2020

Analysis of Coevolving Residues of xCDxCDx-PHD, A Distinct type of PHD-Finger

Shraddha Basu
South Dakota State University

ANALYSIS OF COEVOLVING RESIDUES OF xCDxCDx-PHD, A DISTINCT TYPE OF PHD-FINGER

By

SHRADDHA BASU

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy Major in Biochemistry

South Dakota State University

2020

DISSERTATION ACCEPTANCE PAGE

Shraddha Basu

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Dr. Suvobrata Chakravarty
Advisor Date

Douglas Raynie
Department Head Date

Dean, Graduate School Date

DEDICATION

I would like to dedicate this dissertation to my loving parents Mr. Sib Ratan Bose and Mrs. Susmita Bose for all the unconditional love and sacrifices they have made throughout their lives to see this dream come true. I am forever grateful to you.

I like to dedicate my dissertation to my elder brother Mr. Souvik Bose, without your continuous support and love I couldn't make it possible. Thank you so much for being such a wonderful brother.

This dissertation would not have been completed without the continuous trust, love and support of my love Mr. Anindya Sen Gupta. Thank you for being the reason of my dream. This achievement is dedicated to you.

I am thankful to have Mrs. Payel Bose as my sister-in-law whose idealism always encourages and motivates me a lot. This achievement is not completed without my little niece Ms. Priyasha Bose.

ACKNOWLEDGEMENT

I am thankful to the almighty God for giving me this wonderful opportunity of pursuing the Ph.D. and continuous support for surviving this journey and finally making it successful. I would like to express my deepest gratitude to my advisor Dr. Suvobrata Chakravarty. Without his continuous guidance and supervision, I could not even shape up my research work. 'Thank you' is not enough for expressing the heartfelt appreciation to my advisor. Your advice, mentorship, and criticism enriched me tremendously. Thanks a lot for being there for me continuously and reminding me of my potentiality. I am forever thankful to you Dr. Suvo. I am thankful to my precious four (Mr. Sib Ratan Bose, Mrs. Susmita Bose, Mr. Souvik Bose, and Mr. Anindya Sen Gupta) for their continuous support, encouragement, love, and enormous faith on me. I am forever indebted to my parents for being there for me throughout my life and made me successful. I thank Mrs. Payel Bose and Ms. Priyasha Bose for being a significant part of my life. I like to thank my sister Mrs. Mousumi Bose for being a wonderful teacher. I am thankful to Mr. Tapan Sen Gupta, and Mrs. Rekha Sen Gupta for blessing me all the way. A special thank you goes to my past lab mates Dr. Daniel Boamah and Dr. Tao Lin for teaching me in and out of the research. From bottom of my heart, I would like to thank my precious friends and sisters Fatima Tuz Zafrin Tuli, Manisha Das, Farhana Akhter, Anyesha Sarkar, and Jawata Afnan Saba for continuous support of living life of a graduate student and unconditional support.

I thank the Department of Chemistry and Biochemistry, the Biochemical Spatio-temporal NeTwork Resource (BioSNTR) and the National Institute of Health (NIH) for their financial support at different stages of this career. A special thank you goes to my dear committee members (Dr. Douglas Raynie, Dr. Fathi Halaweish, and

Dr. Carotta Christin) for their tremendous help and encouragement at every step of my graduate study.

I am grateful to my amazing brothers Indranil Bhattacharjee, and Subrata Bhattacharya and dear friends Dr. Angelina Sampson, Dr. Felix Acheampong, Dr. Partha S. Saha, Dr. Sonali Pandey, Imtisal Razzaq, Smita Pal Sur, Jadav Balawanth Rao, Tama Saha, Sajjadur Rahman, Nizam Uddin, Brian Kobylkevich, and Ranen Roy for their continuous support, and help in my graduate study during this journey. I would like to thank my lab mates Samuel Adjei, Ujwal Subedi, and Franchesca AnneMarie Poppinga for their continuous cooperation. I thank the department of chemistry and biochemistry for giving me this wonderful opportunity of pursuing Ph.D.

TABLE OF CONTENTS

ABBREVIATIONS----------viii

LIST OF FIGURES----------x

LIST OF TABLES----------xiii

ABSTRACT----------1

CHAPTER I----------2

INTRODUCTION----------2

1.1 GENERAL INTRODUCTION----------2

1.2 INTRODUCTION TO RESEARCH----------13

CHAPTER II----------18

PREPARATION OF REAGENTS FOR DETERMINING THE STRUCTURAL STABILITY OF THE MUTATIONS IN THE PHD FINGER

ABSTRACT----------18

INTRODUCTION----------19

MATERIALS AND METHODS----------21

RESULTS AND DISCUSSIONS----------40

CHAPTER III----------48

DETERMINATION OF THE PROTEIN STABILITY OF THE PHD FINGERS

ABSTRACT----------48

INTRODUCTION----------49

MATERIALS AND METHODS----------52

RESULTS AND DISCUSSIONS----------58

CHAPTER IV----------72

BINDING STUDY OF THE PHD FINGER AND CASE STUDY OF THE WEAK INTERACTION USING QM-PDZ DOMAIN

ABSTRACT----------72

INTRODUCTION----------73

MATERIALS AND METHODS----------74

RESULTS AND DISCUSSIONS----------78

CHAPTER V----------88

ABSTRACT----------88

INTRODUCTION----------89

MATERIALS AND METHODS----------91

RESULTS AND DISCUSSIONS----------95

BIBLIOGRAPHY----------97

ABBREVIATIONS

PHD: *P*lan*t H*ome*od*omain

UHRF1: *U*biquitin-like PHD and *R*ING finger domains 1

KDM5B: Lysine (*K*) specific demethylase 5B

KAT6A: Lysine (*K*) *a*cetyla*t*ransferase 6A

GST: *G*lutathione *S*-*t*ransferase

TL: *T*hermolysin

DPF: *D*ouble PHD Finger

BPTF: *B*romodomain *P*HD finger *t*ranscription *f*actor

AIRE: *A*utoimmune *r*egulator

IPTG: *I*sopropyl β-D-1-*t*hiogalactopyranoside

FPLC: Fast Protein Liquid Chromatography

ITC: Isothermal Titration Calorimeter

CD: Circular Dichroism

PCR: Polymerase Chain Reaction

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

FASTpp: Fast parallel protein proteolysis

SDM: Site directed mutagenesis

NMR: Nuclear Magnetic Resonance

HSQC: Heteronuclear single quantum coherence spectroscopy

HNOE: Hetero nuclear overhauser effect

LB Media: Lysogeny or Luria Broth

TB Media: Terrific Broth

PDZ: Postsynaptic density, Drosophila disc large tumor suppressor, and Zonula occludens-1

HAT: Histone acetyltransferases. HDACs: Histone deacetylases

LIST OF FIGURES

Figure 1.1. *Writers, erasers*, and *readers*----------3

Figure 1.2. Taxonomic affiliations of the PHD finger subtypes----------5

Figure 1.3. Topology of PHD finger----------7

Figure 1.4. Sequence alignments of the PHD finger subtypes----------8

Figure 1.5. Structural deviations among PHD finger subtypes----------9

Figure 1.6. PHD_nW_DD associated Human disorders----------12

Figure 1.7. Cloning hUHRF1 PHD----------24

Figure 1.8. Recombinant protein expression----------27

Figure 1.9. Chromatogram of the GST-hAIRE showing FPLC profile----------28

Figure 2.0. Chromatogram of the ^{15}N-KDM5B-Wt showing FPLC profile----------34

Figure 2.1. Chromatogram of the ^{15}N-KDM5B-D331A showing FPLC profile-----35

Figure 2.2. Chromatogram of the ^{15}N-KDM5B-L326A showing FPLC profile-----35

Figure 2.3. Chromatograms of the ^{15}N-KDM5B-V346A showing FPLC profile----35

Figure 2.4. Chromatogram of the ^{15}N-UHRF1-Wt showing FPLC profile----------36

Figure 2.5. Chromatogram of the ^{15}N-UHRF1-M345A showing FPLC profile------36

Figure 2.6. Chromatograms of the ^{15}N-UHRF1-V365A showing FPLC profile-----37

Figure 2.7. SDS PAGE gel of ^{15}N labelled KDM5B Protein purification----------37

Figure 2.8. SDS PAGE gel of ^{15}N labelled UHRF1 Protein purification----------38

Figure 2.9. Steps of Site directed mutagenesis (SDM)----------39

Figure 3.0. Standard chemical shift of folded and unfolded proteins----------41

Figure 3.1. NMR-HSQC spectra of KDM5B and its mutants----------44

Figure 3.1. T1, T2 & HetNOE of KDM5B wild type and its L2 mutant----------44

Figure 3.2. Schematic diagram of the FASTpp experiment----------53

Figure 3.3. Controlling factors of FASTpp experiment----------59

Figure 3.4. Role of Zinc atom in folding nature of PHD finger----------60

Figure 3.5. Stability test (FASTpp) of Reference proteins at Gradient Temp----------62

Figure 3.6. Comparative Stability test (FASTpp) of different PHD fingers----------62

Figure 3.7. Comparative Stability test (FASTpp) of wt and mt KDM5B ----------62

Figure 3.8. Comparative Stability test (FASTpp) of 4M Urea treated wt and mt KDM5B-62

Figure 3.9. FASTpp of the PHD finger wt and mt as a function of time----------63

Figure 4.0. FASTpp of RNase A as a function of Melting Temperature (Tm)--------64

Figure 4.1. FASTpp of GST as a function of Melting Temperature (Tm)----------64

Figure 4.2. Circular Dichroism (CD) of KFM5B-PHD wt and mt----------65

Figure 4.3. CD of KDM5B-PHD wt and mt at different temperatures----------65

Figure 4.4. CD of 1mM DTT treated wt and mt KDM5B-PHD finger----------67

Figure 4.5. Protein unfolding as observed by Tryptophan fluorescence----------67

Figure 4.6. Fluorescence intensity of the Tryptophan of KDM5B----------69

Figure 4.7. Comparison of the chemical shift as a function of temperature----------69

Figure 4.8. Statistical analysis of the melting comparison of wt and mt KDM5B-------70

Figure 5.0. Chromatogram showing FPLC profile of His-BPTF purification----------75

Figure 5.1. Schematic diagram of how ITC works (mechanism)----------78

Figure 5.2. ITC binding studies between KDM4B/UHRF1-PHD vs H3T3S----------80

Figure 5.3. Diagram of the Arginine specific mutations on the Histone H3 peptide--81

Figure 5.4. ITC binding studies between PHD finger and mutated Histone peptide--82

Figure 5.5. ITC binding studies between BPTF wild type Vs H3K4Me3----------85

Figure 5.6. ITC showing contribution of the key factors in the tight packing---------85

Figure 5.7. ITC between PHD finger (F to L mutation) Vs H3K14AC----------87

Figure 5.8. ITC binding studies between DPF3-PHD finger Vs H3K14AC----------88

Figure 5.9. Diagram of the Anion-Quadrupole interaction----------90

Figure 6.0. Chromatogram showing FPLC profile of GST-QM wild type----------93

Figure 6.1. Chromatogram showing FPLC profiles of QM mutants----------94

Figure 6.2. Thermodynamic quantification (ITC) of anion-quadrupole interaction-96

LIST OF TABLES

Table 1.1: PCR Reagents for hUHRF1 cloning----------21

Table 1.2: PCR Reaction cycle----------22

Table 1.3: Nucleotide sequence of cDNA and sequences of Primers----------25

Table 1.4: Reaction cycle of Mutagenesis----------28

Table 1.5: Reagents for Site Directed Mutagenesis----------29

Table 1.6: Protein sequence of the hUHRF1-PHD finger mutations----------30

Table 1.7**:** Sequence specificity of different proteins including PHD fingers--------48

Table 1.8: List of mutations considered for binding studies of the PHD fingers-----83

Table 1.9: Nucleotide sequence of the cDNA of QM Protein----------91

Table 2.0: Primers and Protein sequence of the mutants created on QM----------92

ABSTRACT

ANALYSIS OF COEVOLVING RESIDUES OF xCDxCDx-PHD, A DISTINCT TYPE OF PHD-FINGER

SHRADDHA BASU

2020

*P*lant homeodomain (PHD), a Zinc-finger scaffold, is a conserved protein module in eukaryotes that typically recognizes unstructured histone tails, and thus, PHDs play a crucial role in chromatin signaling. The sequences of Zinc-fingers, in general, diversify during the course of evolution often giving rise to subtypes (e.g., RBR-RING (Dove 2017) or zf-CxxC subtypes (Long 2013) who typically contain specific sequence signatures. We recently discovered that PHD fingers also contain a distinct subtype, namely the xCDxCDx-PHD. xCDxCDx-PHD has a distinct composition of specific amino acids that coevolved (coevolving residues) in the course of evolution, and xCDxCDx-PHD also shows a unique mechanism for interacting with histone H3 (i.e., unique function). It, however, is unclear if the set of coevolving residues were selected by nature exclusively for function alone or the selection was for the maintenance of the structure, stability and/or dynamics in order to perform the specific function. Here, we systematically characterize the contributions of the evolving residues in the PHD scaffold with detailed experimentation (e.g., using NMR, CD, FASTpp and ITC) to discover that the coevolving residues of xCDxCDx-PHD contribute only towards histone recognition (i.e., function). As somatic and germline mutations of xCDxCDx-PHD are often associated with a number of cancers and inherited diseases, this investigation provides a platform to further probe the consequence of disease-causing mutations in xCDxCDx-PHD.

CHAPTER I

INTRODUCTION:

General introduction:

Chromatin is the physiological state of a eukaryotic genome. That is, the genetic blueprint of DNA molecule, in combination with different proteins, is organized hierarchically to constitute chromatin with nucleosomes forming the unit of the chromatin structure. Nucleosomes are composed of four positively charged histones proteins (H2A, H2B, H3, and H4) forming a compact structure in association with 146 base pairs of negatively charged genomic DNA (Luger K 1997). The characteristic histone fold is the key structural feature of all four histone proteins that are utilized for strong interaction with DNA. In addition to the histone fold, all four histones also consist of disordered N- and C-terminal tails that serve as critical information retrieval network hubs for chromatin signaling. For their structural malleability, the unfolded tails are bound by several proteins to orchestrate chromatin signaling pathways, which collectively regulate chromatin structure and thereby exert ultimate control over all the diverse biological outcomes associated with chromatin (Yun M 2011).

The tails serve as docking sites for proteins engaged in chromatin signaling and the recognition of disordered peptide segments of histones is a crucial component of these pathways. Histone peptide segments, with or without site-specific post-translational modifications (PTMs), provide the binding surface for regulatory protein complexes to anchor onto chromatin. In general, peptide-mediated signaling operates with a three-part toolbox (Musselman CA 2012), consisting of modular protein domain families of (a) enzymes that catalyze the addition of site-specific PTMs, called marks, (b) peptide-anchoring modules that bind peptides with or without the marks and (c) enzymes that catalyze the removal of these marks. In other words, *writers* (methylases and

acetylases) write specific marks, *readers* (peptide binding domains) interpret the marks, and *erasers* (demethylases and deacetylases) erase the marks. Anchoring of unmodified histones (histone peptides bearing no marks) by readers is a crucial component of chromatin regulation, as the introduction of marks at these sites can disrupt such anchoring and thereby regulate information flow. However, compared with the number of reported readers of histone site-specific PTMs, there are fewer reports on the readout of unmodified histones. Hence, to better understand the mechanism of readout of unmodified histones, we focused on one specific type of reader, the *P*lant homeodomain (PHD) finger which specifically binds the unmodified histone H3 N-terminal peptide (Qiu Y 2012).

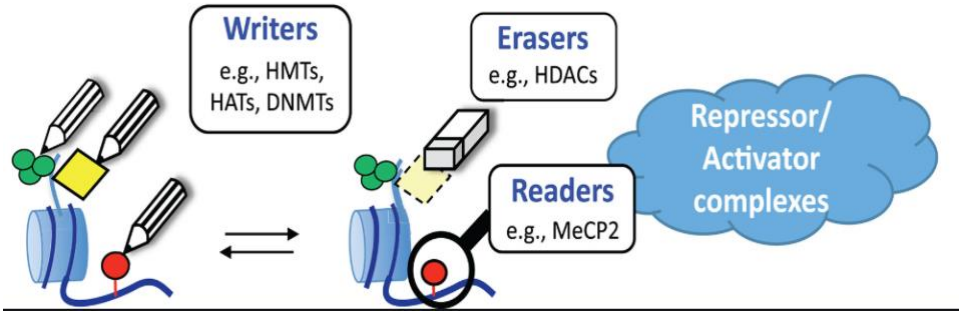


Figure 1.1: Chromatin signaling orchestrated using *writers*, *erasers*, and *readers* (Schang AL 2018)

The PHD finger scaffold, consisting of 60 – 70 amino acids with a distinct Zinc-anchoring amino acid sequence pattern of Cys and His residues, was first reported in 1993 in HAT3 protein of *Arabidopsis thaliana* (Schindler 1993). For its first report in plants, resulted in the PHD. However, PHD finger was noted to be highly conserved in all eukaryotes and found in proteins typically localized in the nucleus (Bienz 2006). These early observations had provided an important clue about the nuclear functions of PHD finger, speculatively in the maintenance of the genome. The histone peptide-

binding molecular function was first reported in 2006 (Schuetz A 2006), establishing PHD finger as an important *reader* of histone marks, particularly for the readout of methylated histone H3 Lys4 (H3K4me3). Since the first report of H3K4me3-readout, studies on PHD fingers, over more than a decade, have shown PHD finger to be a very versatile reader of histone H3. That is, capable of reading not just the H3K4me3 mark but also acylation marks (e.g., acetylation, crotonylation, etc.) and most importantly unmodified histone H3 marks. Among the myriad of readers of modified histones, PHD finger remains an important scaffold known also for the readout of unmodified histones. For its functional versatility, PHD fingers have been subject to structural manipulation to probe the detailed molecular mechanism of chromatin signaling. For example, PHD readers have been reengineered to switch their specificity for probing and altering DNA methylation landscapes in the living stem cells (Musselman AC 2009) and also for probing the pH-dependent chromosomal association/dissociation of readers (Tencer AH 2017).

Reader re-engineering involves the manipulation of a few binding-site residues of an existing reader while leaving the rest of the reader sequence unaltered. Therefore, structure-guided dissection of the key determinants of histone peptide recognition mechanisms will continue to play an important role, not just in probing chromatin signaling mechanisms but also in the design of reagents for diagnostic applications (Chakravarty S 2015). To dissect the key determinants, we previously probed the energetics for recognition of the histone H3 N-terminal peptide by readers belonging to the same structural scaffold, the PHD finger, that recognizes the same peptide sequence. In that study, we found that the readout of the same peptide sequence could differ between readers, and the energetic contributions of the peptide amino acids correlated with the sequence features of the readers. For example, H3 Lys4 makes negligible small

energetic contributions to N-terminal H3 recognition by PHD readers featuring a treble clef xCDxCDx motif.

