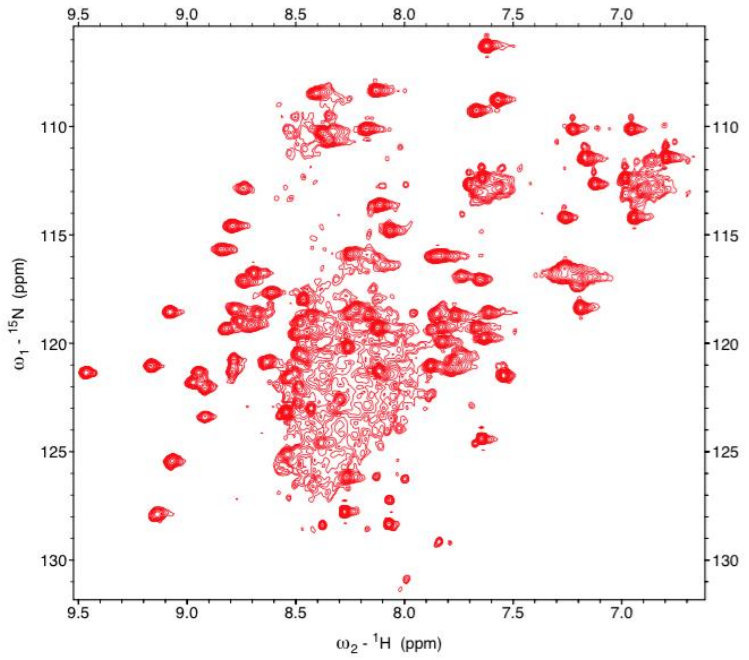

Figure 3.6 HSQC spectrum for purified H1.0g at pH 6.5 and temperature 297.9K

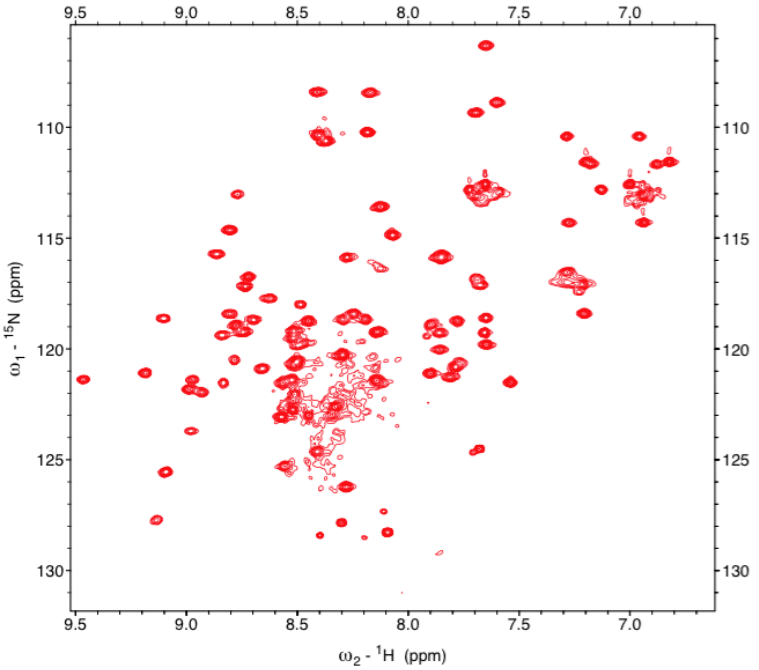

Figure 3.7 HSQC spectrum for purified H1.0g at pH 6.8 and temperature 287K

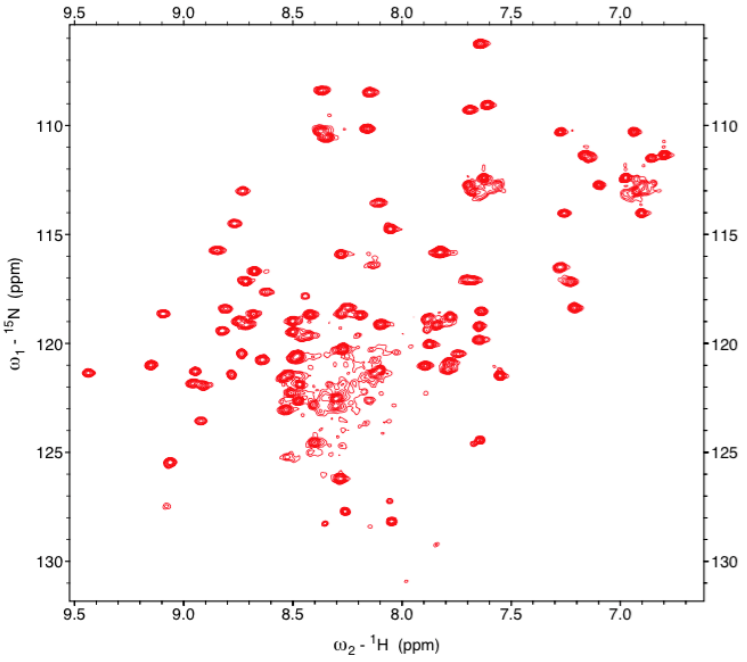

Figure 3.8 HSQC spectrum for purified H1.0g at pH 6.8 and temperature 293K

3.7 Circular Dichroism (CD)

Circular Dichroism (CD) is very useful in understanding secondary structure in DNA and proteins. Each macromolecular structure has a specific signature in CD and this technique can be used to understand and identify the structural elements and their changes. Various secondary structural elements of the protein such as α-helix, β-sheet are most widely studied using CD. Secondary structure prediction is one of the applications of CD. Any change in the structure of the protein can be monitored through circular dichroism. The most commonly studied structure changes via CD include change in temperature, pH, ligands, or denaturant concentration.

In our work we have investigated the changes in structure of histone H1.0 globular domain under different pH conditions. All the experiments were conducted on