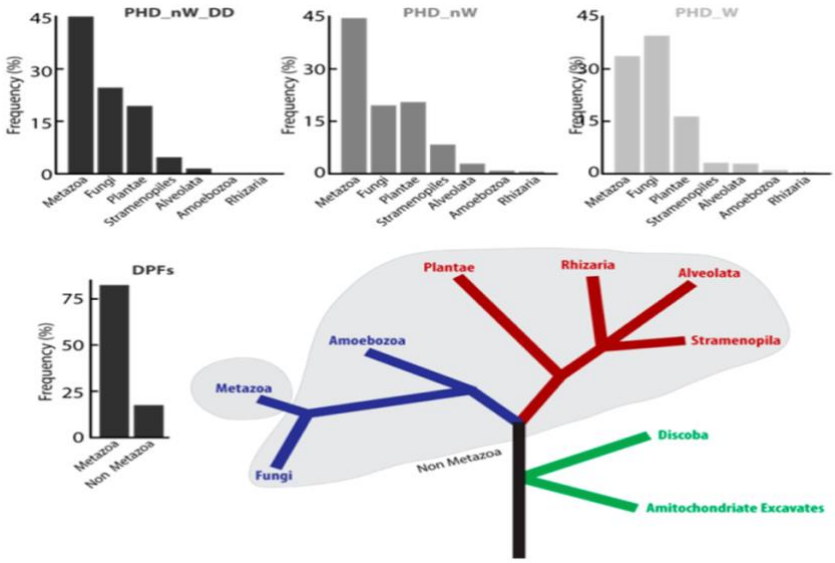


Figure 1.2: Taxonomic affiliations of the PHD finger

Unstructured histone tails are read by the PHD finger in a site-specific way. PHD fingers are zinc fingers and conserved protein domain that plays an important role in chromatin signaling. The sequences of zinc fingers may have numbers of mutations (Hoffman CR 1993), and thus it gives rise to different subtypes of the PHD fingers, such as: PHD_nW_DD, PHD_nW, and PHD_W, among which a particular motif xCDxCDx has been observed in the PHD_nW_DD subtype, that made it very unique and behaving differently than other subtypes. The binding site of this subtype is composed of bulky nonpolar and negatively charged amino acids which are not found in other types of the PHD finger.

There are two zinc atoms in the PHD finger that are held with the cystein and histidine residues in the tetrahedral coordination geometry (Capili 2001) and helps in protein folding and maintaining the stability (Cox HE 2000). Since there is a stability

in the core inside of the zinc finger, there is a possibility of finding out novel functions by altering the amino acid sequences. Thus, it has been realized that the zinc finger is a large super family because there are ample possibilities to observe different types of sequence alterations.

We have looked at different types of the PHD finger sequences. We saw that there are subgroups that possess coevolving residues at their binding sites (denoted by 1,2,3,4,5 positions in the cartoon) (Figure 1.2). Among 20 amino acids, they always tend to have few sets of selected amino acids in these positions and they are coevolving together. Means that, at position 1, if amino acids D or E occurs, it has been found that D or E also occurs at position 2. The occurrence of L or F amino acid is also found to be correlated with the occurrence of D or E at position 2. Likewise, D is present on the 4th position if L or F is present at 3rd position. Same trend has been found for the 5th position which is correlated with the 4th position (Figure 1.3). If we look at other subtypes of the PHD finger, this frequency of coevolving residues is much lesser. So, it becomes important to understand that how the zinc fingers diversified, some of them choose to retain a set of amino acids when others didn't.

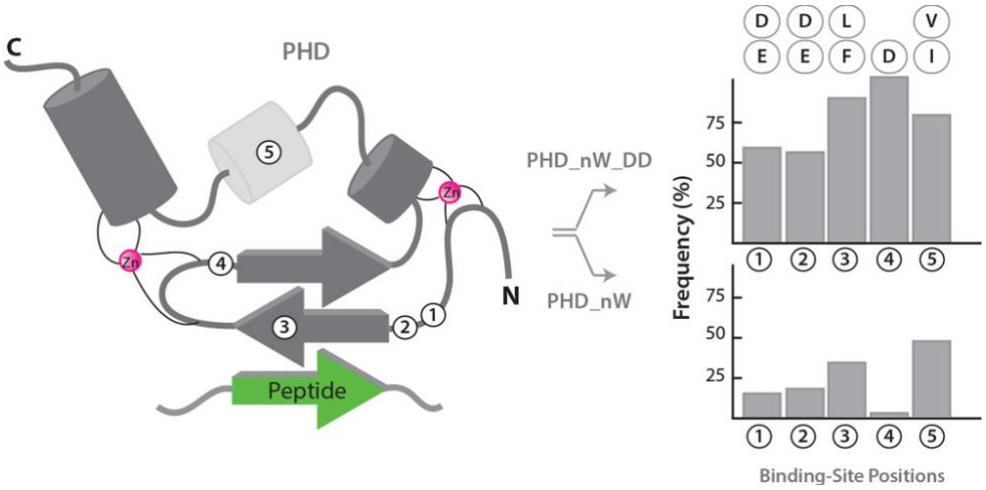
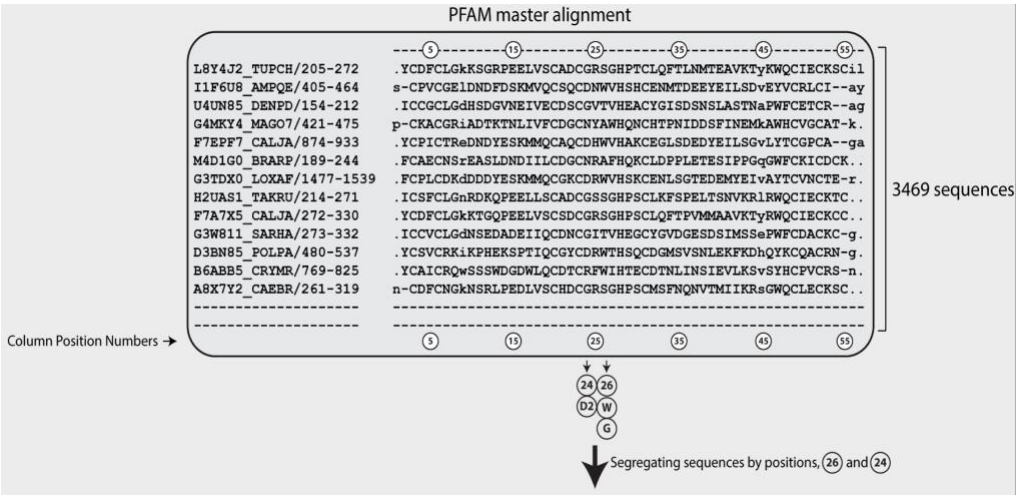


Figure 1.3: Topological pattern of PHD fold in which Cys and His residues are holding the binucleated Zn atoms.

Master alignment of the PHD finger superfamily (10,858 PFAM aligned sequences) has been downloaded from PFAM ftp server. After downloading, using CD-Hit (Li 2002), PFAM alignment has been filtered for eliminating the particular positions that have the symbol "." Filtered sequences were being further extracted to have a non-redundant set and this non-redundant set contained 3,469 sequences. A part of the condensed PFAM master alignment has been shown in Figure 1.4. Without altering the master sequence alignment, it has been segregated into subtypes: PHD_W (1426 sequences), PHD_nW (1351 sequences), and PHD_nW_DD (692 sequnces) subtypes.



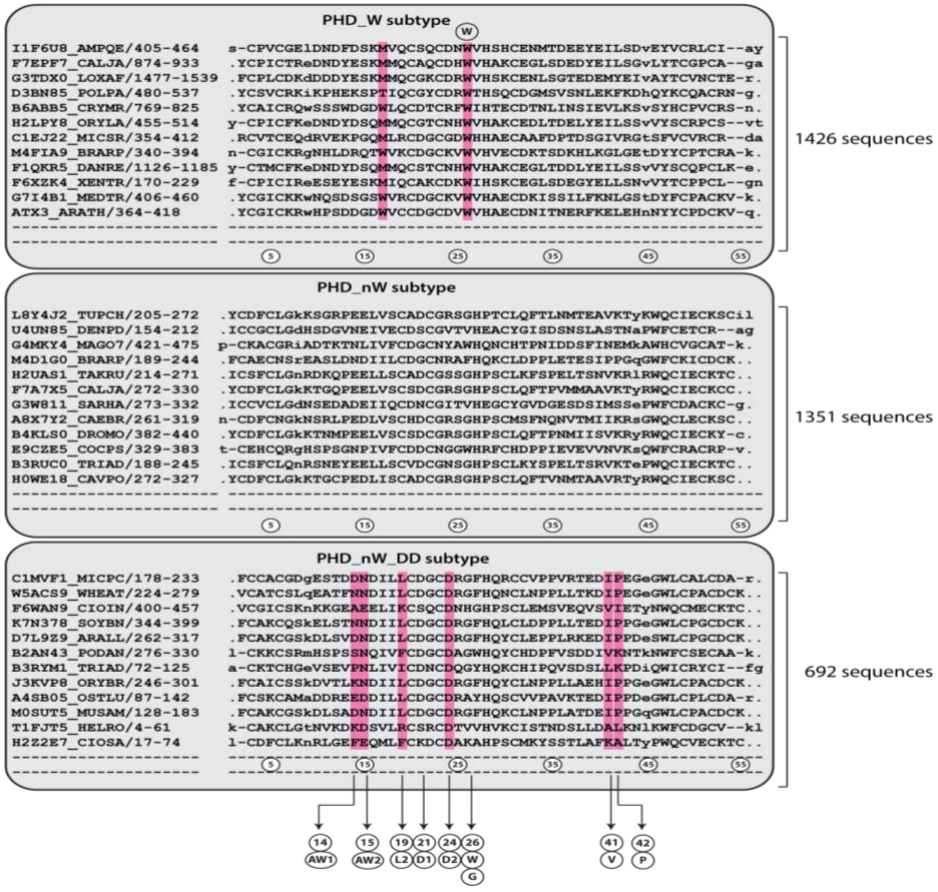


Figure 1.4: Sequence alignment of the PHD Superfamily

Fungi and metazoan are found to dominate the sequence space of PHD_W, whereas in the case of PHD_nW_DD and PHD_nW only metazoan lineage is the dominating one (Boamah D 2018). A more striking proportion has been found in case of DPFs because around 80% of DPF sequences have been found in the metazoan lineage. It clearly indicates that the sequences of PHD_nW_DD and PHD_nW started prevailing in the metazoan lineage. Thus, it is highly possible that N terminal of the unmodified histone H3 tail can be read out by either one of PHD_nW_DD and PHD_nW subtype. So, it is expected that these two subtypes have a significant contribution in the evolution of the metazoan lineage.

It has been found that the subtype PHD_nW_DD that contains the sequence xCDxCDx, behaves differently compared to other PHD fingers that do not have a

sequence feature. In this Figure 1.5A, proteins that belong to PHD_nW_DD subtype, are showing far less structural deviation from each other despite having only 30% sequence identity. Besides this, structural deviations are higher in the case of other subtypes (that do not have xCDxCDx motif) such as PHD_nW (Figure 1.5C) and PHD_W (Figure 1.5D) even with higher sequence identity (~40%). Therefore, the question arises whether the coevolving residues are playing an important role in maintaining the structure or are responsible for binding histone peptide substrates.

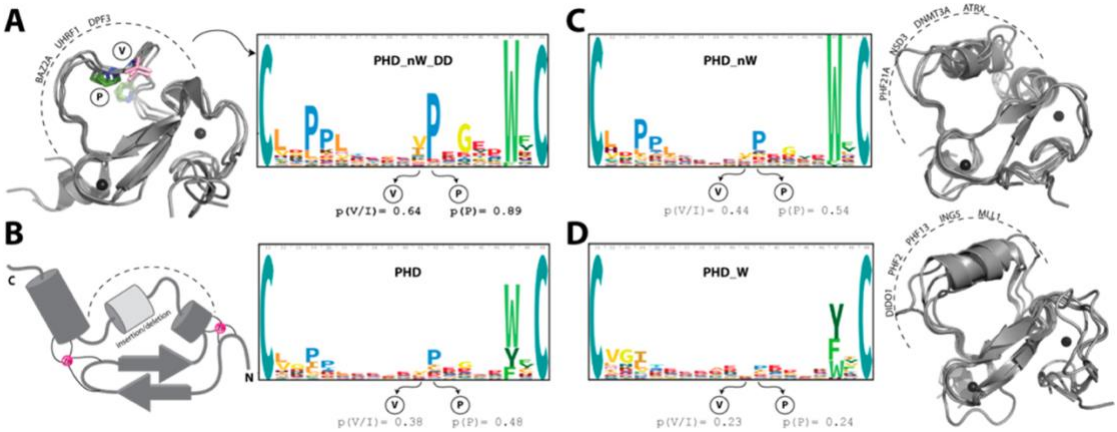


Figure 1.5: Structural deviations among the subtypes of the PHD finger.

From our analysis, it has found that PHD_nW_DD behave differently mostly due to the following reasons: (i) presence of negatively charged and bulky amino acids in the binding site (ii) there is a mutual and positive correlation in the binding site of this subtype, which is not present in other subtypes. (iii) there is small structural deviation despite having low sequence similarity. Considering these three facts, we wanted to determine the contributions of the negatively charged and bulky amino acids in their binding for folding and stability and toward reading histone peptide. Therefore, we made mutations of those negatively charged bulky amino acids keeping the rest of the sequence unaltered. From the previous study, it has been found that the substitution of one amino acid can cause a lot of differences in the binding behavior. For example,

remarkable compromise in the binding has been observed while D1 was substituted in KDM5B (Chakravarty S 2015), UHRF1 (Rajakumara 2011), and KAT6A (Dreveny 2014). We have performed thermodynamic studies for determining the histone recognition mechanism of the PHD finger that plays a very important role in the chromatin signaling. We believe that if we can understand the mechanism made by nature, we can apply the concept for further designing the reagents for diagnostic purposes. We have probed the energetic contribution of the important residues at the binding site of a number of PHD fingers (hKDM5B, hUHRF1, hKDM5D, hAIRE, hKAT6A, hDPF3, hBPTF). We have used the Isothermal Titration Calorimetry (ITC) for studying the binding thermodynamics of the mutations. ITC can directly measure the heat generated during each injection of ligand for the associated change of state after each injection. The heat is proportional to the increment in the concentration of complex (protein-ligand) in the calorimetric cell after the first injection (Velazquez-Campoy 2015).

There are two categories of mutations that can affect the peptide-binding site: (i) positions in the proteins that are in contact with the peptide and (ii) positions distal from the peptide binding site. Here we have probed type (i) mutations using recombinant protein chemistry and biophysical thermodynamic studies (Chakravarty S 2015). We have selected wild type xCDxCDx-PHD (or PHD_nW_DD) and their mutants for studying the binding energetics. Next-generation sequencing also suggests that mutants in xCDxCDx-PHD are associated with different human disorders (Figure 1.6).

When PHD_nW_DD subtype recognizes unmodified H3, the question raises if the recognition is distinct from that of the well-known H3K4me0 reader subtype PHD_nW. In our earlier study, we have discussed and tried to address three questions:

i) how PHD_nW_DD subtype is different from other subtypes of the PHD finger, ii) what are the characteristics of this special type of PHD finger and iii) how does it evolve in course of time? Followed by this study and after addressing the above questions, we are now able to focus on more detailed questions like how a mutation could affect the structure, stability, and peptide recognition behavior of the PHD_nW_DD subtype.

Since the nonpolar residues of the binding site of the PHD finger have likely coevolved with the xCDxCDx motif, we wanted to understand whether these coevolving residues are contributing to (a) folding of a scaffold, and/or (b) function (e.g., substrate binding) or do they contribute for both (a) and (b). We have learned that the coevolving residues of xCDxCDx-PHD are necessary for peptide binding and contribute little toward the folding of xCDxCDx-PHD scaffold.

Previously we have studied the detail of the binding mechanism of the PHD_nW_DD subtype considering few proteins from this group as a model. After that we started taking interest in deeper insight into their molecular structure, stability, and folding pattern. We have chosen the first PHD finger of KDM5B, UHRF1 as representative of this subtype of the PHD finger that harbors xCDxCDx motif. I got the construct of KDM5B-PHD from my past laboratory mate and I have cloned UHRF1-PHD in pGEX4T3 vector taking the template DNA insert from him (UHRF1 was in pET28A vector). Then I have made the mutations in D2, L2, and V positions in UHRF1-PHD and KDM5B. Except for H3 binding, the behavior of these mutants was similar to that of the wild type suggesting that the coevolving residue arose as an adaptation for specific substrate binding. We already concluded from the previous study that mutations can affect their binding ability. Now we wanted to confirm whether mutations are equally playing an important role in other aspects like stability, structure, and protein dynamics. Nature has beautifully crafted the biochemistry of the

proteins and efficiently controlling its mechanism. Upon characterizing the proteins of this particular subtype of the PHD finger, we wanted to understand the underlying mechanism of how nature has crafted and made them special, so that they can smoothly orchestrate the signals through the chromatin signaling pathway. Our findings can provide a detailed insight into the characteristics of xCDxCDx motif and how it is associated with many disorders caused by mutations or protein aberrations. With this understanding, researchers can think of designing the drugs that can be applied for addressing the pathological disorders related to somatic or germline mutations.

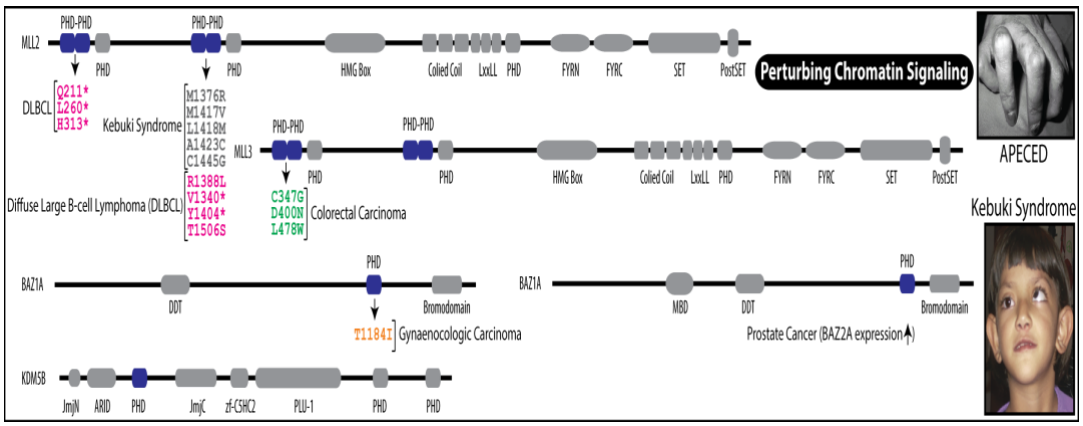


Figure 1.6: PHD_nW_DD associated Human disorders: gynecological carcinoma (Kaustov 2011), colorectal carcinoma (Zeng L. 2010), kebuki syndrome (He 2010, Li W 2016), intraocular medullo epitheliomas, diffuse large B-cell lymphoma (DLBCL), non–small cell lung cancer (NSCLC) (Karpen 1992) mutations are highlighted in orange, green, grey, and pink respectively. '↑' signifies overexpression of the protein in prostate and ovarian cancer. Non-sense mutations are denoted by '*' and protein names are given on the left side.

Through the series of experiments, we confirmed and concluded that the zinc finger is very stable and sturdy. Our research encourages future study for probing the expression levels of zinc-finger proteins in living cells as proteolytic resistance could confer resistance to proteasomal degradation and thereby could alter zinc-finger protein recycling and intracellular concentration. Due to its stability, zinc finger can be used as a scaffold to deliver drugs or other proteins inside the cell.

In my project, we tried to understand whether the coevolving residues are equally responsible for folding, stability, motion, and maintaining the structure of the protein. To find out the answer, we made ^{15}N labeled protein of KDM5B wild type and its mutations and we performed NMR HSQC to determine their molecular structure. To strengthen our hypothesis that mutation affects only the binding and not on the folding, we have repeated the same experiment with another protein, UHRF1 wild type, and its mutations.

Introduction to research:

How mutagenesis affects the function of a protein:

Mutations can affect the protein in several ways. Due to the mutation, protein can be less stable. When a protein folds, there is an equilibrium between the folded and unfolded state. If stability drops, the equilibrium shifts towards the unfolded state and the proteasome removes the unfolded fractions immediately. Therefore, a certain drop in the protein concentration will be noticed. Therefore, simply by stability drop, we can affect the biology of the protein in many ways. It can be removed from the cell very easily and quickly. Mutation can disrupt the binding ability, that is, if the mutation occurs on the surface of the protein, it can directly impact the ligand binding. If the mutation occurs in the core of the protein, it can affect the binding of the protein allosterically. The motion of the protein can be affected by the mutation. The function

of the protein depends on how it moves in the solution. If the motion of the protein is enhanced or compromised, the function of the protein will also be affected and no longer be able to function as effectively as it supposed to be. Mutation can affect the protein folding pathway following which protein either misfolds or becomes aggregated and causes disease like Alzheimer's. In general, in this particular zinc finger where we have the particular motif (xCDxCDx) and noticed coevolving residues in its sequence, we wanted to understand that are these coevolving residues contributing to all of these factors or contributing to one of them. Through next-generation sequencing, it has been found that a lot of somatic mutations like cancer genome sequences are producing. Different populations of the human genome are letting us know that mutants are the rare forms of the allele. If there is any mutation happens in the nucleotide sequence or DNA sequence, that will be carried forward to the protein. For example, due to mutations, oncogenes can be activated and the proteins from that gene may lead to carcinoma (Croce 2008). Likewise, if the tumor suppressor genes become mutated, corresponding proteins will not be produced, as a result, cancer might be developed (Sherr 2004). Realizing the extremely important role of mutations in causing many human disorders including carcinomas, researchers already started working on it using advance scientific facilities like genome sequencing (Greenman and Bignell 2007) (Jones 2009).

To function properly, each cell needs to depend on thousands of proteins to perform its functions properly. Upon modification in the genetic information, when a mutated protein is being created, it disrupts the normal development and thus it causes the disordered condition. We wanted to have a broader view of how mutations can affect the function of a protein. The function of a protein can be affected by its structure, its stability, and its motion. We have studied the effect of the mutation using zinc finger

as a model. Our laboratory was already interested in understanding protein-protein interaction related to chromatin signaling and epigenetic regulations. Therefore, I continued my research on the PHD finger which is also a type of zinc finger, and started probing the effect of mutations in their functions. We have characterized the xCDxCDx PHD finger by systematically probing every possibility of getting affected by the mutation one by one.

Chromatin structure is regulated by some regulatory protein complexes that bind to histone tails and these protein complexes are known to control over chromatin-signaling pathway. Thus it controls over all the possible diverse biological effects that are associated with chromatin signaling (Smith 2010, Allis 2016). It's important to note that genetic disorders are mainly caused by mutations that make a gene improperly functional. For example, while someone is known to have "the cystic fibrosis gene," it means that the person is carrying a mutated version of CFTR gene, that causes the disease. All people, including those without cystic fibrosis, have a wild type CFTR gene (Moskowitz MS 2008).

Proteins or peptides that bind with the histone peptide, have negatively charged amino acids. Since histone proteins are highly positively charged, therefore there is a possibility of very tight packing. The packing is so tight, even if there is any removal of a tiny methyl group happens, the binding may be compromised. We found that mutating threonine amino acid may lead to a complete disruption of the interface binding. In some cases, we found that even though Lysine is a bulky amino acid, but still does not disrupt the binding. But threonine has CH_3 group and OH group, so generally, it can be assumed that it is the OH group which is mainly responsible for binding. Parallelly CH_3 group can also play an important role in binding. To confirm this, we have mutated threonine to serine to show that even this tiny CH_3 group can

disrupt the tight binding. In our lab's previous work, we had characterized a subfamily of histone binding proteins. There also we found that tight packing is one of the main observations. We confirmed it by the complete disruption caused by the mutation of Alanine to Glycine where CH_3 group is lacking. It also strengthens the significant effect of methyl group in tight packing. We used peptide that was threonine to serine in which OH is there, but methyl group is absent. We purified wild type and mutant proteins and performed ITC with this peptide. We found that the methyl group also can disrupt the binding. So if we want to design diagnostic reagents for histone peptides like an antibody, we can do it in such a way that negatively charged amino acids can be replaced and tight packing can be observed and even these amino acids can be engaged with tight packing and get tight binding. Therefore, in a nutshell, tight packing is responsible for histone binding, even if they don't have tryptophan, tyrosine, and phenylalanine because of charges and where there are tiny amino acids because of the tight packing packed so well, so if we remove them, it will create empty space and the complex doesn't stay very well. In previous work, it has been found that there is a relationship between two kinds of amino acids, negatively charged and non-polar amino acids. Negatively charged amino acids are related to electrostatics and non-polar amino acids are related to tight packing. Co-relationship based on the nucleotide sequences between these two kinds of amino acids indicates that PHD finger is a huge superfamily. PHD finger superfamily is well known for its site-specific histone readouts. PHD finger consists of 50-80 amino acids and usually found in the nucleus of more than 100 human proteins. They are highly important because they are involved in chromatin-mediated gene regulation. Through the ages, it has been believed that a protein folds when a hydrophobic driving force leads the non-polar amino acids to the core inside of the protein. Real folding or stability comes from the core structure of the protein which is

made up of the non-polar amino acids. In case of PHD fingers, they have zinc atoms held by cysteine, and histidine residues, playing the central role of nucleating agent and folding properly. Therefore, zinc fingers turn out quite stable and sturdy protein.

After a detail introduction of my research work in chapter I, in chapter II, I will be discussing the preparation of all the reagents that have been used for the series of experiments addressing different types of questions regarding structural, folding, motion, and stability of the PHD finger. In chapter III and IV, I will be discussing about how to prob the structural and melting properties of the PHD finger using NMR study and whether there is any effect of mutations in the structure of the protein or mutations can only affect the binding ability of the protein. We also looked into weak anion-quadrupole interaction and its impact on human disorders. To carry forward this experiment, we have started working with QM protein interacting with PDZ domain and we have a series of proteins in the list to be studied.

CHAPTER II

Preparation of reagents for determining the structural stability of the mutations in the PHD finger

Abstract

The preparation of the reagents for biochemical and biophysical experiments is extremely important. Further study completely depends on the quality and the purity of the reagents. Because we study structural biology, the quality of the sample proteins should be very pure, otherwise, it may intervene in the results while we perform NMR-spectra. It is equally important for the CD-spectra and other experiments as well. To understand the underlying mechanism of how nature has beautifully crafted and orchestrated the mechanism, characterization of the protein molecules is highly necessary. There are different ways of how a protein should be characterized for biochemical and biophysical studies. To characterize the proteins, we have performed a series of experiments, out of which ITC is one of the main experiments that is extremely important for the thermodynamic studies. Through ITC, we have measured the amount of released heat when there was a binding noticed between the protein and the ligand. Since mutational analysis holds major importance in our study, therefore we needed to perform the comparative study of the binding patterns of the mutant and the wild type proteins. One of our major concerns was to understand the stability and the folding pattern of the sturdy and robust protein like PHD finger that evolved a special type of sequence called xCDxCDx motif through the course of time. Through NMR-HSQC spectra and CD-spectra, we made sure that mutations cannot impact the folding pattern of these proteins. They are equally well folded even after having mutations. We have also compared the thermal stability of the wild type and mutant type proteins through FASTpp and we didn't find any remarkable differences in their behavior. We

have also performed a melting comparison study and found both wild as well as mutant proteins behaving the same way. Through all of our experiments, we wanted to make sure whether mutations are playing a significant role only for substrate binding, or they are also equally important for protein folding, and stability. We have successfully probed that in this case of PHD finger (that has xCDxCDx motif), the mutation can only affect the substrate binding and not the folding or stability of the proteins. We also proved that due to the course of time this particular motif (xCDxCDx) makes the PHD finger more robust and stable.

Introduction

The structure of life is strongly related to proteomics that mainly can be explained by the structural and functional study of the protein. Understanding the protein stability is essential for determining the structure and the folding pattern of the protein. To characterize a protein, we need to perform numbers of biochemical as well as biophysical experiments, for all of the experiments, we need to have a pure form of protein and peptide. Towards getting the pure form of protein, there are some very important steps. Such as making recombinant proteins is a major step in this field. We can manipulate the construct of the recombinant DNA based on which it would be easier to purify the protein and we can get a good yield of the protein as well, which we will use later for different kinds of experiments like ITC, FASTpp, NMR-HSQC, and CD experiments respectively. There are several factors that can affect the structure and the stability of the protein. The amino acid sequence of the protein determines the structural stability and the folding ability. Entropy may play a role by reducing the flexibility (cystine bridges, and by increasing proline content), or by releasing water from the residues buried upon folding and association (RainerJaenicke 2000). Sometimes proteins need a conjugation with small molecule or metal to get properly

folded following which it can become biologically functional (Murata 2013). These conditions may affect the conformational distribution as well as the stability of the protein. Without having the conjugated form, proteins may lose its structure as well as function. Therefore, according to the need of the study we have made the recombinant proteins and respective mutations. Here we have studied structural stability of the PHD finger which is also a type of zinc finger. Zinc fingers are typically small proteins with 60-70 amino acids. Along with the pattern of cysteine residues, we also observed a couple of negatively charged amino acids in the motif of xCDxCDx PHD finger. This motif helps us to identify a subgroup. But in this subgroup, there are several positions that are correlated to the aspartic amino acids (Ds). Among millions of protein-protein interactions (PPI) that happen in a living cell, we wanted to look at the PPI in the nucleus, specifically on the chromosomes. Among them, we are particularly interested in proteins that bind to histones because their nature of the interaction is having a direct impact on the genes. Therefore, we are interested in understanding the structural stability of the zinc finger that has this particular motif xCDxCDx. Understanding this phenomenon would help us to recommend a sturdy, and stable zinc finger that can be used as a scaffold to deliver drugs or other proteins inside the cell. There are several ways of probing the structure and stability of the protein. We used FASTpp in different conditions to check the stability of the PHD finger and its mutants. We also checked the thermal stability of the PHD finger and its mutants by Circular Dichroism (CD) at different temperatures significantly higher than 65°C as NMR experiments were limited up to 65°C for structural determination. To determine the biophysical stability of the PHD finger and its mutants, we have performed ITC. It helped us to measure the bioenergetics especially the heat that is released while protein binds with the peptide.

Materials and methods

DNA Constructs:

DNA of hUHRF1 wild type was cloned into Ampicilin resistant pGEX4T3 vector. Then M435A, and V365A, and D30A mutations were created in the hUHRF1 by site-directed mutagenesis. L326A, D331A, and V346A mutations were created in the hKDM5B by the same method. KDM5B wild type construct was obtained from my previous lab mate.

Molecular Cloning:

Synthetic DNA of hUHRF1 PHD (residues 299-367 UniProt ID Q96T88/UHRF1_HUMAN) was obtained from Genscript and was amplified by PCR (polymerase chain reaction) (Table 1.1) using Thermo Scientific Phusion High Fidelity DNA Polymerase with the forward and reverse primers (IDT Technologies) (Table 1.3) using the thermal cycler (Table 1.2).

Table 1.1: PCR Reagents for hUHRF1 cloning. The reaction was set up in the ice to amplify the insert DNA and the chemicals were added in the following order:

Component.	Volume (In 50µl reaction)
Double distilled water	35.5µl
5X Phusion buffer	10µl
10mM dNTPs	1µl
Forward Primer	1µl
Reverse Primer	1µl
Template	1µl
Phusion DNA Polymerase	0.5µl

Table 1.2: Reaction cycle of PCR

Cycle steps	Temperature	Time	Cycles
Initial Denaturation	98°C	30Sec	1
Denaturation	98°C	10 Sec	
Annealing	62°C	30 Sec	30
Extension	72°C	1 Min 45Sec	
Final Extension	72°C	10 Min	1
Final Hold	4°C	∞	

Then the amplified DNA was run in the DNA gel electrophoresis for 25 minutes at 120mA voltage. DNA bands were purified from the DNA gel by gel extraction kit (QIAGEN). Double digestion reaction was performed right after gel purification. Reaction was set for two hours at 37°C with BamH1 and Xho1 restriction enzymes (New England Biolabs). After double digestion, the sample was purified by PCR purification kit (QIAGEN). Thus, gene of insert was prepared for ligation reaction. pGEX4T3 empty vector was purified by QIAprep spin miniprep kit (QIAGEN) from overnight grown *E. coli* culture (Ampicilin added LB culture). After that, it was double digested by BamH1, and Xho1 restriction endonucleases at 37°C for two hours. Then the sample was run in DNA gel and purified by gel extraction kit (QIAGEN). Once the insert and the vector were ready for ligation, based on their concentration (ng/µl), ligation reaction was set up according to 1:3 ratio of insert and vector for three hours at 16°C. After ligation, expected recombinant DNA was transformed (1:1 ratio) into high-efficiency chemically competent *E.coli* (XL10Gold) cells. It was incubated at 37°C for overnight (about 16 hours) in the Ampicilin resistant LB plate (Figure 1.7). Colonies were observed in the following morning. Individual colonies were picked up for colony

PCR to confirm the right insert of the clone. From the same colonies, cultures were grown for about 18 hours at 37°C in Ampicilin added LB media. DNA was extracted from each culture by QIAprep miniprep kit and sent for DNA sequencing to confirm the nucleotide sequence of the clone. After verification of the DNA sequence (from Functional biosciences), the recombinant plasmid was transformed into chemically efficient *E. coli* - BL21(DE3) cells and incubated overnight (about 16 hours) at 37°C. After getting colonies in the following morning, small scale expression test was done with the four individual colonies grown in the antibiotic added LB media and treated with and without IPTG to check whether the recombinant protein expressed or not, samples were then run in SDS PAGE gel for an hour at 120mA voltage. After the staining and de-staining process, the SDS PAGE gel picture was taken. From the gel picture, the most expressed colony was selected and continued for purification in a large scale (4liters).

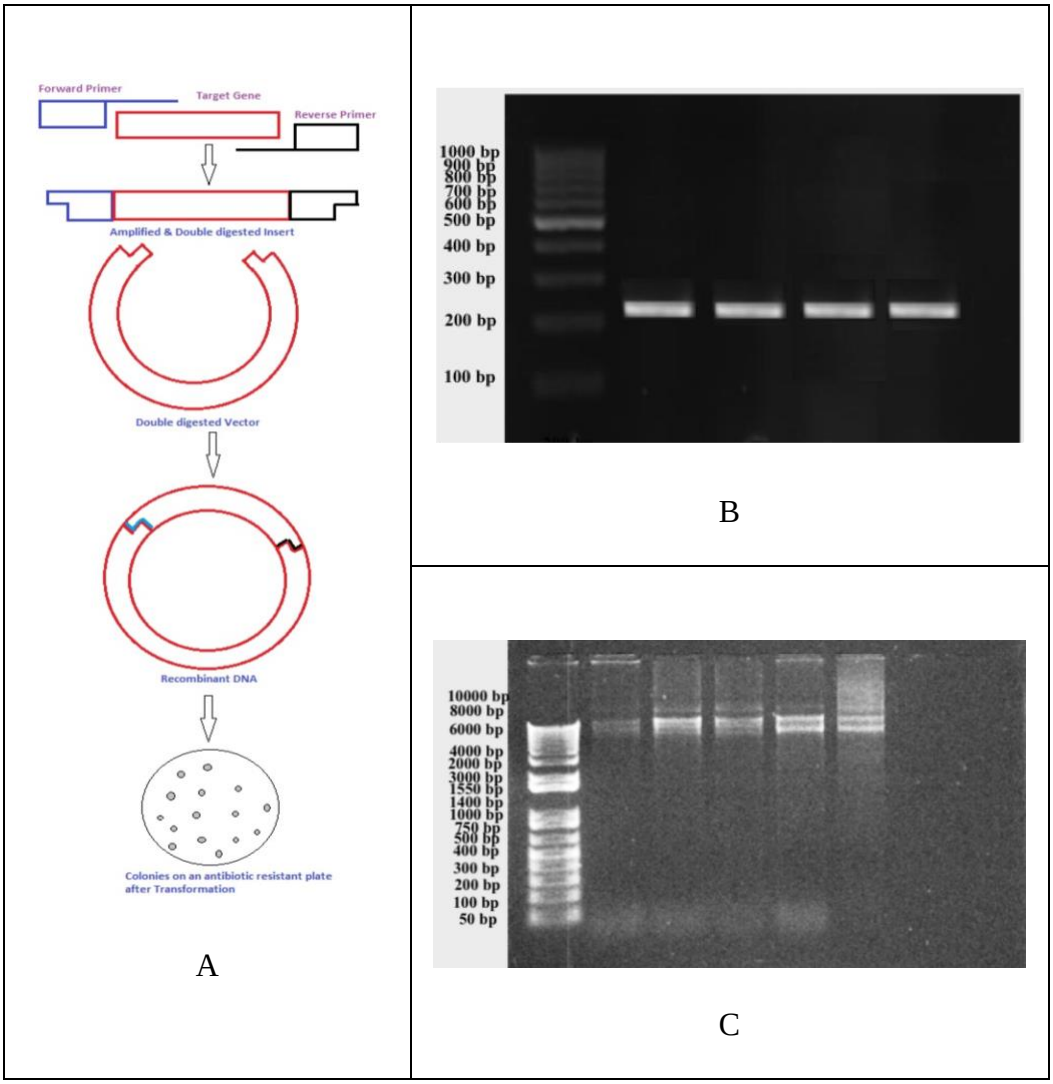


Figure 1.7: A. Design of the Cloning strategy of hUHRF1; B. DNA gel (0.8%) showing bands of insert (UHRF1) at 207bp; C. DNA gel (0.8%) showing bands of vector (pGEX4T3) at 4968bp.

Table 1.3: Nucleotide sequence of cDNA of the PHD finger protein and its related Primers.

	Nucleotide Sequence
hUHRF1	CGCGGATCCGGTCCGAGCTGCAAGCACTGCAAA GACGATGTGAACCGTCTGTGCCGTGTTTGCGCGT GCCACCTGTGCGGTGGCCGTCAGGACCCGGATA AGCAACTGATGTGCGACGAGTGCGATATGGCGT TCCACATCTATTGCCTGGACCCGCCGCTGAGCAG CGTGCCGAGCGAGGATGAATGGTATTGCCCGGA ATGCCGTAACGACGCGTAACTCGAGCGG **Protein Sequence** GPSCKHCKDDVNRLCRVCACHLCGGRQDPDKQL MCDECDMAFHIYCLDPPLSSVPSEDEWYCPECRND A
Primer	Primer Sequence
Forward Primer (5’ to 3’)	GACCCGCCGCTGAGCAGCGTGCCGAGCGAGGAT GAATGG
Reverse Primer (5’to 3’)	CCATTCATCCTCGCTCGGCACGCTGCTCAGCGGC GGGTC

Transformation:

After ligation, the ligated product (1µl) was transformed into *E. coli* XL10Gold competent cells. To confirm the success of cloning, we have performed colony PCR with the colonies that we got from this transformation. We have cultured (overnight) each colony in the LB media along with Ampicilin and collected the DNA samples through Miniprep (Qiagen). Then we sent the DNA samples for sequencing. After DNA sequencing, the second transformation has been done in BL21 cells of *E.coli*, other reagents required for transformation, is as follows:

LB-agar plate with antibiotics (Ampicilin for pGEX4T3 vector and Kanamycin for pET28A vector), sterile SOC media, and 42°C water bath.

Following are the steps of Transformation:

• 30µl BL21 competent cells of *E.coli* were taken out from -80 °C freezer and was put on ice for 10min.

• 1-5µL of DNA was added to the cells and mixed gently.

• The tube was then incubated on ice for 30 minutes.

• The SOC media was Pre-warmed at 42°C.

• The cells were heat-shocked for 45 seconds in a 42°C water bath without shaking.

• The tube was then placed on ice for 2 minutes.

• 300µL of SOC was added to the cells and incubated at 37°C for 1 hour at 225rpm.

• The Kanamycin/Ampicilin agar plate was pre-warmed in the incubator at 37°C.

• 100µL of cells was inoculated and incubated at 37°C for overnight (12-16hours).