Olis DSM 20 spectropolarimeter. A nominal concentration of 30μM protein is maintained for all the experiments. Protein samples are purified as demonstrated in chapter II materials section, and dialyzed overnight in appropriate buffers over a multiple pH range (4.5, 5.5, 6.5, 7.5, and 8.0). CD spectra were collected over a range of 180 to 280nm in a 1mm path length cuvette at room temperature. The spectra represent the average of four scans. CD experiments were used to corroborate any changes in the structure due to the changes in pH. Figure 3.9 shows a CD spectrum for the H1.0 globular domain recorded at multiple pH values. From the figure below, it is shown that a change in pH causes a change in secondary structure of the protein. Interestingly, the spectrum obtained at pH 4.5 has a well-pronounced helical hump; this hump corresponds to α-helix formation and the remaining spectra lack this prominent helical hump. CD spectrum for pH 6.5 is noisy with reduced helical hump when compared with spectrum pH 4.5. CD spectrum at pH 7.5 has a helical shoulder at 220nm. The spectrum at pH 5.5 is mostly distorted. As the spectral quality for pH 5.5 is poor it is hard to estimate the structural behavior of the protein and it needs further investigation. A considerable change in the structure of protein at different pH range is clearly observed. This could be a potential area of study in the future in order to understand the protein behavior at physiological pH versus other pH values.