Protein Expression and Solubility test:

After verification of the sequence from the functional biosciences, the recombinant DNA of UHRF1-PHD was transformed into BL21 (DE3) *E.coli* competent cells. Four colonies on the Ampicilin agar plate were randomly selected for protein expression and solubility tests. These colonies were then inoculated into 4ml LB media and started culturing at 225rpm at 37°C for few hours (usually 5-6 hours) till it reached to 0.2-0.4 optical density at 600nm (OD600). Half of each culture (2ml) was then transferred to a new falcon tube and added 2µl 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and let them to shack for another 4 hours in the 225rpm incubator shaker at 37°C. After that, all the induced and uninduced samples were run into SDS PAGE at 120V for an hour followed by staining and de-staining process (Figure 1.8). After seeing the bands of expression in the expected range, I have performed the following steps for solubility test:

The cell cultures were pelleted in pre-weighted Eppendorf tubes by centrifuging at 13000rpm for 10mins. 0.5g of the pelleted cells were completely resuspended in a 1ml solution of 1x bug buster. The suspension was incubated on a shaker at 4°C for 20mins and then centrifuged at 13000rpm for 10mins. The supernatant, which now contains the soluble protein, was separated from the pellet and run in the SDS-PAGE gel to determine the solubility of the protein.

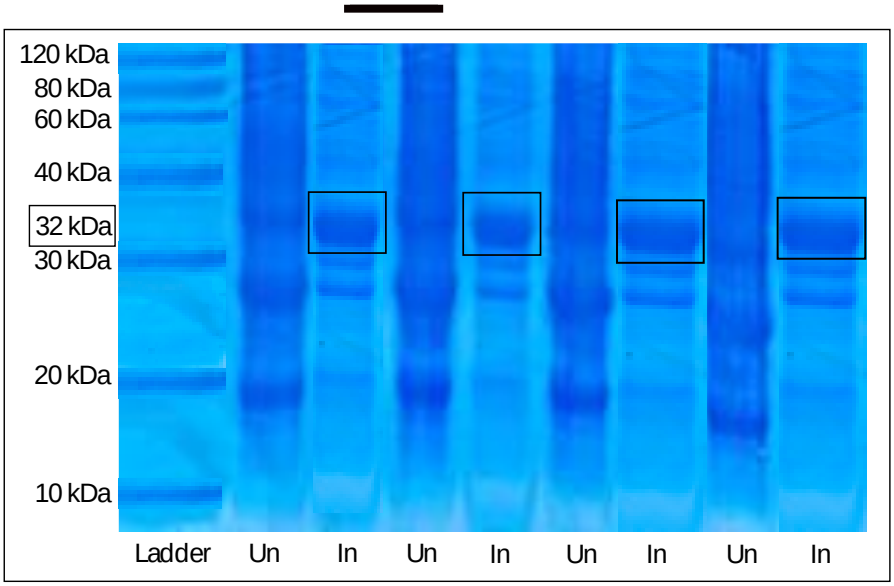


Figure 1.8: SDS PAGE of the Expression of the recombinant protein (UHRF1wt-GST) in the lane of induced (In) at 32kDa.

Site-directed mutagenesis

I have created mutations in the PHD finger (hUHRF1) using the mutated primers (Table 1.6), following the site-directed mutagenesis (SDM) protocol (Table 1.5) (Figure 2.9) and following the programming explained in Table 1.4. I got the mutants of hKDM5B from my past lab mate. SDM is a method of creating changes in the double-stranded plasmid DNA leading into the altered amino acid sequence in the protein.

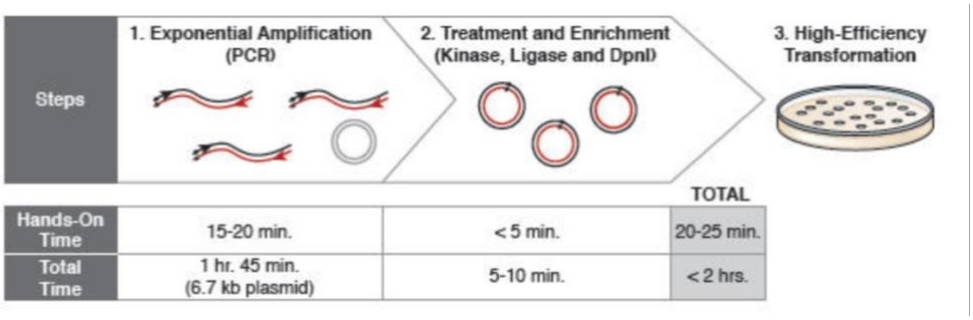


Figure 1.9: Steps of Site directed mutagenesis (SDM) (Kunkel 1985)

We have considered the positions of mutants based on their accessible surface area with the peptide, ΔASA_{SC}, $\geq 10 Å^2$ of the side chain atoms because we wanted to determine the binding interaction of the PHD fingers wild type as well as its mutants with the histone H3 peptide as most of the PHD fingers we looked at it were playing the role of histone reader and directly contributing in chromatin regulation (Sanchez 2011). We have considered negatively charged distal amino acids because we wanted to understand how they are contributing to the substrate binding.

Table 1.4: Reaction cycle of Mutagenesis. It took 1hour 45minutes to complete the entire reaction.

Cycle step	**Temperature**	**Time**	**Cycles**
Initial Denaturation	98 °C	30 sec	1
Denaturation	98 °C	10 sec	1
Annealing	Varying	30 sec	18
Extension	72 °C	1 min 30 sec	
Final Extension	72 °C	10 min	1
Final Hold	4°C	∞	

Table 1.5: Reagents for Site Directed Mutagenesis: The reaction was set up in ice and the reagents were added in the following order:

Component	**Volume in 50µL Reaction/µL**
Distilled water	35.5
5X Phusion buffer	10
10mM dNTPs	1
Forward Primer	1
Reverse Primer	1
Template DNA (pGEX4T3_UHRF1-PHD)	1
Phusion DNA Polymerase	0.5

Table 1.6: Protein sequence of the mutations created on hUHRF1 PHD finger and corresponding primers:

Mutant	Forward Primer (5'- 3')	Reverse Primer (5'-3')	Protein Sequence
D350A	ATGTGCGACGAGTGCGCTAT GGCGTTCCAC ATC	GATGTGGAACGCCATAGC GC ACTCGTCGCAC AT	RGSGPSCKHCKDDVNRLCRVCAC HLCGGRQDPDKQLMCDECAMAF HIYCLDPPLSSVPSEDEWYCPECR NDA
V365A	GACCCGCCGCTGAGCAGCGC GCCGAGCGAGGATGAATGG	CCATTCATCCTCGCTCGG CGCGCTGCTCAGCGGCGG GTC	RGSGPSCKHCKDDVNRLCRVCAC HLCGGRQDPDKQLMCDECDMAF HIYCLDPPLSSAPSEDEWYCPECR NDA
M345A	GACCCGGATA AGCAACTGGC GTGCGACGAG TGCGATATG	CATATCGCACTCGTCGCA CGCCAGTTGCTTATCCGG GTC	RGSGPSCKHCKDDVNRLCRVCAC HLCGGRQDPDKQLACDECDMAF HIYCLDPPLSSVPSEDEWYCPECR NDA

After mutagenesis PCR reaction, Dpn I digestion was performed at 37°C for 2 hours to destroy the template (hUHRF1-PHD) wild type DNA. Dpn I digestion of the parental supercoil dsDNA was followed by enzyme inactivation at 80°C for 20mins. The mutants (Dpn I digested PCR products) were transformed into XL 10-gold competent cells and then incubated on a warm Ampicilin agar plate at 37°C for overnight (14-16hours). Colonies were obtained on the Ampicilin agar plates in the following morning. These colonies were cultured and allowed to grow at 37°C for 18hours in an Ampicilin /LB media. DNA samples were extracted from the overnight grown culture using miniprep kit (QIAGEN). To determine the presence of the mutants, DNA samples were run in 1% DNA agarose gel in 1x TAE buffer (pH 8.0). The samples were then sent for DNA sequencing (Functional Biosciences) for determining the right nucleotide sequence for each of the mutants.

Transformation, Expression, and Purification of Mutants

After getting the sequencing results and confirming the right DNA sequences for all the mutants, each of them was transformed into HI-Control® BL21 (DE3) Chemically Competent cells (see transformation). Several colonies were observed in the following morning. To select the best colony of each of the mutant, we have performed solubility test followed by small-scale expression test. 4-liter culture was prepared for each mutant and purified as described above (see protein purification of the wild type protein).

TB media preparation:

For preparing TB media, 1000ml distilled H_2O was added to the following: 50.8g Terrific broth powder, and 4ml Glycerol and sterilized by autoclaving.

M9 media preparation:

For preparing ^{15}N labeled protein, we made M9 minimal media supplemented with $^{15}NH_4CL$. First of all, M9 salts were prepared in 800ml H_2O, and the following chemicals were added: 60g Na_2HPO_4-$7H_2O$, 5g NaCl, 30g KH_2PO_4, and 5.0g $^{15}NH_4Cl$ and Stirred until get completely dissolved. The total volume of the solution was adjusted to 1000ml with distilled H_2O and sterilized by autoclaving. Then 700ml sterilized distilled H_2O was measured and 200ml of M9 salts were added with it along with other trace elements like 2ml of 1M $CaCl_2$ (sterile), 2ml 1M $ZnCl_2$, 2ml 1M MgSO4 (sterile), 10ml 100xMEM vitamins (or 1.2ml 5% Thiamine) and 20 ml of 20% glucose and adjusted to 1000ml with distilled H_2O.

Protein purification:

After the determination of small-scale expression and solubility of the protein, 4liter cultures were prepared from overnight grown 30ml cells (culture in LB/Ampicilin media overnight at 37°C, 225rpm, LB media was discarded and palleted cells were added only). We prepared 4liter M9 media for the proteins intended for NMR study. Cells for other experiments (FASTpp, CD, ITC) were grown in 4liter TB/LB media along with Ampicilin/Kanamycin antibiotic respectively as required based on the plasmid vector. After inoculation, cells were incubated at 37°C for 8-9 hours at 225rpm until OD_{600} reached to 0.7-0.8. 1M IPTG (isopropyl-β-D-thiogalactopyranoside) (500µl) was added to the culture to induce the cells to express the proteins. After induction, cells were transferred to 16°C shaker incubator (N-Biotek) for 18 hours at a rotating speed of 225rpm. Cells were then collected by centrifuging at 8000rpm for 45 minutes at 4°C. To make ^{15}N labeled protein, $^{15}NH_4Cl$ was added to the M9 media. This media was supplemented with $MgSO_4$, $ZnCl_2$, MEM vitamin, and 20% glucose.

We have lysed the overexpressed cells with 10mg/ml of lysozyme in 1x phosphate buffer solution (25mM phosphate buffer, 250mM NaCl at pH 7.6). Lysozyme helps in breaking the cell wall of the bacteria and thus helps in releasing the soluble protein in the buffer. The cells were also being treated with 0.1mg/ml of IGEPAL- CA 600 (a nonionic, non-denaturing detergent that prevents the surface-induced protein from aggregating in the buffer). We made sure that the number of bacteria cells in the resuspension phosphate buffer should be around 30%. Resuspended cells were ice incubated for 30 minutes before I started processing through the Misonix ultrasonic liquid processor (Sonicator S-4000) for 5 minutes at 10 seconds and 20 seconds pause, that further made sure that the cells are properly lysed. After sonication, lysates were cleared through centrifugation (18000rpm for 30min at 4°C, Beckman J2-MI centrifuge). Then the supernatant (containing the soluble protein) was loaded on a phosphate buffer equilibrated glutathione agarose beads (Thermo scientific) and let it rock overnight, so that the GST tagged protein can bind with the glutathione beads. Following morning, beads with the protein were washed 3 times with 1x PBS followed by 3 times wash with 1M NaCl and then 3 times wash with 1x PBS to get rid of any unbound proteins. Finally, proteins were eluted by the elution buffer (10-20mM glutathione reductase solution) at pH 7.6 and ran through size exclusion chromatography where we got a peak of Protein of interest + GST. We collected the samples from tubes of that peak and concentrated it down and added Human thrombin protease (9mg/ml, HTI) for incubating at 4ºC for overnight to cleave the peptide bond between the protein of interest and the GST, leaving the GST with the glutathione beads. Finally passed it through FPLC (AKTA prime plus using HiLoad 26/600 superdex column from GE healthcare) (Figure 2.0) and got purified form of protein of interest. We have performed SDS PAGE (Bio Rad) to confirm the quality of the purified

protein (Figure 2.7). To purify the GST tagged protein in the quickest way, we have performed column digestion method of purification. In this method instead of eluting out the protein with the 20mM Glutathione reductase solution, we have applied Thrombin protease to cleave the bond between the GST protein and our protein of interest in the column, as a result, GST was left behind with the glutathione beads and we got our protein after running through the FPLC. We have collected our protein of interest and concentrated it down to ~2mg/ml and stored it to -80ºC immediately after freezing it with the help of liquid nitrogen (Figure 2.1-2.7). to confirm the purity and the quality of the protein, we have performed SDS PAGE gel (Figure 2.8- 2.9).

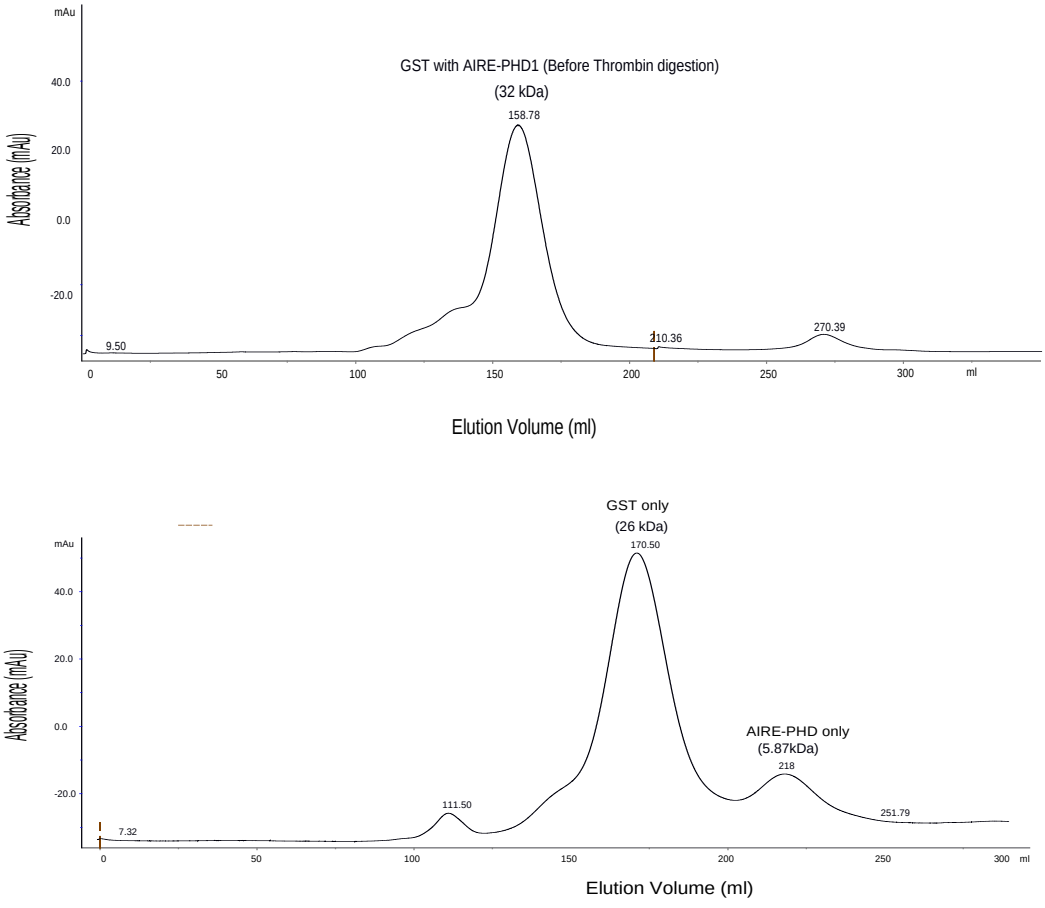


Figure 2.0: Chromatogram of GST-AIRE and only AIRE (without GST) showing FPLC profile: Protein Purification through traditional method (Through elution and then thrombin digestion).

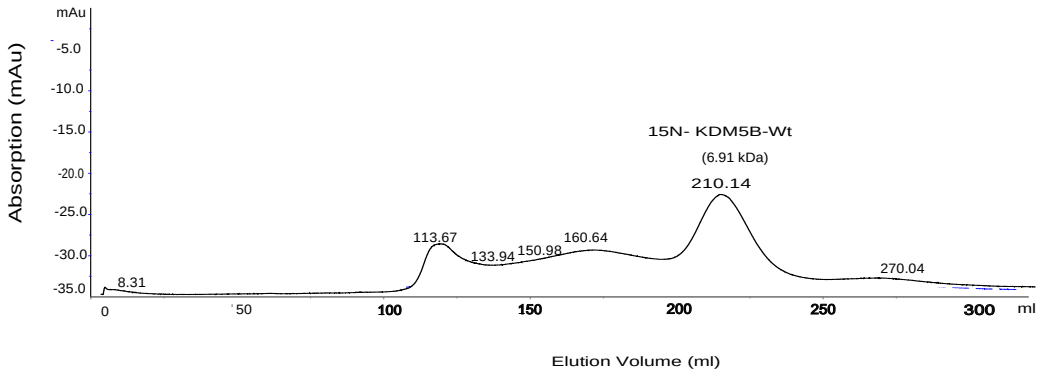


Figure 2.1: Chromatogram of the $_{15}$N-KDM5B-Wt showing FPLC profile: Protein purification through column digestion method.

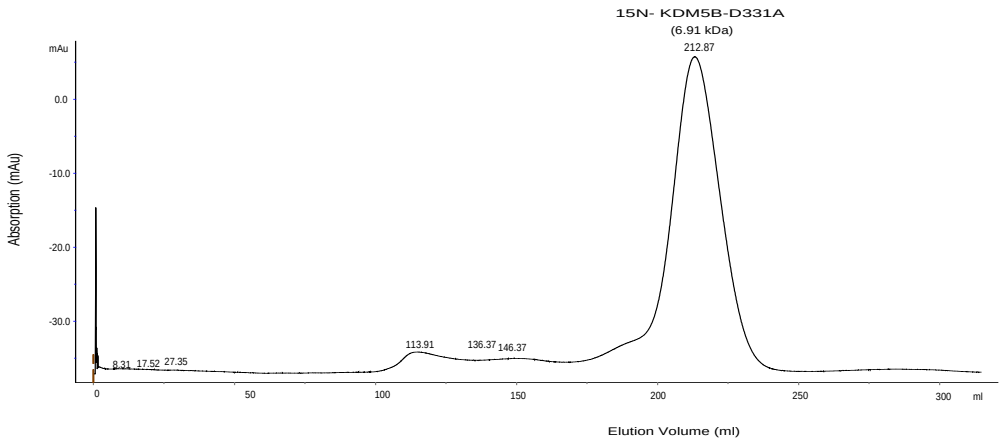


Figure 2.2: Chromatogram of the $_{15}$N-KDM5B-D331A showing FPLC profile: purified by column digestion method.

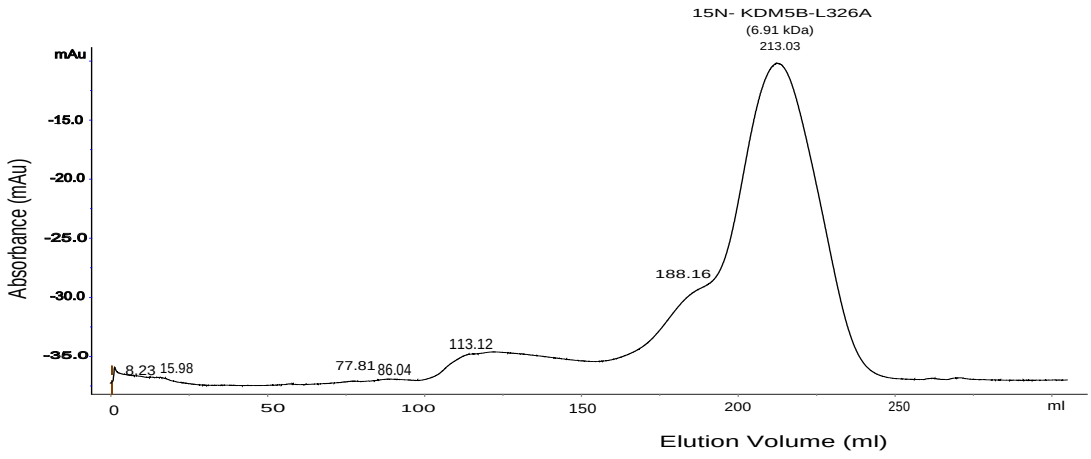


Figure 2.3: Chromatogram of the $_{15}$N-KDM5B-L326A showing FPLC profile: purified by column digestion method.

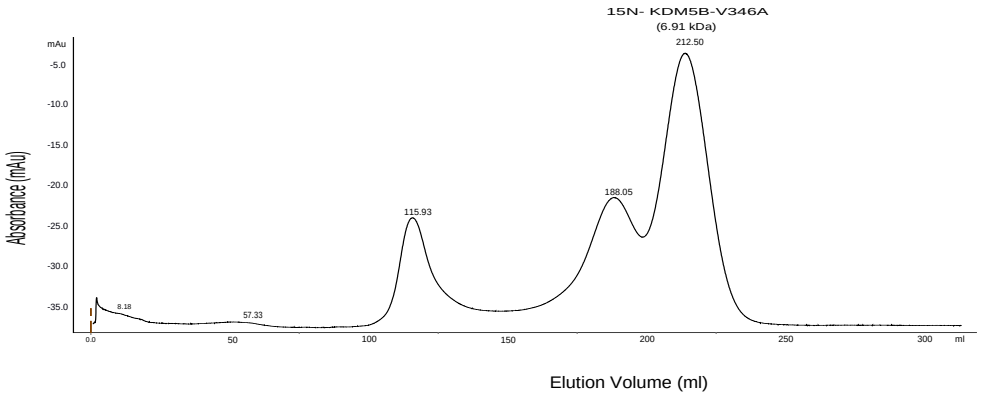


Figure 2.4: Chromatograms of the $_{15}$N-KDM5B-V346A showing FPLC profile: purified by column digestion method.

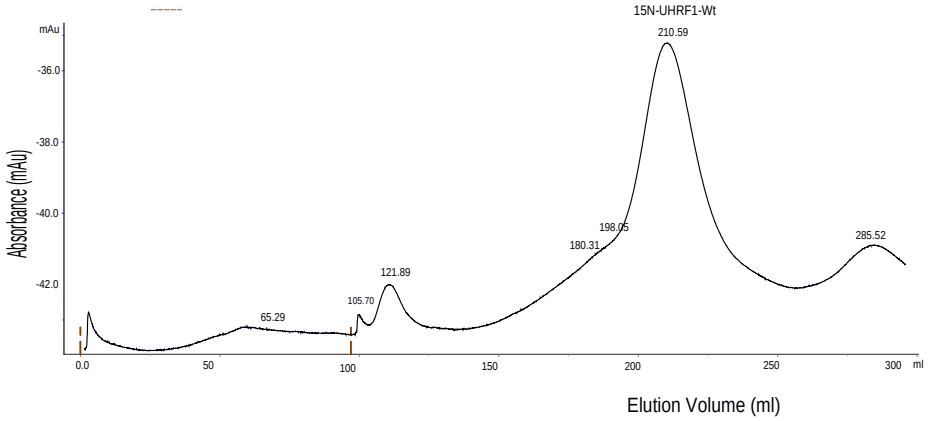


Figure 2.5: Chromatogram of the $_{15}$N-UHRF1-Wt showing FPLC profile: purified by column digestion method.

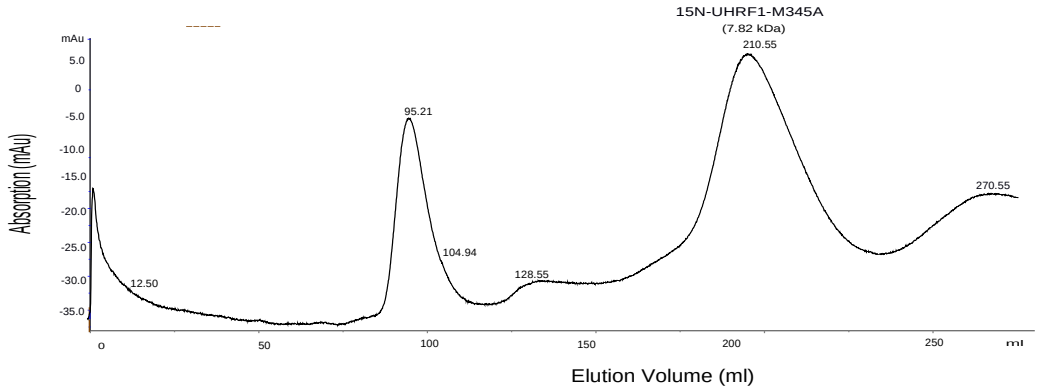


Figure 2.6: Chromatogram of the $_{15}$N-UHRF1-M345A showing FPLC profile: purified by column digestion method.

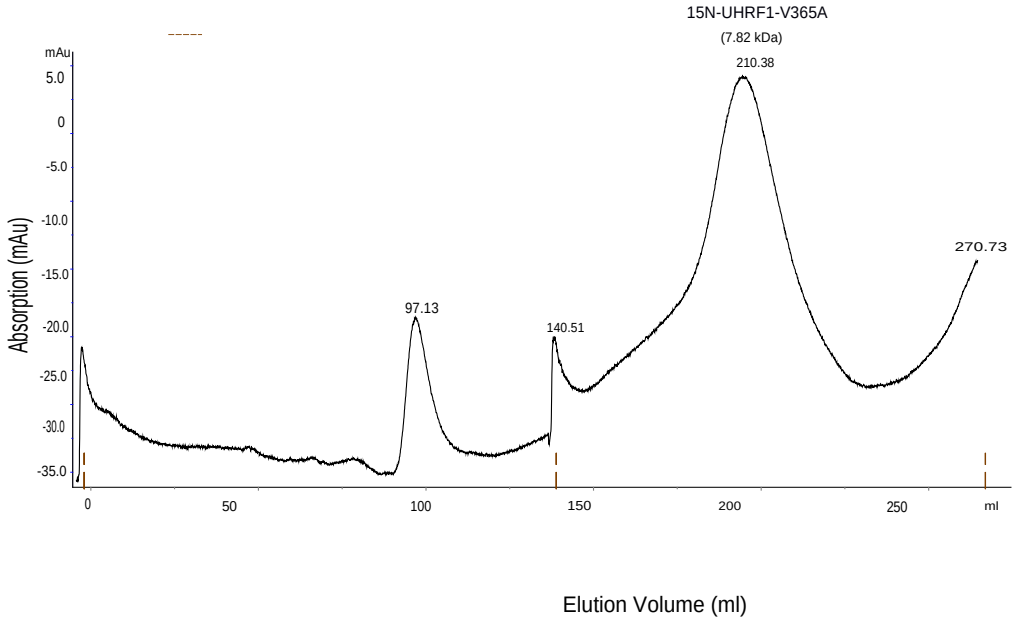


Figure 2.7: Chromatograms of the $_{15}$N-UHRF1-V365A showing FPLC profile: purified by column digestion method.

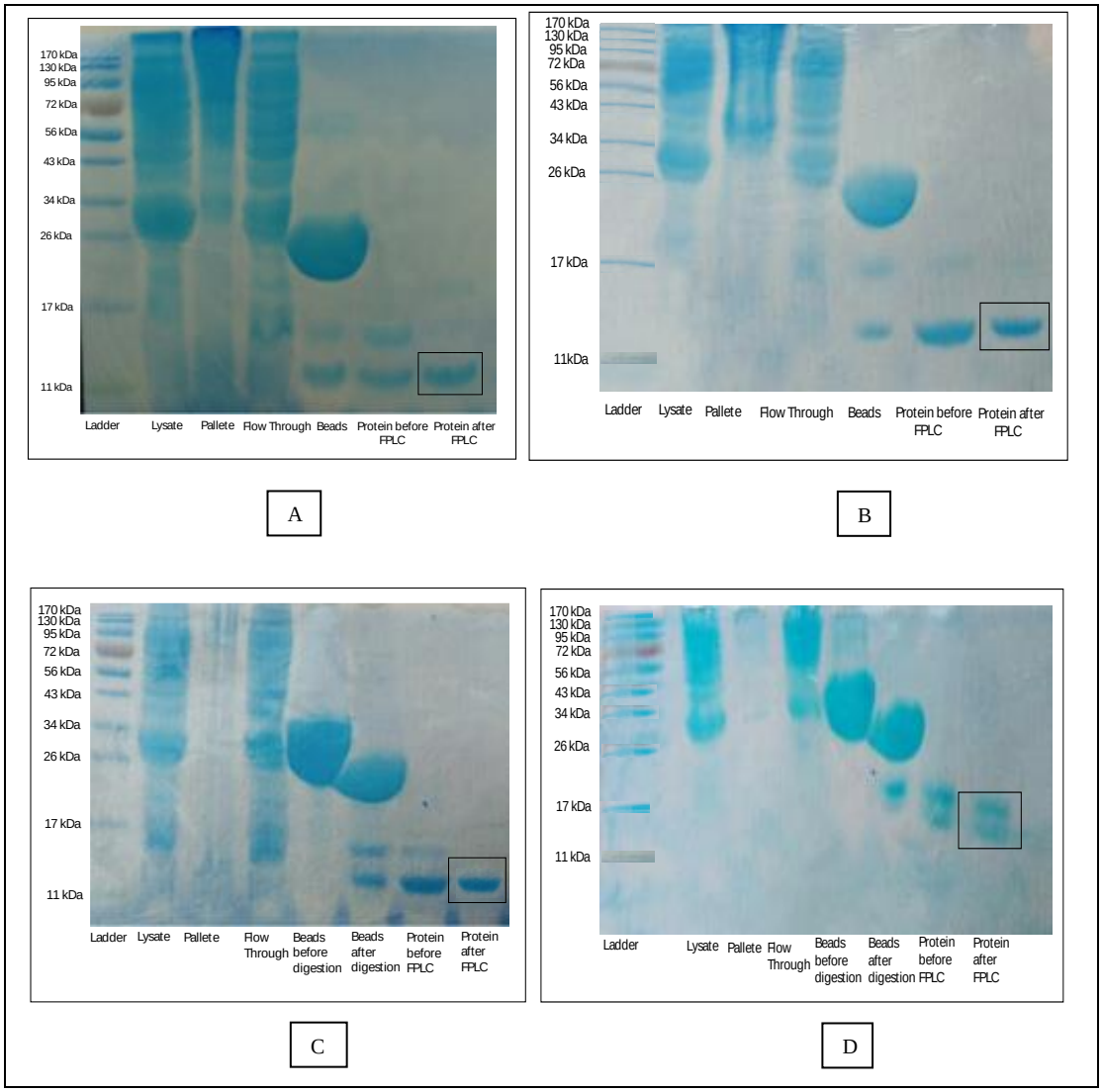


Figure 2.8: SDS PAGE of ^{15}N labelled KDM5Bwt protein purification. A. Purification of KDM5B wild type. B. Purification of KDM5B D331A. C. Purification of KDM5B L346A. D. Purification of KDM5B V346A.

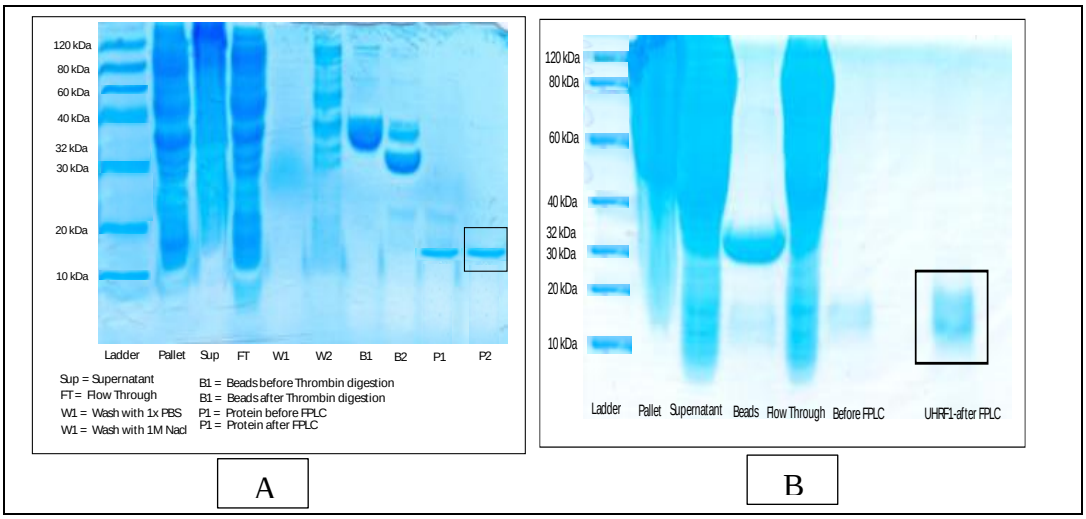


Figure 2.9: SDS PAGE with the samples of $_{15}N$ labelled UHRF1wt protein purification. A. Protein (wild type UHRF1) purification in traditional method by elution first and then thrombin digestion. B. Protein (Mutant type UHRF1) purification by Column digestion method.

NMR experiments:

For performing 2D-NMR-HSQC spectra, we have purified the protein following the method mentioned above. M9 minimal media supplemented with $_{15}NH_4Cl$ along with D-glucose-$_{13}C_6$-$2H_7$ in 80% H_2O was used to produce $_{15}N$ labeled protein (0.5mM and 100µl) for nuclear magnetic resonance experiments for determining the molecular structure of the protein. All the experiments were done at 300 K on Varian Inova 800 or 600 MHz NMR spectrometers equipped with triple resonance cryogenic probes along with z-axis pulse field gradients. Backbone assignments were based on TROSY versions of HSQC, HNCA, HN(CO)CA, HNCAB, HNCO, HN(CA)CO, a 1H/1H NOESY-HSQC (τmix = 100 ms) as well as a non-TROSY version of a nitrogen-edited HSQC-NOESY-HSQC (τmix = 100 ms) (Larry R. Masterson 2009).

Results and Discussions:

Folding study by NMR-HSQC spectra:

Information about the dynamic motion of the proteins at different period of time and at different length scale, is highly important to know more about the folding mechanism of the protein (Alexandrescu A. T. 1993, Schwartz M. P. 1999, Cavagnero S. 2001). Besides that, investigation about the conformational dynamics of the partially structured proteins may provide significant knowledge about the potential energy landscape of the interaction and it is extremely useful for better understanding the stability of the protein. It helps to rationalize the protein design (Dobson 1998). Understanding the molecular structure and dynamic motion of the protein is extremely useful for the researchers to design the drugs. From this aspect, NMR spectroscopy is a highly important and very quick method to probe the landscape of the protein folding because it gives a wide chance of studying the conformational dynamics of the native, partially folded, or unfolded states at a residual level (Wong K. B. 1996, Schulman B. A. 1997). Previous trend of research has shown that NMR-15N relaxation measurements have successfully probed the structural and dynamic characteristic features of the different states of the proteins (Bai Y. 1993, Chamberlain A. K. 1998, Li R. H. 1999, Wijesinha-Bettoni R. 2001). We have prepared all the reagents to perform a series of experiments mainly for addressing a few basic questions regarding the structural and functional aspects of the PHD finger. We also wanted to know whether and how mutations can influence the structure and function of the protein. We wanted to determine whether mutations are affecting the folding pattern of the PHD finger. From our previous study, we have found that mutations in the PHD finger significantly perturbs its binding ability and no longer it can bind effectively to the histone H3 peptide. To have a comparative overview of the molecular structure of a

special type of PHD finger (PHD_nW_DD) and to confirm the importance of a few specific amino acids that make this subtype different from other subtypes of the PHD finger, and how much they are contributing in the protein folding, we have first mutated those amino acids and then performed NMR-HSQC experiment in collaboration with NMRFAM of University of Wisconsin, Madison. With the purified ^{15}N labeled proteins, NMR-HSQC spectra were run at 800mHZ (Figure 3).

In the binding site of the PHD finger, there are few sets of amino acids that may play a crucial role in the binding of histone H3 peptide and especially found to be present in a very high frequency in a particular type of the PHD finger that contains xCDxCDx motif (PHD_nW_DD) whereas the frequency of those amino acids is quite negligible in another type of PHD finger (PHD_nW) that doesn't contain the xCDxCDx motif (Figure 1.3). We wanted to confirm whether these amino acids are only important for binding or they equally contribute to folding and stability of the PHD finger as well. Therefore, we have performed a series of experiments that help us to determine the structural stability and folding behaviors of the PHD_nW_DD.

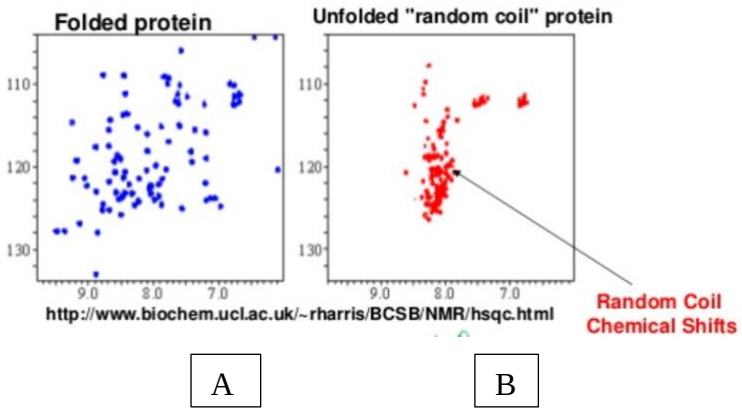


Figure 3: Standard chemical shift of folded and unfolded proteins

---

When protein folds properly, the chemical shift of the NMR-HSQC should spread out very well, so that each proton can be distinguished from another and the

distances can be measured as well as we can see in the figure 3A. Upon getting unfolded or misfolded, we can expect a chemical shift that appears like a clumsy random coil (Figure 3B).