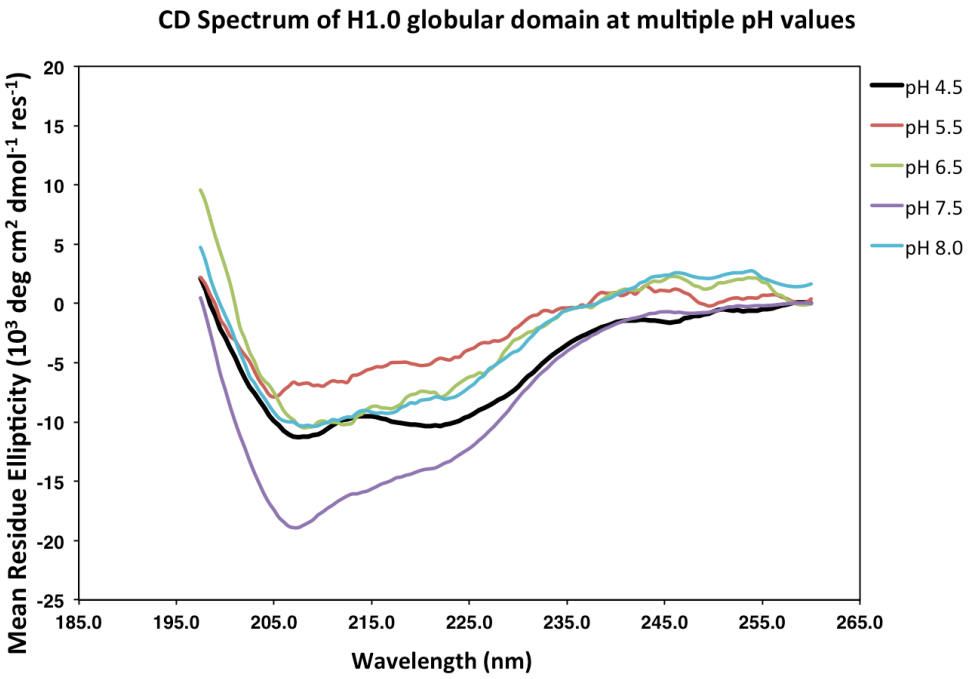

Figure 3.9 CD Spectrum of H1.0g at pH 4.5, 5.5, 6.5, 7.5 & 8.0

3.8 Histone/DNA Binding Studies

After determining the assignments of histone H1.0 globular domain, we examined the binding interactions of the globular domain with DNA using NMR titration experiments. Figures 3.10 & 3.11 represent the HSQC spectra of globular domain without DNA (red) and with DNA (blue), respectively, and figure 3.12 is an overlaid spectrum for the H1.0 globular domain and histone/DNA complex. In these experiments, we used calf thymus DNA to assess potential histone-DNA interactions. This is the approach used by Maccha, *et al*.,[10] and it allows for screening structural changes. Calf thymus DNA (CT-DNA) is a mixture of homogenized DNA molecules from bovine calf thymus extracts; therefore, many potential histone binding sites exist, and DNA size ranges from 587 to 831 bp. The estimated histone binding site size H1.0 is 36bp.[10] In

the experiments shown below, the protein concentration was fixed at 30 μM, whereas the CT-DNA was added to a final concentration of 3.0 μM. There are no distinguishable chemical shift perturbations changes observed in the NMR spectra of the histone/DNA complex, but several peaks are observed to broaden slightly in the presence of DNA. These residues are V60, A96, S49 and S45 and R47, Q83, Q67, K82 and K40.

The existence of two binding sites on globular domain was suggested by Drave's *et al*.[11] Thomas *et al*. has investigated on nucleosome binding site on histone H5 and identified that Lys-85 constitutes a strong binding site for DNA. They have studied these binding sites by methylation of lysine residues and observed that these lysine residues are weakly protected against chemical modifications.[12] Glutamine or glutamic acid were used as a replacement for Lys-85 in a site-specific mutagenesis study, and it was observed that binding of histone H5 to DNA was abolished.[13] The residues in the second binding site of histone H5 were also identified, and it was observed that the basic residues involved in binding are commonly conserved through the histone H1 family.[14, 15] All the above studies were performed for histone H5, and it is believed that same kind of behavior also applies to histone H1, because the structure of histone H1 and H5 are similar.[16] In our study, we also observed a change for the residues R47, K82 & K40 suggesting that our data supports earlier results from the literature. However, our experiments still cannot provide a complete picture of structural changes induced by DNA as hypothesized by Machha *et al.*[10] As mentioned in the pH study section above, we assume that there might be an intermediate exchange process and consequences of these investigations are yet to be investigated. Here we propose two NMR techniques CEST (Chemical Exchange Saturation Transfer) and DEST (Dark state Exchange

Saturation Transfer) for measuring the saturation transfer for small molecules.

Experimental details about these techniques are listed in chapter IV.

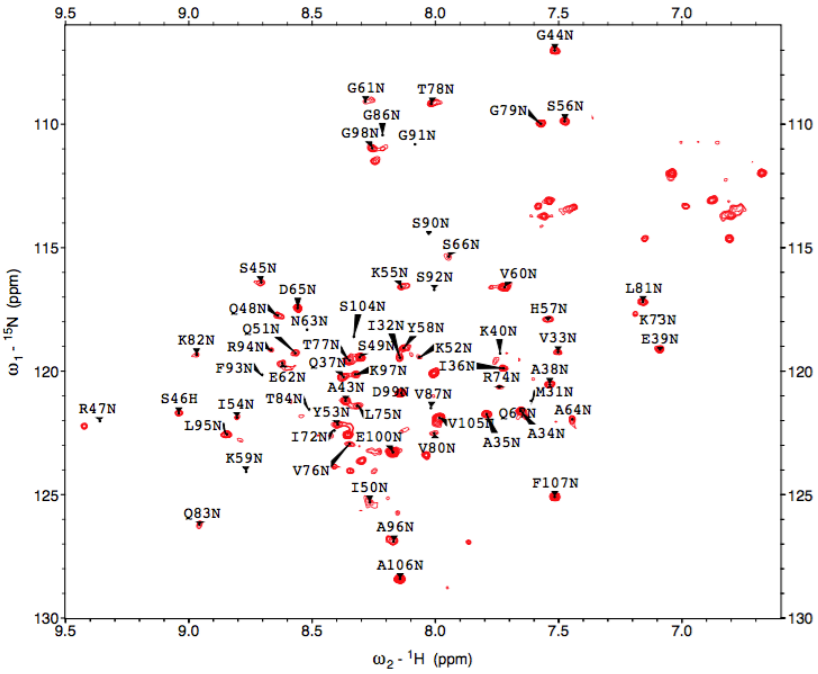

Figure 3.10 H1.0globular domain HSQC without DNA

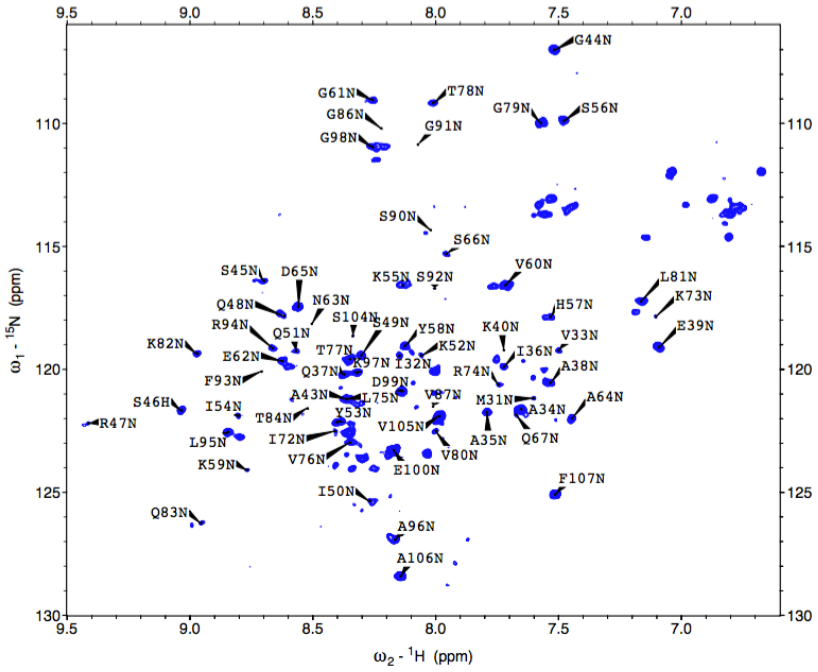

Figure 3.11 H1.0globular domain HSQC with DNA

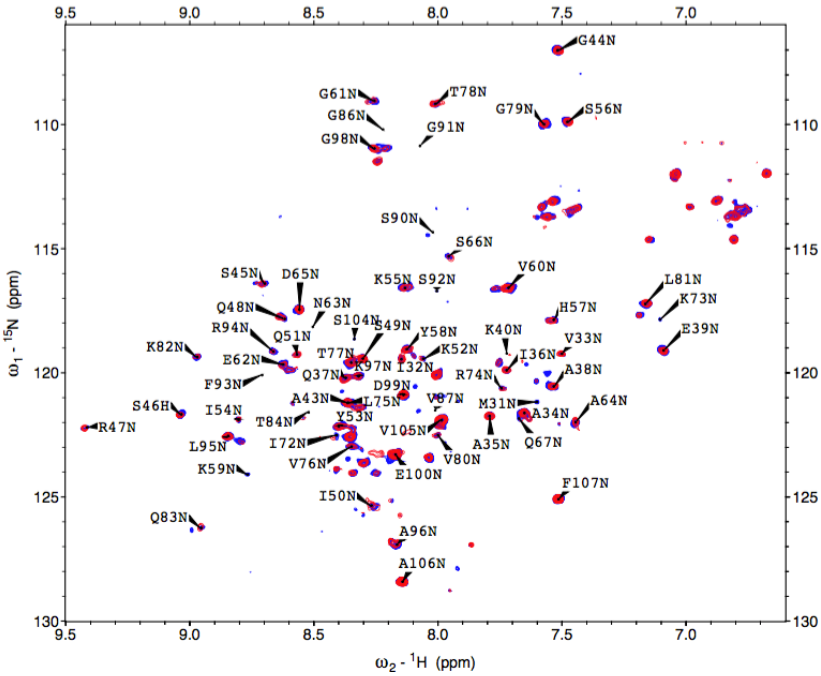

Figure 3.12 Overlaid spectrum of H1.0g

With DNA (blue) and without DNA (red)

3.9 Summary

Linker histones, H1 and its isotypes are thought to stabilize the chromatin fibers.[17] Like all histones proteins, H1 is known to be actively involved in regulation of gene expression. Although various subtypes/isoforms of the H1 histones and their dynamics are well studied, very less is understood about the function of histones.[18-20] In the recent years scientists have studied the energetics and dynamics of the histone H1 and its subtypes using various spectroscopic studies.[10, 21, 22]

The main objective of this study is to attempt to answer a few questions on interactions of histone upon binding with its partner DNA. In the first step we have determined the assignments of H1.0 globular domain using NMR. Data sets are collected and processed using nmrPipe, nmrDraw and SPARKY software's.

Histone H1.0 is 87 residues long with a molecular weight of 9 kDa. Throughout Several modifications were made during purification to increase the yield of the protein, to reduce the preparation time, and to remove the impurities noticed in the initial NMR spectrum. The molecular weight of the protein obtained from the new prep has been verified with SDS-PAGE and ESI-Mass Spectroscopy. In order to improve the spectral quality several two-dimensional HSQC and TROSY experiments were conducted at different pH and temperatures. Comparing the spectra, it is observed that a concentration of 140μM protein in 20mM NaPi, 50mM NaCl, at pH 6.8 and 293K are optimal for acquisition. 