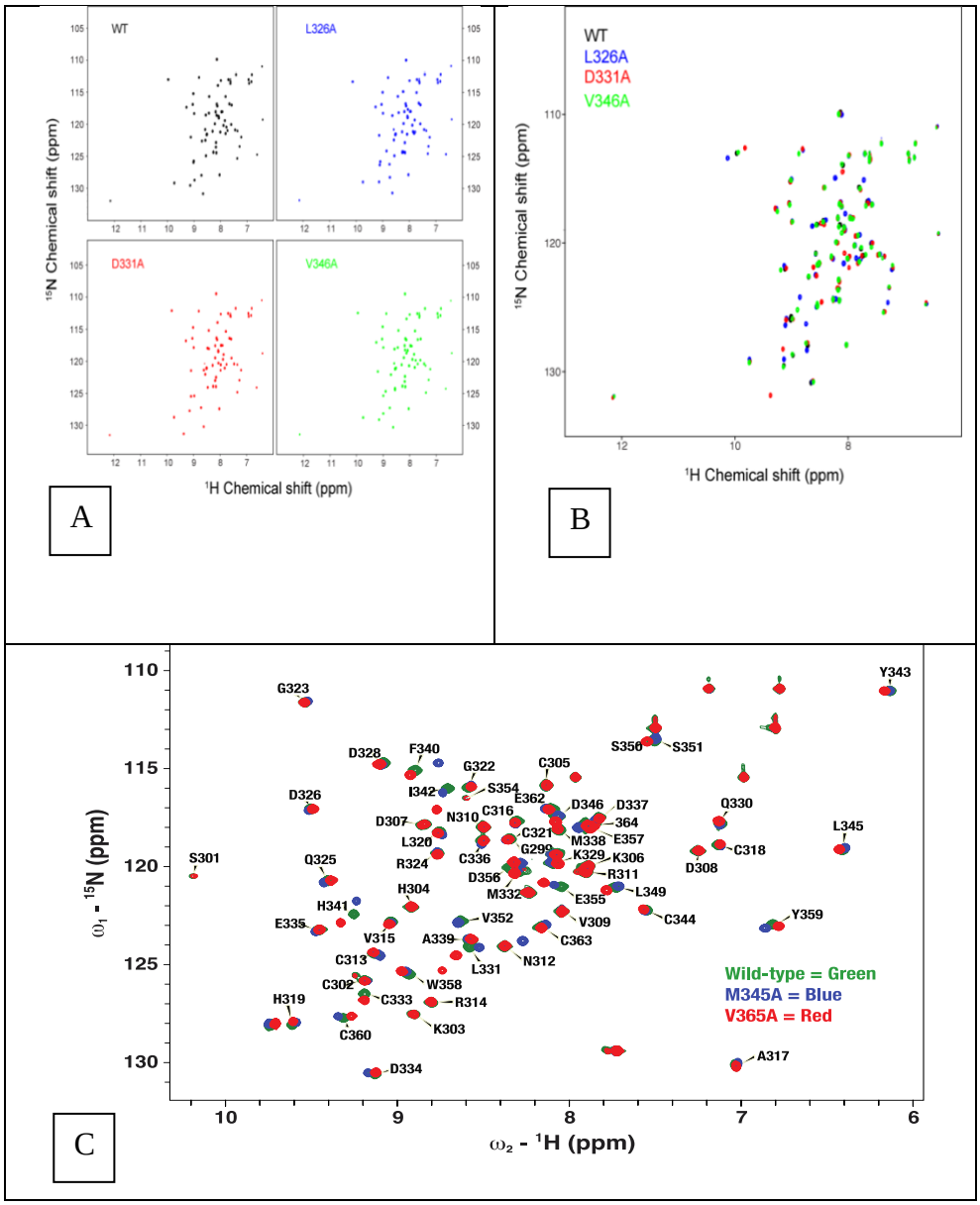


Figure 3.1: (A) NMR-HSQC spectra of KDM5B and its mutants (B) Overlapped picture of NMR-HSQC spectra of KDM5B and its mutants (C) Overlapped picture of NMR-HSQC spectra of UHRF1 and its mutants

From the above figures 3.1 of NMR-HSQC spectra, it has been found that all the peaks (representing each amino acid) are well dispersed and the chemical shift is well spread out as expected for properly folded protein (Figure 3a). Distances between the peaks of mutant types are exactly similar to the wild type. From the well-dispersed chemical shift, it can be concluded that mutated proteins (KDM5B- D331A, L325A, V346A; UHRF1- V365A, and M345A) are equally folded as the wild type does. That means, mutations have very less/ or no impact in their folding ability.

Protein dynamics study by NMR-T1, T2, and Het-NOE:

We have performed different types of relaxation experiments: spin-lattice (T1 or longitudinal relaxation that happens in the z direction) and spin-spin (T2 or transverse relaxation that happens in the x-y plane). T1 relaxation corresponds to the population distribution of energy by radiofrequency from α and β spin state. T2 relaxation corresponds to the decay of transverse magnetization. From our experiment, it has been found that the T2 is less than T1 because loss of magnetization occurred in the x-y plane while magnetization was returning to the z-direction. Besides this, Het-NMR is also a very powerful tool to investigate the intramolecular conformational dynamics (Farrow N. A. 1994, Orekhov V. Y. 1995, Dyson H. J. 1998). Through these NMR relaxation experiments, we have tried to quantify the backbone internal motion and also tried to determine the rate of overall molecular tumbling. Three relaxation parameters like T1, T2, and Het-NOE were being noted for each amide-bond vector (Figure 3.5).

In these experiments, it is highly necessary that the backbone amide protons are ^{15}N labelled proteins. Estimation of the side chain and the backbone dynamics through NMR depicts a unique perception about the motion of the protein (Dyson JH. 2004). This structural activity plays an important role in designing a potent inhibitor.

The subsequent binding activity can also be detected by this NMR relaxation experiment (Reddy 2010). Numbers of techniques are being suggested to determine the dynamics of the unfolded proteins. Among all of them, the most commonly used technique is NMR that includes T1, T2 relaxation experiment, and Het-NOE for studying the backbone resonances. To understand the principle of NMR-relaxation, it is highly recommended to consider the fully relaxed thermal equilibrium state. In this state, two conditions matter: (i) population of both α and β spin state and (ii) Loss of the phase coherence among the nuclei.

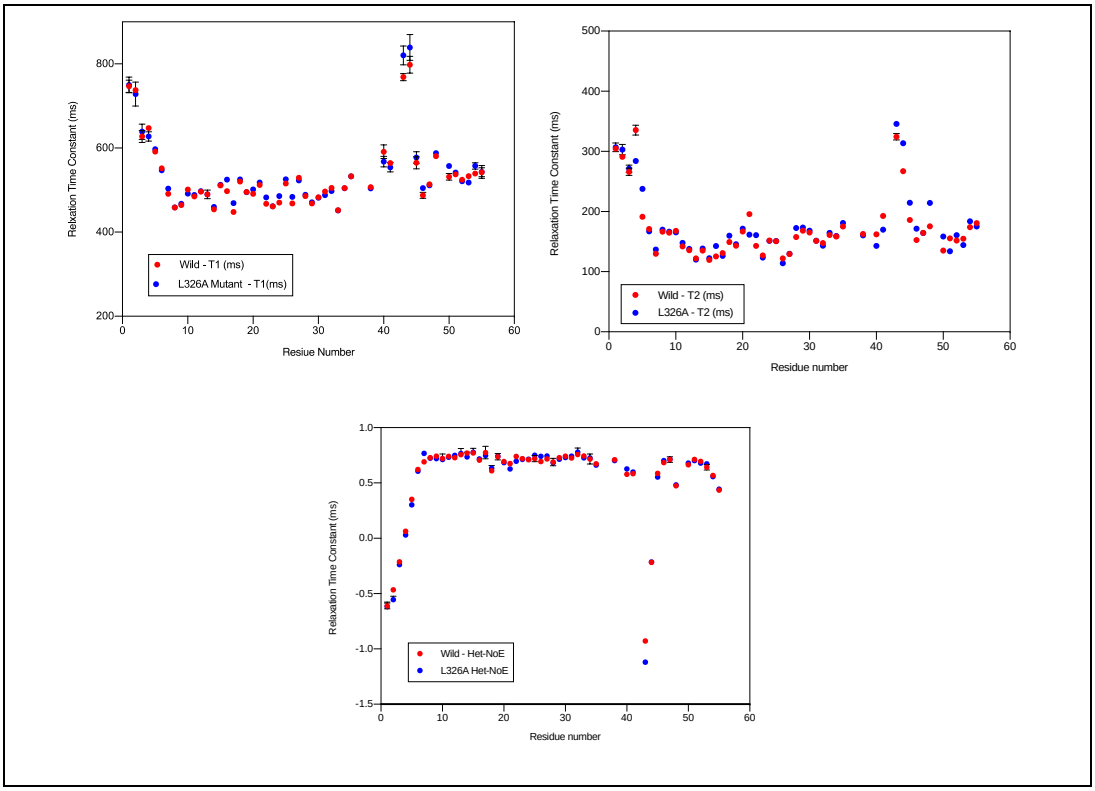


Figure 3.1a: T1, T2 & HetNOE of KDM5B wild type and its L2 mutant

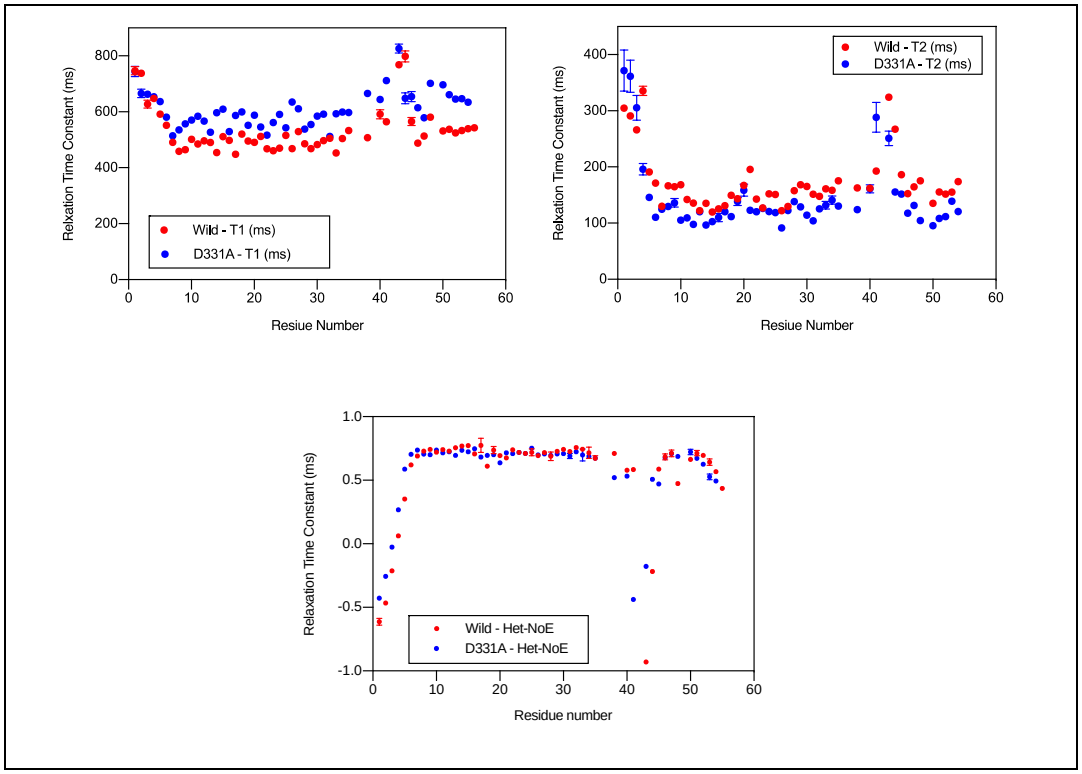


Figure 3.1b: T1, T2 & HetNOE of KDM5B wild type and its D2 mutant

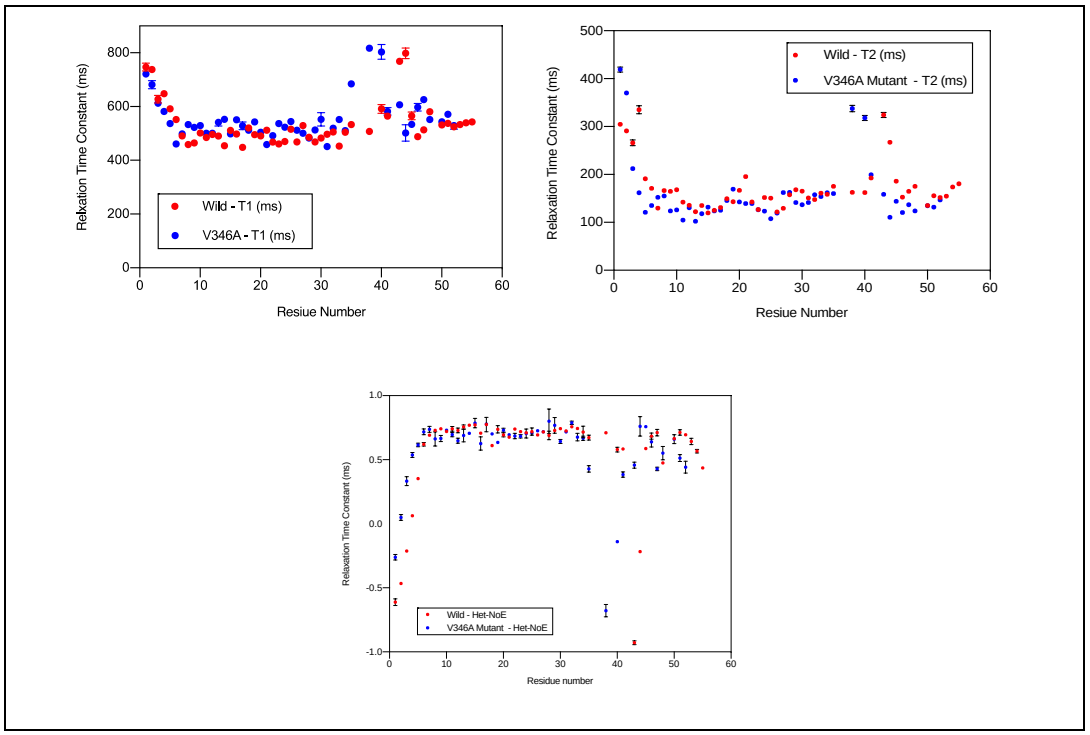


Figure 3.1c: T1, T2 & HetNOE of KDM5B wild type and its V mutant

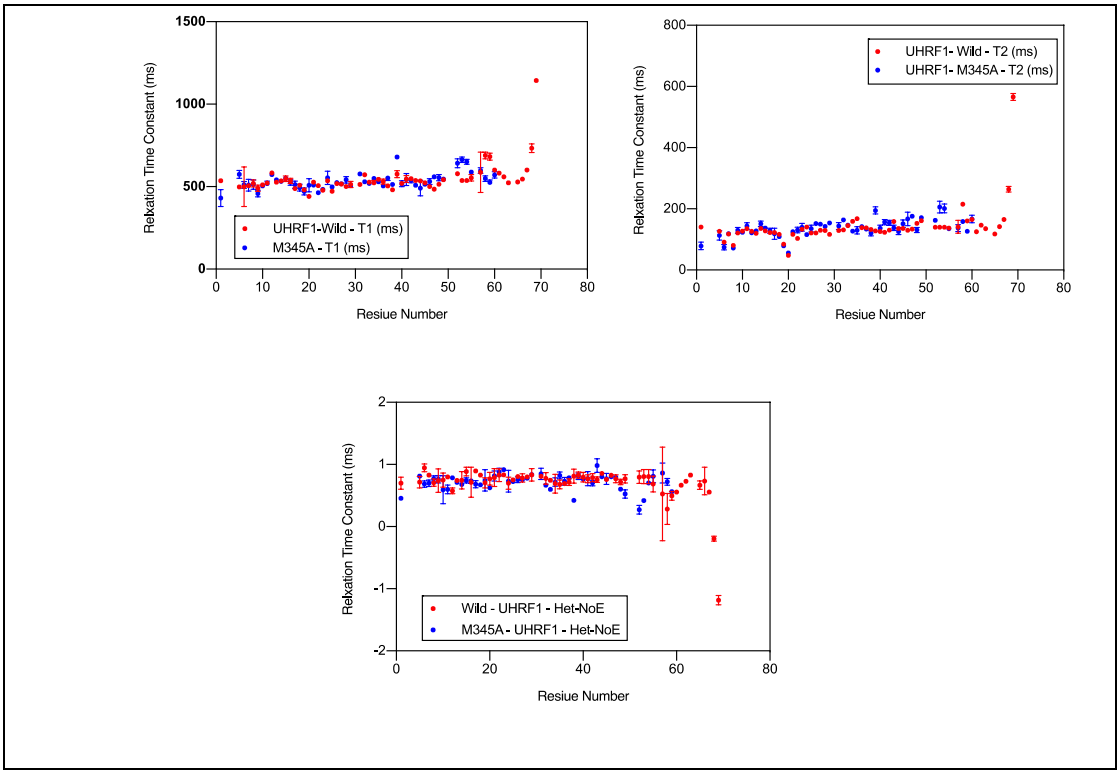


Figure 3.1d: T1, T2 & HetNOE of UHRF1 wild type and its mutant

---

Protein dynamics studies were done by NMR-T1, T2, and Het-NOE with the same wild type and mutant type samples of KDM5B and UHRF1 and it has been found that dynamically mutant proteins are behaving exactly the same way as the wild type does (Figure 3.1a-d).

We have compared our results of protein dynamic motion experiments with the findings of (Chatterjee 2019) in which they found different relaxation time than the wild type upon making mutation (F to L) (Figure 3.1.e). But in our case, we didn't see any remarkable changes in protein dynamic motion upon making mutation. It helps us to conclude that the coevolving residues are not contributing significantly for the dynamic motion of the PHD finger.

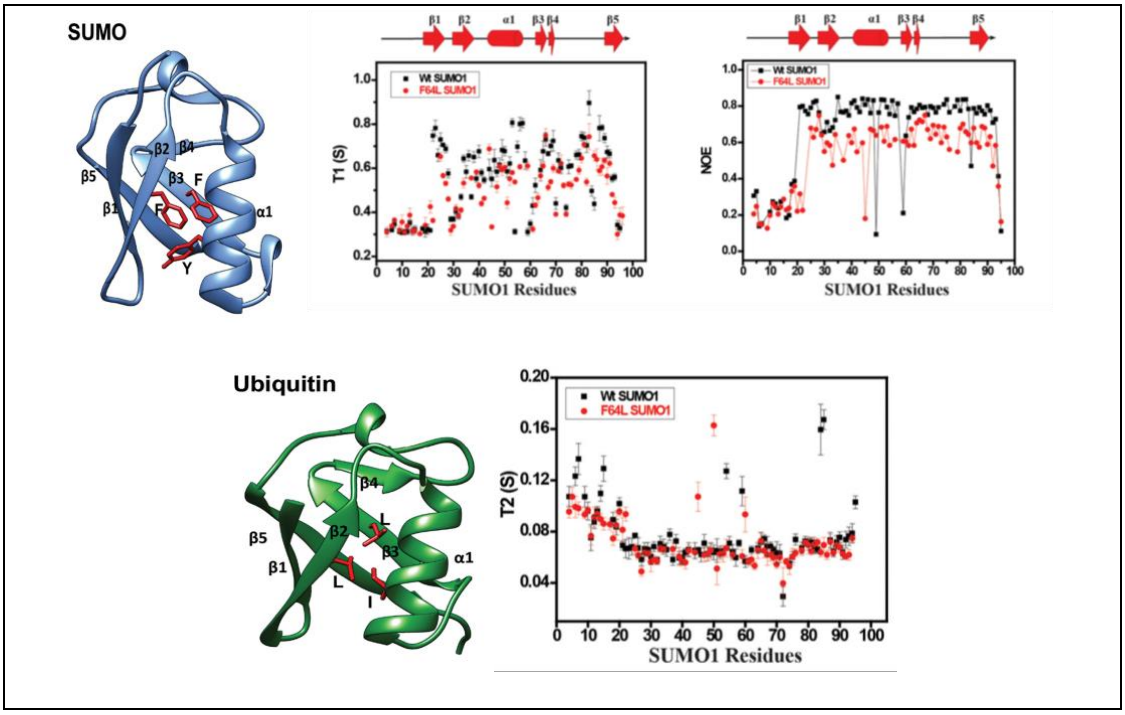


Figure 3.1.e: Reference study of protein dynamics (Chatterjee 2019)

We have performed a comparative analysis of the fractions of the preferred amino acids over the total numbers of the amino acid residues. and we found that this ratio is very less (0.26 and 0.22 respectively) in KDM5B and UHRF1 PHD finger that contains xCDxCDx motif compared to other control proteins that don't have this sequence motif. Therefore, it clearly indicates that even though the control proteins for example QM protein has this ratio pretty high compared to KDM5B and UHRF1, but still stability wise, QM protein is extremely weak as we found as a result of the FASTpp experiment (Figure 3.5). Thus, it can be concluded that the stability doesn't only depend on the presence of the preferred amino acids (Table 1.7).