90% of residues are assigned and chemical shift values for H1.0globular domain at pH 6.8 are tabulated in table 3.1. In the next step of our study we measured the spectral changes of Histone H1.0 in the presence of calf thymus DNA. For this a fresh sample of histone H1 globular domain is prepared and dialyzed in phosphate buffer at pH

6.8 overnight. HSQC spectrum is collected for the dialyzed sample and is compared with the assigned spectrum to check for spectral differences. In a 1:2 ratio of protein to DNA, only a few changes were observed for the following 9 residues V60, A96, S49, S45, R47, Q83, Q67, K82 and K40. We also expect that there might be an intermediate exchange-taking place during the binding interactions. The experimental techniques that can be used to study the exchange process are discussed in chapter IV, "Future Recommendations."

During the course of study, we have also noticed that the NMR spectral quality changed as pH was lowered. Several HSQC and TROSY experiments were conducted for histone sample at pH 5.9, 6.5, 6.8 and 7.0. Initially, an HSQC was collected at physiological pH, and we observed fewer peaks. In order to understand the behavior of the protein, we have performed several experiments at different pH values. During this process we observed substantial structural changes as reflected both by CD and NMR spectroscopy. Figure 3.9 shows a representative CD spectrum for globular domain at various pH. Preliminary data obtained from CD shows the changes in structure of H1.0 at different pH values. To date many researchers have investigated on the pH and ion strength effects of core histone proteins.[23] It could be a potential area of study in future to investigate the pH and ionic strength effects on the stability of histone H1 and its isoforms.

In conclusion, we have assigned 90% of the residues for histone H1 globular domain of mice and investigated on the interactions of histone/DNA complex using NMR. We have observed peak broadening and decrease in intensity of peaks for 9 residues in our preliminary results. In addition, we have performed set of CD experiments

to observe the changes induced in histone globular domain structure at various pH and all the NMR spectra collected and CD spectrum were shown in figures 3.1 to 3.12.

3.10 References

1. Kamakaka, R.T. and S. Biggins, *Histone variants: deviants?* Genes Dev, 2005. **19**(3): p. 295-310.

2. Arnaudo, A.M., R.C. Molden, and B.A. Garcia, *Revealing Histone Variant Induced Changes Via Quantitative Proteomics.* Critical reviews in biochemistry and molecular biology, 2011. **46**(4): p. 284-294.

3. Cohen, I., et al., *Histone Modifiers in Cancer: Friends or Foes?* Genes & Cancer, 2011. **2**(6): p. 631-647.

4. Wisniewski, J.R., et al., *Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue.* Mol Cell Proteomics, 2007. **6**(1): p. 72-87.

5. Vaquero, A., et al., *Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin.* Mol Cell, 2004. **16**(1): p. 93-105.

6. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.

7. Cavanagh, J., et al., *CHAPTER 10 - SEQUENTIAL ASSIGNMENT, STRUCTURE DETERMINATION, AND OTHER APPLICATIONS*, in *Protein NMR Spectroscopy (Second Edition)*, J. Cavanagh, et al., Editors. 2007, Academic Press: Burlington. p. 781-817.

8. Kneller, T.D.G.a.D.G., *Sparky 3.*

9. Kwan, A.H., et al., *Macromolecular NMR spectroscopy for the non-spectroscopist.* FEBS Journal, 2011. **278**(5): p. 687-703.

10. Machha, V.R., et al., *Calorimetric studies of the interactions of linker histone H10 and its carboxyl (H10–C) and globular (H10–G) domains with calf-thymus DNA.* Biophysical Chemistry, 2013. **184**: p. 22-28.

11. Draves, P.H., P.T. Lowary, and J. Widom, *Co-operative binding of the globular domain of histone H5 to DNA.* J Mol Biol, 1992. **225**(4): p. 1105-21.

12. Thomas, J.O. and C.M. Wilson, *Selective radiolabelling and identification of a strong nucleosome binding site on the globular domain of histone H5.* The EMBO Journal, 1986. **5**(13): p. 3531-3537.

13. Buckle, R.S., J.D. Maman, and J. Allan, *Site-directed mutagenesis studies on the binding of the globular domain of linker histone H5 to the nucleosome.* J Mol Biol, 1992. **223**(3): p. 651-9.

14. Ramakrishnan, V., et al., *Crystal structure of globular domain of histone H5 and its implications for nucleosome binding.* Nature, 1993. **362**(6417): p. 219-223.

15. Crane-Robinson, C. and O.B. Ptitsyn, *Binding of the globular domain of linker histones H5/H1 to the nucleosome: a hypothesis.* Protein Eng, 1989. **2**(8): p. 577-82.

16. Cerf, C., et al., *Homo- and heteronuclear two-dimensional NMR studies of the globular domain of histone H1: sequential assignment and secondary structure.* Biochemistry, 1993. **32**(42): p. 11345-51.

17. Hashimoto, H., et al., *Histone H1 null vertebrate cells exhibit altered nucleosome architecture.* Nucleic Acids Res, 2010. **38**(11): p. 3533-45.

18. Brown, D.T., B.T. Alexander, and D.B. Sittman, *Differential effect of H1 variant overexpression on cell cycle progression and gene expression.* Nucleic Acids Res, 1996. **24**(3): p. 486-93.

19. Sancho, M., et al., *Depletion of Human Histone H1 Variants Uncovers Specific Roles in Gene Expression and Cell Growth.* PLoS Genetics, 2008. **4**(10): p. e1000227.

20. Parseghian, M.H. and B.A. Hamkalo, *A compendium of the histone H1 family of somatic subtypes: an elusive cast of characters and their characteristics.* Biochem Cell Biol, 2001. **79**(3): p. 289-304.