Table 1.7: Sequence specificity of the proteins:

Proteins	Leu (L)	Phe (F)	Val (V)	Ile (I)	Ala (A)	Met (M)	Fraction of Preferred Residues [(L+F+V+I+A+M)/Total Number of residues]
KDM5B	9	1	3	1	2	0	0.26
UHRF1	5	1	3	1	3	2	0.22
AIRE	5	1	2	2	4	0	0.26
BPTF	4	1	2	4	2	0	0.21
QM	9	3	6	5	7	2	0.36
GST	28	9	10	14	10	9	0.34
RNase A	2	3	9	3	12	4	0.27

CHAPTER III

Determination of the Protein folding and stability of the PHD fingers

Abstract

Protein folding plays a key role in the biological system. It happens through various steps of molecular events and finally gets a three-dimensional shape which exposes binding sites, receptors and channels through which proteins interact with each other. Understanding this mechanism of how protein interacts with another protein can lead us to better understand the more complex processes. By understanding the protein folding and thermodynamic stability, we can get the detailed insight of the forces and energetics that helps to determine the molecular structure of the PHD finger. This study also helps us to better understand the protein domain organization and how interdomain interactions take place. More specifically, we have looked at the effects of the mutations on its overall structure and stability. Secondary and tertiary structures of the protein are resulted from the three-dimensional folding of the protein, in which there are several types of structures like α-helixes, β-sheets, turns and loops. We have had a comparative

study of these structures of the wild and mutant types of the PHD finger through CD at room temperature as well as in a wide range of temperature up to 110ºC. We didn't see any differences between the ellipticity of the wild as well as mutant types of the PHD finger. We have checked the stability of the PHD finger through melting study of the NMR-HSQC and we have observed the same pattern of chemical shift in either case of wild and mutant type. We have ensured the thermal stability by another quick and fast method called FASTpp. Using this technique, we have also determined the melting temperature of the PHD fingers and other proteins like GST, RNase A, and QM. Through this assay, we have confirmed the importance of the zinc atom in holding the structure of the PHD finger as we see PHD fingers to lose its structure while the zinc atom has been taken off. In this chapter we have also discussed about the extreme robustness of the PHD finger as it sustained through 4M Urea treatment when other proteins like RNase A, and GST started degrading at very low temperatures. Through these in-depth observations, we came to know about detail information of the PHD finger which definitely going to be a significant addition in the structural biology and disease prevention where protein folding, and stability are the main concerns.

Introduction

Stability of the protein depends on its structure and how the protein folds. It mainly concerns about how the amino acid sequence determines the three-dimensional structure of the protein. Protein folding is extremely important process because this is associated with functional aspect of the protein. Usually proteins become functional at its native state. Because in this state interactions between the residues are much more stable and persistent compared to non-native state, though in some cases it has been also found that proteins could be functional even at its thermally accessible non-native state (Henzler-Wildman K. 2007, Baldwin J.A. 2009). Protein folding is a very complex

process that gets through various degrees of structural organization such as numbers of partially ordered states that starts from denatured ensembles to intermediate and transition states respectively (Dill KA 1991, Mittag T 2007, Sosnick TR 2011, Zhuravleva 2017). During this process, if any mutation in the amino acid sequence occurs, if there is any translational error happens, protein might be misfolded or aggregated which may develop diseases like cancer, diabetes, autoimmune disorders, and neurodegenerative disorders (M.C. 2003). How the protein folding emerges from the fundamental condition, and how does the structural transition take place, can be investigated using various methods such as NMR spectroscopy (Dobson 1998), CD spectroscopy, and FASTpp. NMR is the quickest and best way to perform the structural analysis of the protein. It helps to get the detail characterization of the protein folding towards its native state. Method of NMR is one of the most reliable structural indicators of the protein. First chemical shift was observed for the 14N nuclei through NMR in the year 1950 (Proctor WG 1950). After few years, in 1957, 1H nuclei was observed using NMR facility (Arnold JT DSS 1957). For measuring the chemical shift, enormous advancement has been established by the researchers in past few years by simply increasing the sensitivity as well as resolution. Apart from structural information, chemical shift also provides other information too (Wishart DS 1991, Asakura T 1995, Asakura T 1999, Cornilescu G 1999), such as hydrogen bonding interaction (Osapay K 1991), oxidation states (Pastore A 1990) (MP 1990), ionization state, influence of aromatic amino acids, and the dynamic motions of the protons (Dalgarno DC 1983, E. 1995, L. 1995, Sharma D 2000).

We used NOE based NMR for characterizing the structure of the proteins. Study of ps-ns dynamics was performed using the NMR spin relaxation measurements because it can help to corroborate the accuracy of the chemical shift. Chemical shift is

one of the most effective parameters that explains the detailed structural as well as electronic properties of a protein. Chemical shift dispersion of the NMR spectrum (^{1}H and ^{15}N spectrum) can provide the information whether the associated protein is folded or unfolded. We have also performed melting comparison NMR-HSQC experiment of wild as well as mutant proteins as a function of temperature starting from 5ºC to 65ºCat 5ºC interval. We did this experiment to make sure if there are any changes in the melting behavior observed in the PHD finger after having mutations. Basically, we compared the chemical shift of the wild and the mutant types of the PHD fingers along with the increasing temperatures. For studying the melting behavior beyond 65ºC (maximum capacity of the NMR), Circular dichroism (CD) was performed up to 95ºC by our collaborator, Dr.Fuentes's research group of University of Iowa. CD is a very important tool for determination of secondary structure of a protein. Using CD, we can confirm whether the protein is properly folded or not. CD helps to determine the changes in the secondary structures of the wild type as well as mutant proteins along with the varying temperatures (WC 1999). Different structural elements of the secondary structures of the proteins have different CD spectra. Such as, negative bands supposed to be observed at 222 nm and 208 nm and positive band at 193 nm for α-helical proteins (Holzwarth G 1965). Proteins with β-pleated sheets supposed to have negative bands at 218 nm and positive bands at 195 nm, whereas, in case of disordered proteins, ellipticity was pretty low (approximately 195-210nm) (Greenfield NJ 1969), (Venyaminov S 1993). Spectra of the proteins are very much dependent on the conformational structure of the proteins. Thus, if there are any conformational changes due to mutations, temperature, denaturants, heat, or due to binding interactions, it can be monitored by CD spectra. CD doesn't give specific secondary structure of a particular residue like NMR can provide, but the best advantage of using CD in our case was that it could reach higher

temperature that can explain us whether any conformational changes happening with the increasing temperatures or not. Another advantage of using CD is also that data can be collected and analyzed from very less amount of protein like 20 µg or less in the aqueous buffers under physiological conditions (Greenfield 2006). Protein becomes most active when it gets properly folded. When this folded protein goes inside of the cell, it also needs to be stable for further functionality. To determine the stability, we have performed the NMR that successfully explains us structural conformation of the residue. We also performed CD for determining the secondary conformational structure, though this one is not that much specific like NMR but preferred for few factors like when there is a need of measuring CD spectra beyond a certain temperature which was perhaps not possible through NMR. We have also checked thermal stability by Isothermal Titration Calorimetry (ITC). It determines the binding interaction and the stability of the protein by monitoring the changes in the entropy and enthalpy. Additionally, to determine the stability and the robustness of the PHD fingers and control proteins, we have also performed series of Fast parallel protein proteolysis (FASTpp) experiments in different conditions.

Material and Methods

Protein stability test by FASTpp:

Function of a protein depends on its structure and stability. To find out the biophysical activity of a protein, it was highly required to check the thermal stability using FASTpp (Minde DP 2012). We performed FASTpp using a temperature gradient thermal cycler (Techne TC-3000) with simultaneous proteolytic cleavage (by thermostable protease Thermolysin) of the unfolded part of the proteins (Figure 3.2). Temperature plays a crucial role in protein stability as well as functionality. Different proteins get degraded at different temperatures. Fundamental idea of this experiment is

that proteins are being exposed to simultaneously higher temperatures using thermal cycler with different pulses of time and expected to get unfolded with the increasing temperatures, and the unfolded part of the proteins get digested by the thermostable protease. Thus, to see the ratio of the folded fraction of the protein left over in the wide range of temperature, we have performed this experiment for the wild as well as mutant forms of the PHD fingers.

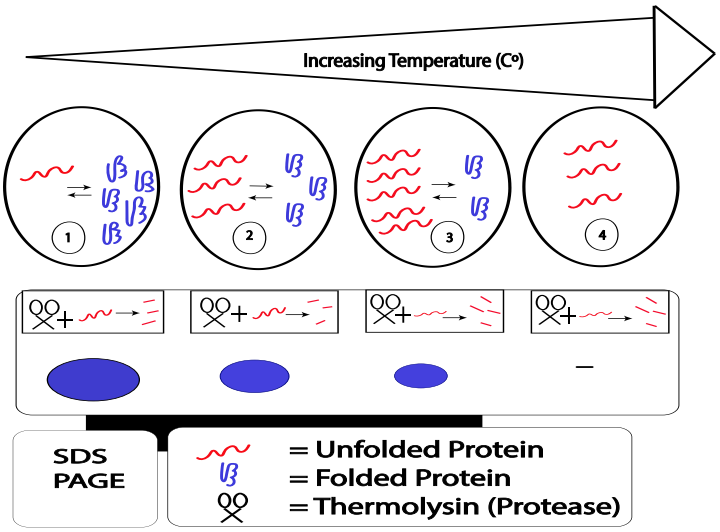


Figure 3.2: Schematic diagram of FASTpp experiment to check the thermal stability of the protein in a thermal cycler on a row of temperature gradient.

Without further purification, commercial Thermolysin (TL) from *Bacillus thermoproteolyticus rokko* (Sigma Aldrich) was used to perform the experiment. A stock solution of TL (5mg/ml) was prepared in 2.5 M NaCl containing 10 mM $CaCl_2$ (Park C 2005). The concentration of TL (5mg/ml) was determined spectrophotometrically at 280 nm and the molecular mass was 34.6 kDa. We have performed FASTpp experiments with PHD finger and control proteins (6 kDa to 26 kDa respectively). Our total reaction time for each sample was 2 minutes pulse. First,

we made the master-mix of the protein sample and the TL protease and then aliquoted into PCR tubes. After that we started heating the aliquoted sample mix during the heating time (h_t) using the gradient PCR and exposing the sample mix to a variable range of melting temperature (t_m) with a subsequent cooling time (t_c) that brings the reaction down to 4ºC until reaction was quenched with 50mM EDTA. We have performed FASTpp in different conditions. First we have exposed the protein samples (RNase A, GST as control, and KDM5B wild type, KDM5B mutants as the PHD fingers) in different temperatures (10, 20, 30, 40, 50, 60, 70, 80)°C for two minutes to determine the melting temperature of the protein. Later on we have narrowed down the window of melting temperature (40, 43, 46, 49, 52, 55, 58, 61, 64) ºC and (43, 46, 49, 52, 55, 58, 61, 64, 67) ºC range for RNase A and GST. To determine the robustness of the PHD finger in compare to other control proteins (BPTF, AIRE, RNase A, GST, and QM), we have also performed FASTpp in which we have treated the protein samples (0.25mM) with 4M Urea and then run through different temperatures (10, 20, 30, 40, 50, 60, 70, 80)°C using thermal cycler for 2 minutes. Besides this, keeping the temperature constant at 25ºC, we have performed FASTpp at different time frame (2m, 5m, 10m, 30m, 1hr, 2hrs, 3hrs, 5hrs, and 24hrs) to determine the sustainability of the PHD finger in comparison to other control proteins like RNase A and GST. At the end of each run, we have added 50mM EDTA (5ul) to stop the reaction immediately. Then we have run the samples in 15% SDS PAGE gel. We have performed this experiment with PHD-Finger- KDM5B wild type and three of its mutants (V346A, L326A, D331A), and control proteins were: AIRE (5.9kDa), BPTF (7.5kDa), RNAse A (13kDa), QM (11kDa) and GST (26 kDa).

We have chosen TL suitable for FASTpp for several reasons: (i) TL is thermostable up to 70-80°C. (ii) TL can cleave peptide bonds near N terminal end of

hydrophobic, bulky and aromatic amino acids, specifically Phe, Leu, Ala, Val and Ile. Since TL has a preference for large hydrophobic and aromatic amino acids, thus it has been considered very specific for FASTpp. Usually these types of hydrophobic amino acids buried themselves inside the hydrophobic core of the folded proteins and only come out while the proteins become unfolded. After being unfolded, these amino acid residues get exposed and immediately digested by TL. (iii) TL is very stable in a wide pH range of 5.5 to 9. (iv) TL is also stable at high concentrations of chaotropic reagents like 8 M urea and in the presence of EDTA-free protease inhibitors cocktails. (v) TL is instantly inhibited by addition of EDTA, which removes TL' s essential Ca^{2+} ion.

Using FASTpp, we have also determined the melting temperature (Tm) of GST, and RNase A protein. To have a comparative overview about the robustness of the PHD finger, and other control proteins like RNase A and GST, we have performed FASTpp keeping the temperature constant at 25°C, but variable pulse of time (2m, 5m, 10m, 30m, 1hr, 2hrs, 3hrs, 5hrs, and 24hrs).

Circular Dichroism:

CD was performed using 95% pure proteins. For obtaining high quality CD data, concentration of the protein should be accurate. Proteins were purified as described before (0.3 mM protein in 8 mM Phosphate buffer pH 7.6, 50 mM NaCl). Then samples were diluted to a final concentration of 20 µM in Phosphate buffer (25 mM 1XPBS [pH 7.6],125 mM NaCl, 5% glycerol) and analyzed using a 1-mm cuvette equipped with the peltier thermostating cuvette holder in a JascoJ-815 CD spectrometer under the constant nitrogen flow. Each spectrum was recorded in a 0.2-cm cell at a protein concentration of 0.05mg/ml, and averaged over three times with a final correction for the buffer signal. The raw data were revised and corrected to get the pre and post transition slopes. Quantitative amino acid analysis using the concentration of

the stable amino acid such as alanine and lysine to calculate the concentration of the whole protein, is the most accurate way of determining the protein concentration. This method is extremely sensitive, and we have performed it on the aliquots of the actual CD sample. Melting curves were obtained by measuring $[ø]_{222}$ as a function of temperature as depicted in (Hecht 1984). We have used molar extinction coefficient 8.48 for calculating the concentration of KDM5B wild type and its mutant. Using nanodrop, and after blanked with the buffer (10mM phosphate, pH 7.6, 50mM NaCl), concentrations of the protein samples were being measured and standardized to get the good signal of CD. Protein samples were diluted according to the need of the CD experiment.

NMR HSQC, melting comparison and protein dynamics study:

NMR spectroscopy was used to characterize the structure and the dynamics of the PHD fingers and its mutants. ^{15}N labelled proteins (PHD fingers) have been purified following protocol mentioned in chapter II and performed different NMR experiments such as HSQC, melting comparison, and protein dynamics studies with the collaboration of NMRFAM of University of Wisconsin, Madison. Data analysis was done by several packages like Sparky developed by NMRFAM. Several layers of modernization in this software have been done by NMRFAM using codes written in C++ and extended with Python. They have released a new version called NMRFAM-SPARKY which is extremely useful as well as user friendly for graphical NMR assignments of biomolecules (Lee W 2015) and this has been used for our data analysis. NMR-HSQC has been done for studying the detail molecular structure of the protein molecules. The spectrum that we got from the NMR, is two dimensional. NMR chemical shifts are highly important for determining the structural motifs of the protein. In one axis there is proton (^{1}H), and in another axis there is a heteronucleus

(other than proton) and in our case it is ^{15}N. There are unique peaks for each proton in the spectrum and each proton is attached to ^{15}N labelled heteronucleus. Secondary chemical shift being used for probing the α-helices and the extended conformation (Wishart DS 1995).

The fundamental idea of this experiment is to transfer the magnetization of one proton to another ^{15}N via insensitive nuclei enhanced by polarization transfer. After T1 and T2 time delay, magnetization is being transferred back to the earlier proton via reverse insensitive nuclei enhanced by polarization transfer and the signal has been recorded by that time. NMR experiments, Heternuclear Over Hauser Effect (Het-NOE) were also performed to determine the rigidity of the protein backbone. If the protein is properly folded, then it must have secondary structures. In that case, heteronuclear ^{1}H-^{15}N NOE values should be larger than 0.7 while secondary structures are associated with the rigid molecules. But HetNOEs show values very close to zero or negative while proteins have disordered backbone and lack of secondary structures (Hellman M 2011).

In case of NOE, transverse spin magnetization has been created by first pulse. Longitudinal magnetization, which is equal to transverse magnetization, was being produced by the second pulse. Between these two pulses, spins processed during time T1. Each magnetization was orthogonal to pulse direction. Thus, NOE magnetization was transferred through cross-relaxation. During third pulse, transverse magnetization was being created and observed as a function of time delay T2. Following third pulse, data acquisition started. In respect to T1 and T2, and with the help of 2D Fourier transform, NOESY spectrum was being generated. Thus, the experiments of NMR-NOE have been done in order to determine the protein dynamics and motions. Rapid build-up rates of NOE were being observed between the closest protons.

To determine the sturdiness of the PHD fingers and its mutants along with the increasing temperature, we have performed comparative melting temperature experiments through NMR at 800MHz spectrum. A range of 5 to 65ºC temperature was being used to observe the chemical shift as a function of the melting behavior of the proteins of interests.

Results and Discussions:

The temperature at which protein starts unfolding, works as a definite indicator of the protein stability. Factors that are affecting the stability of the protein, must affect the unfolding temperature (Ru¨diger S 2002, Mayer S 2007). If there is a mutation in the amino acid sequence, there could be a compromise in the structure of the protein. In that case, the temperature can be lower at which protein starts unfolding whereas ligands that only binds to folded state, shifts the temperature of unfolding to higher values (Bullock AN 2000, Friedler A 2002, Hoelen H 2010). To determine the fraction of folded part over the wide range of temperature, a thermostable protease called Thermolysin was being used to digest the unfolded part of the protein. In this regard, we have performed FASTpp to find out the thermal stability of the PHD fingers and whether mutants play any role in compromising the robustness of it. Proteins were being exposed to different temperatures (above and below the melting temperature) in presence of the thermostable protease-Thermolysin. To determine the stability of the protein, time dependent degradations were being noticed. Precision of FASTpp depends on few key factors: heating time (t_h), maximum time of exposure to the temperatures (t_e) and the period of cooling down (t_c) (Figure 3.3).

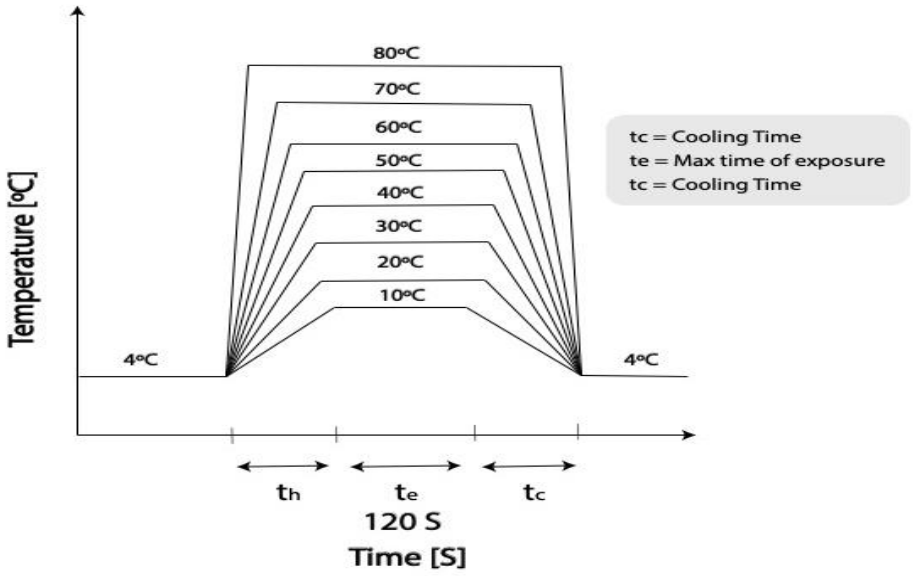


Figure 3.3: Controlling factors of FASTpp. t_h = heating time; t_e = maximum time of exposure to the temperature; t_c = period of cooling down.