21. Machha, V.R., et al., *Exploring the energetics of histone H1.1 and H1.4 duplex DNA interactions.* Biophys Chem, 2014. **185**: p. 32-8.

22. Machha, V.R., *Thermodynamic characterization of linker histone binding interactions with ds-DNA*, in *Department of chemistry*2014, Mississippi State University. p. 95.

23. Karantza, V., E. Freire, and E.N. Moudrianakis, Thermodynamic studies of the core histones: pH and ionic strength effects on the stability of the (H3-H4)/(H3-H4)2 system. Biochemistry, 1996. **35**(6): p. 2037-46.

CHAPTER IV

RECOMMENDATIONS FOR FUTURE STUDIES

4.1 Introduction

In our study we have investigated the structural changes in histone H1.0 using NMR. During the course of study, we observed an intermediate exchange in the protein signal. Intermediate exchange/chemical exchange is a phenomenon of NMR that refers to a process in which a given nucleus exchanges between two or more chemical environments. Chemical exchange can be either intra-molecular (folding/unfolding of proteins, tautomerization etc.) or inter-molecular (ligand binding to macromolecules, protonation/deprotonation, enzyme catalyzed reactions etc.). Chemical exchange occurs when the microsecond (μs) to millisecond (ms) motions cause a change in isotropic chemical shifts.[1]

Any system can undergo slow, intermediate or fast exchange process. Understanding the structural bases of a protein requires the understanding of their dynamic properties, which are essential in describing the biological functions such as binding, catalysis and regulation. NMR spectroscopic techniques stand as a unique potential tool for measuring protein dynamics.[2]

Here we propose two NMR techniques for measuring the saturation transfer for small molecules. NMR saturation techniques were initially developed in early 1960's.[3] CEST (Chemical Exchange Saturation Transfer) and DEST (Dark state Exchange

Saturation Transfer) are the two popularly studied saturation transfer difference (STD) NMR saturation experiments. Chemical Exchange Saturation Transfer (CEST) is one of the powerful techniques to determine the conformational exchange in high molecular weight compounds.[4] This experiment visualizes a minor state (~2-100 ms) in exchange with a major state and results in significant chemical shifts for major sate.[5] Dark state Exchange Saturation Transfer (DEST) is a novel two-dimensional saturation transfer experiment that determines the exchange process between NMR invisible state, technically "dark-state" and a readily detectable monomeric state.[6] It characterizes the larger molecules that are not visible because of their slow tumbling rates and visualizes the large invisible/dark state as long as it stays in exchange with small visible state on an slow or intermediate time scales.[7-9] In both the experiments nucleus is saturated by off-resonance RF radiation.

4.2 Chemical Exchange Saturation Transfer (CEST)

Isotopic labeling of proteins have become some important criteria for many NMR studies. Proteins can be isotopically labeled either uniformly (^{15}N, ^{13}C) or selectively (^{13}C – methyl labeling) or in a combination of ^{2}H, ^{15}N, and ^{13}C. Methyl labeling of protein has developed many applications, in particular for studying protein dynamics. CEST is one of the NMR techniques that provide the information similar to relaxation dispersion for slow-time scale and can be studied for methyl labeled proteins.

4.2.1 Mechanism

CEST technique is relatively new approach for detecting the exchangeable protons indirectly through the water signal with enhanced sensitivity. In this technique

saturation transfer occurs between the readily exchangeable proton and bulk water proton, thus transferring the saturation to the bulk water protons, and this saturation transfer is measured indirectly through the water signal. In general, water protons are much higher than the solute protons for a sample; in this case, each exchangeable solute proton is substituted by unsaturated proton and allowed for saturation. If the time scale of solute protons is in millisecond range (fast exchange), and time of saturation is long (usually partial saturation occurs), thus an extended radio frequency (RF) irradiation causes significant increase in saturation which becomes visible on the water signal, thus enabling to detect the presence of low concentration solutes. Therefore, the saturation is plotted by water saturation signal normalized by signal with no saturation as a function of saturation frequency, which is often called CEST spectrum.[10, 11]

4.3 Dark state Exchange Saturation Transfer (DEST)

Proteins that tumble rapidly have slow R_2 rates and can be visible by NMR. On the other hand, large proteins or any other macromolecules tumble slowly and will have large R_2 rates, and is very difficult or sometimes impossible to detect (invisible) the signal as it gets broadened. However, if this invisible state is in contact with any molecule with slow R_2 rates, it can be possible to detect the signal for this invisible/dark state indirectly by using DEST experiment. Macromolecules with a molecular weight >1MDa can be studied by this experiment.

4.3.1 Mechanism

When a visible state is in exchange with invisible state (dark state), both the states contribute to the overall R_2 rate for a nucleus signal. Thus the transverse relaxation rate

becomes much fast and quickly disappears for the invisible state where as longitudinal magnetization remains intact for visible state. Thus by slightly tilting the bulk magnetization of the invisible state (technically sample) towards the transverse phase and by applying the off-resonance RF field, the signal can be attenuated rather than destroyed. In all conditions the invisible state relaxes quickly than the visible state but the magnetization between the molecules remains. Thus allowing collecting the signal from the visible state by the DEST experiment.[3]

4.4 Conclusions

Many biological processes such as protein-protein interactions, protein folding, protein aggregation, and protein-macromolecule (DNA) interactions involve an intermediate state/intermediate exchange while studying the complexes in solution phase. This intermediate exchange may be because of several reasons such as their invisible states or protein aggregations. NMR spectroscopy has a wide range of applications particularly in investigating/understanding the biological processes. Although many new experimental techniques have been developed in NMR, it still has limitations for high-molecular weight proteins. Recent advances in saturation techniques like CEST & DEST make possible the collection of NMR signal/data for higher molecular weight compounds as well as the detection of invisible/dark states in the intermediate exchange regime.[7-9, 11]

4.5 References

1. Perrin, C.L. and T.J. Dwyer, *Application of two-dimensional NMR to kinetics of chemical exchange.* Chemical Reviews, 1990. **90**(6): p. 935-967.

2. Kleckner, I.R. and M.P. Foster, *An introduction to NMR-based approaches for measuring protein dynamics.* Biochimica et biophysica acta, 2011. **1814**(8): p. 942-968.

3. Anthis, N.J. and G.M. Clore, *Visualizing transient dark states by NMR spectroscopy.* Quarterly Reviews of Biophysics, 2015. **48**(01): p. 35-116.

4. Rennella, E., et al., *(13)CHD2-CEST NMR spectroscopy provides an avenue for studies of conformational exchange in high molecular weight proteins.* J Biomol NMR, 2015. **63**(2): p. 187-99.

5. Libich, D.S., et al., *Probing the transient dark state of substrate binding to GroEL by relaxation-based solution NMR.* Proceedings of the National Academy of Sciences, 2013. **110**(28): p. 11361-11366.

6. Fawzi, N.L., et al., *Atomic resolution dynamics on the surface of amyloid β protofibrils probed by solution NMR.* Nature, 2011. **480**(7376): p. 268-272.

7. Fawzi, N.L., et al., *Atomic-resolution dynamics on the surface of amyloid-beta protofibrils probed by solution NMR.* Nature, 2011. **480**(7376): p. 268-72.

8. Fawzi, N.L., et al., *Kinetics of Amyloid β Monomer-to-Oligomer Exchange by NMR Relaxation.* Journal of the American Chemical Society, 2010. **132**(29): p. 9948-9951.

9. Fawzi, N.L., et al., *Probing exchange kinetics and atomic resolution dynamics in high-molecular-weight complexes using dark-state exchange saturation transfer NMR spectroscopy.* Nat Protoc, 2012. **7**(8): p. 1523-33.

10. van Zijl, P.C.M. and N.N. Yadav, *Chemical Exchange Saturation Transfer (CEST): what is in a name and what isn't?* Magnetic resonance in medicine official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, 2011. **65**(4): p. 927-948.

11. Vallurupalli, P., G. Bouvignies, and L.E. Kay, *A computational study of the effects of (13) C-(13) C scalar couplings on (13) C CEST NMR spectra: towards studies on a uniformly (13) C-labeled protein.* Chembiochem, 2013. **14**(14): p. 1709-13.