We have tested the thermal stability of the PHD finger and its mutants in a wide range of variable melting temperature. FASTpp was a unique method that can show us how the proteins getting degraded with the increasing temperature. PHD finger can be obtained in two states: one is either unfolded without the zinc atom, another is folded with the zinc atom. We have tested whether both states can be distinguished by FASTpp (Figure 3.4).

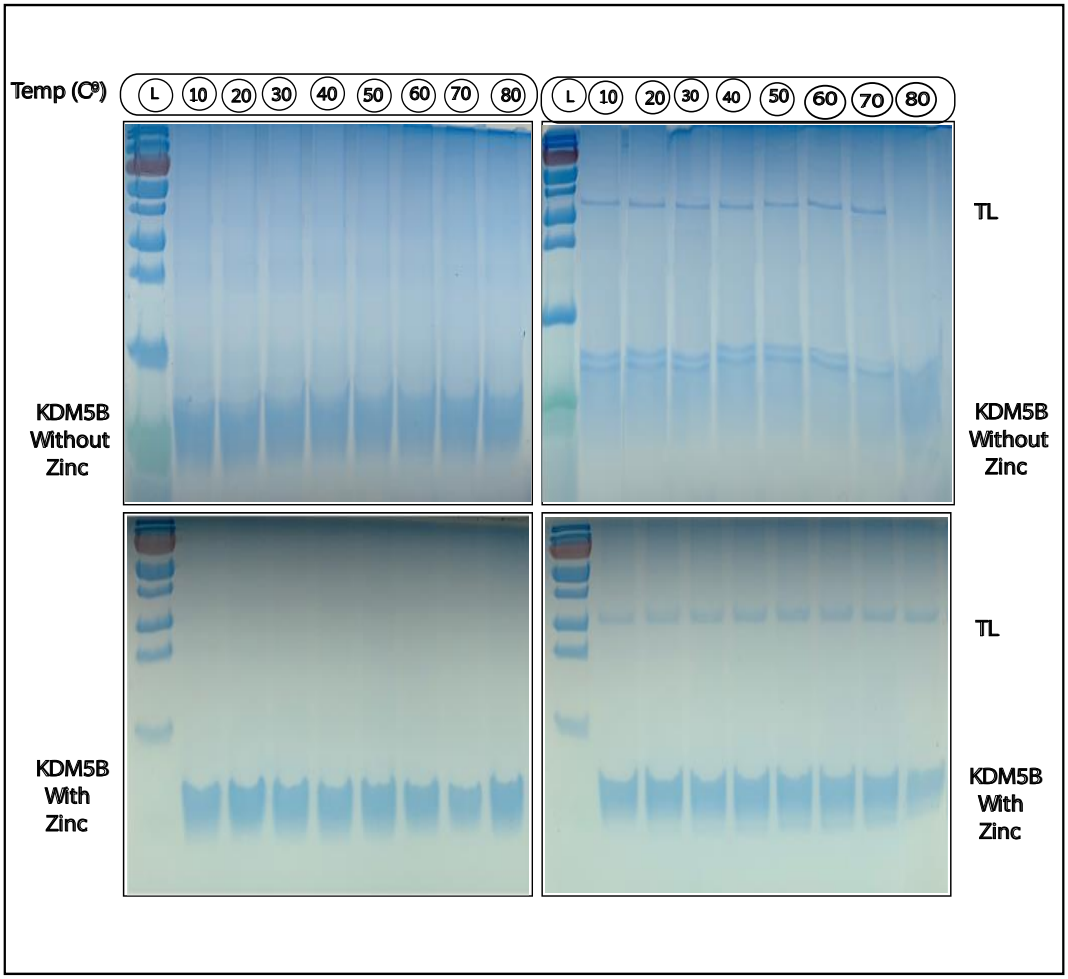


Figure 3.4: Role of Zinc atom in folding nature of PHD finger: Degradation of the PHD finger (KDM5B) without Zinc and folding back with the incorporation of the zinc atom upon treated the protein with the zinc chloride. Tested by FASTpp at gradient temperatures.

Since Zinc atom helps to hold the structure of the PHD finger (Low LY 2002), so we wanted to determine whether the PHD finger (KDM5B) still folds even after losing the Zinc atoms. We have grown the cells in M9 minimal media with and without $ZnCl_2$. Subsequently, we got the KDM5B protein with and without Zinc. We have tested the effect by FASTpp experiment in presence of Thermolysin protease. In absence of the Zn, PHD finger got unfolded and digested by the protease. After providing the Zn atom back to the PHD finger, it was found to be folded back and

found to be very sustainable even at very high temperature (80ºC). To confirm it, we have performed FASTpp in presence/absence of Thermolysin protease.

We have also tested the stability of the PHD finger through urea treatment. Urea denatures the protein by binding to the amide units via hydrogen bonds and thus reducing the hydrophobic effect (Zou Q 1998). We have performed FASTpp of the PHD finger (KDM5B) and its mutants (L2, V, and D2) with 4M urea to find out whether mutations have any effect in the structural stability of the PHD finger. As control experiments, we have performed FASTpp of 4M Urea treated other proteins (RNase A, GST). We found that both the control proteins (RNase A, GST) started degrading after 40ºC while they were being treated with 4M urea. But in case of KDM5B wild and mutant types we didn't notice any degradation until 60-70ºC even after getting treated with 4M urea. It clearly shows that mutations do not have any significant effect in the stability of the PHD finger (Figure 3.8).

We have also compared the thermal stability of the PHD fingers and its mutants with other proteins that don't have Zn^{2+} ions in their structure. In order to do this, we have performed FASTpp of the KDM5B-PHD finger, and its mutants (L2, D2, and V). We have found that mutants are equally stable at very high temperature (80ºC) just like as the wild type PHD finger is stable (Figure 3.7). In presence/absence of Thermolysin, FASTpp (from 10 to 80ºC) was performed with RNase A, QM, and GST proteins which are being considered as control proteins. RNase A is a very sturdy protein, still it started degrading after 50ºC and the unfolded parts getting digested by the protease. GST also behaved the same way. After 50ºC, GST started getting unfolded and getting digested by the protease. In case of QM protein, this is even a very weak and unstable protein as we have found it to be digested by the protease from 20-30ºC (Figure 3.5). We have also compared the thermal stability of the PHD_nW_DD (KDM5B, and its mutants)

and PHD_nW (BPTF, AIRE) (Figure 3.6). In order to do that, we have performed FASTpp in presence/absence of protease We found that both subtypes of PHD finger are almost equally thermostable, rather PHD_nW_DD subtype is slightly more stable than the rest as we can see that KDM5B and its mutants are not being digested by thermolysin protease even at 80ºC while BPTF started getting digested after 60 ºC.

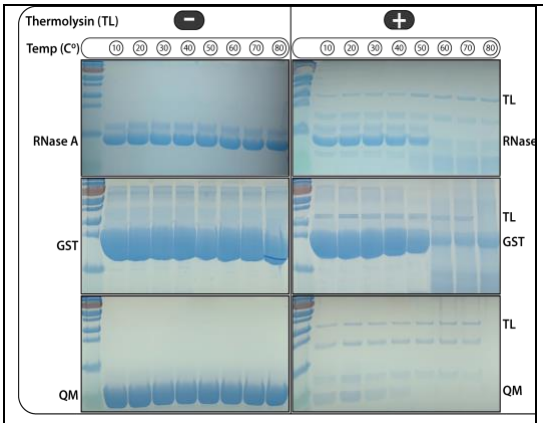


Figure 3.5: Comparative Stability test (FASTpp) of Reference proteins (RNase A, GST, and QM) at different temperatures

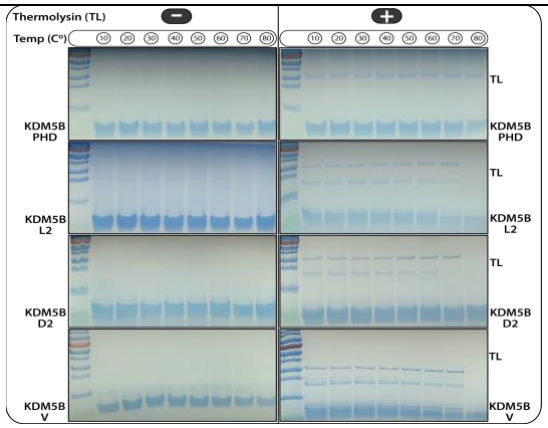
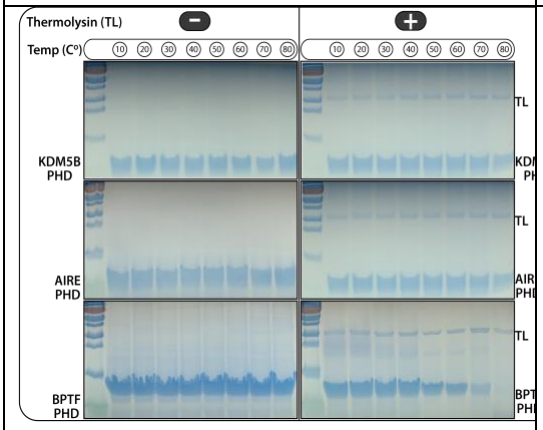


Figure 3.7: Comparative Stability test (FASTpp) of KDM5B wild type and its mutants (L2, D2, and V) at different temperatures



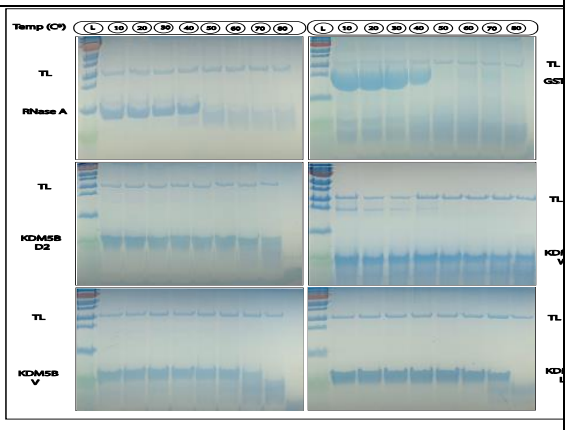


Figure 3.6: Comparative Stability test (FASTpp) of KDM5B-PHD with other PHD fingers at different temperatures	**Figure 3.8:** Comparative Stability test (FASTpp) of 4M Urea treated KDM5B wild type and its mutants (L2, D2, and V) and FASTpp of 4M Urea treated other proteins (RNase A, GST) as control

We have also performed FASTpp of KDM5B-PHD finger and its mutants along with relatively stable control proteins like RNase A and less stable control protein like GST as a function of time (2mins to 1440 mins). We did this experiment at room temperature (25ºC). We wanted to see how long PHD finger wild and its mutant types can retain its folded structure at room temperature. We have found RNase A to be very stable until 5 hours (300mins) as we can see clear band for it in the 15% SDS PAGE gel, after that as soon as the proteins started getting unfolded, Thermolysin started digesting the unfolded parts of the protein. We found the PHD fingers and its mutants very stable also whereas, GST started degrading and getting digested by Thermolysin after an hour (Figure 3.9).

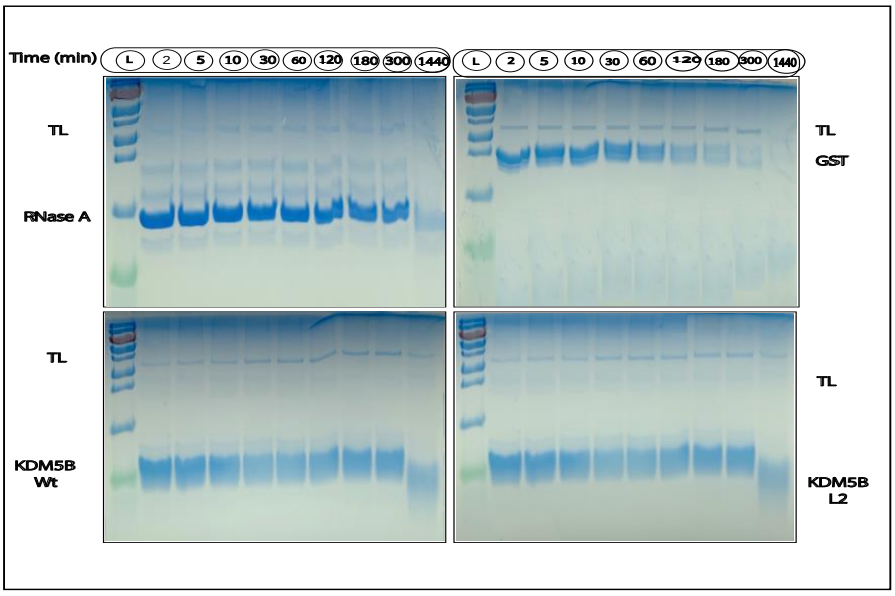


Figure 3.9: FASTpp of the PHD finger and its mutants, RNase A, and GST as a function of time (min).

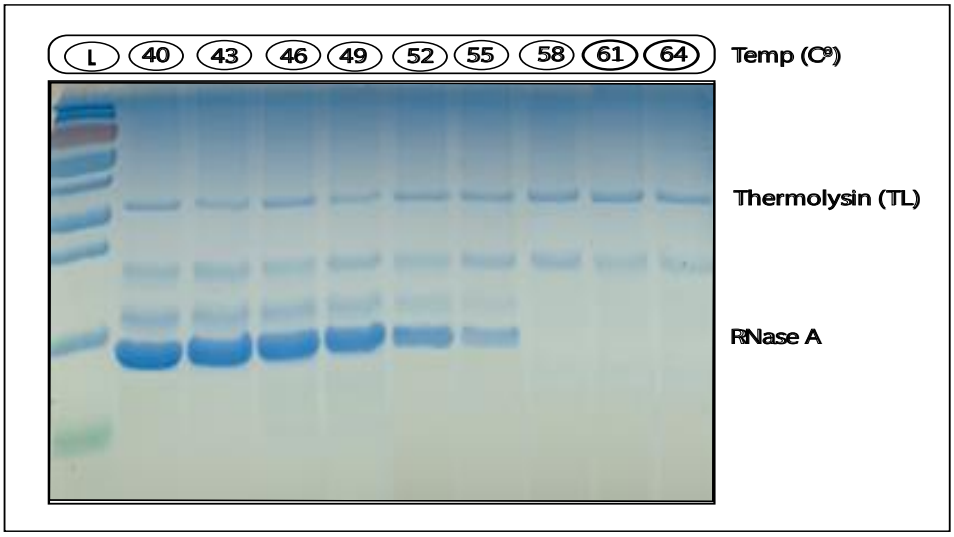


Figure 4.0: FASTpp of RNase A as a function of Melting Temperature (Tm)

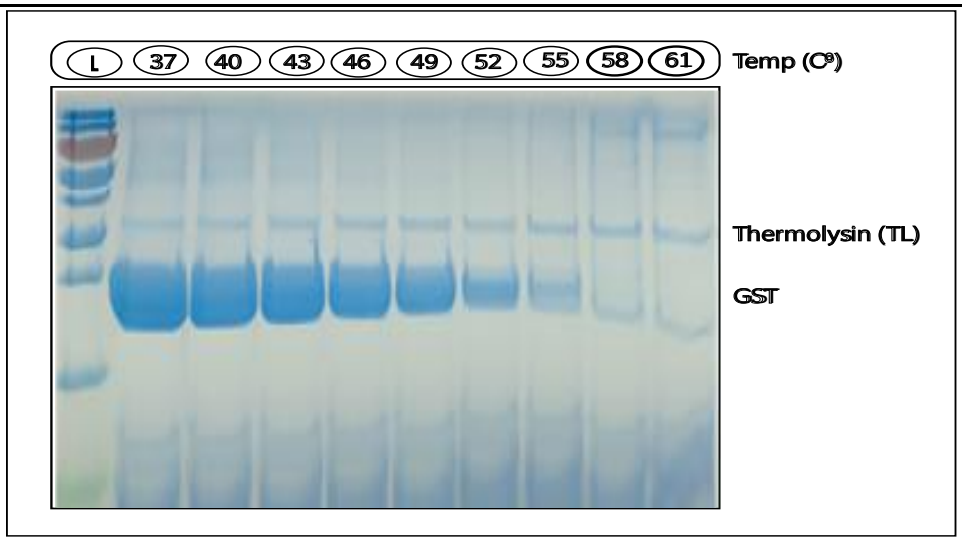


Figure 4.1: FASTpp of GST as a function of Melting Temperature (Tm)

From the FASTpp melting temperature experiment of the RNase A and GST, we found that 52ºC is the temperature after which the RNase A started unfolding and slowly getting degraded and GST also started getting unfolded after 52 ºC as well (Figure 4). Since Thermolysin protease was present in the sample solutions, therefore,

the unfolded parts of the proteins were being digested (Figure 4.1). Comparing the melting temperature of these two control proteins with the PHD finger (KDM5B) and its mutants, it has been found that PHD finger is extremely robust protein. Even though we made mutation in its amino acid sequence and expected to observe changes in the behavior as mutation might have impact over the structure but didn't notice any significant changes in the mutated protein from the aspect of thermal stability.

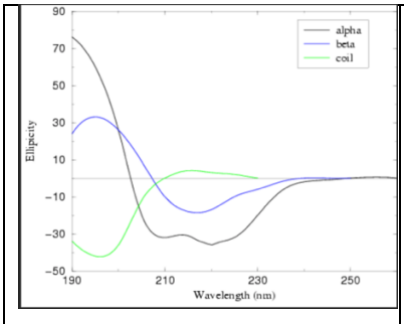
In the field of proteomics, it is necessary to characterize a protein from its structural and stability aspects. CD is one of the most effective ways to evaluate the secondary structures, and folding pattern of the proteins, whereas fluorescence of the tryptophan amino acid is sensitive to the local molecular environment (Lin S 1996). CD involves circularly polarized light, i.e. differential absorption of the left and the right-handed light. Mathematical denotation is given below:

AL = EL. CL; AR = ER.CL;

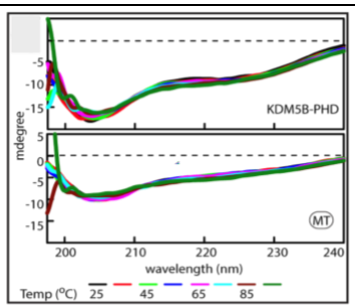
AL – AR = (EL – ER). CL [Since same Protein, so C is same for LCP and RCP]

AL – AR = (EL – ER).

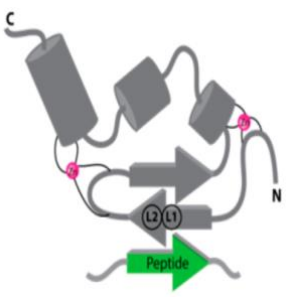
EL and ER are the molar extinction coefficients for LCP and RCP light, C is the molar concentration; L is the path length in centimeters (cm). So, ∆A= ∆E. It appears in the absorption bands of optically active chiral molecules.



A



B



C

Figure 4.2: A. Standard curve of Circular Dichroism (CD). **B.** CD of KDM5B-PHDwt Vs. KDM5B-PHDmt. **C.** Cartoon of PHD finger interacting with peptide.

From the CD result, it has been found that there is no unfolding even at 95°C for any of the samples of mutants as well as wild type PHD finger. The signal at 205nm or 220nm does not decrease with increasing temperature. None of their secondary structures changed with the increasing temperature (Figure 4.2). We have also performed CD at 25ºC for KDM5B wild and its mutants (L2, D2, V) (Figure 4.4). We have found that signals at 205nm and 230nm drops a little bit for KDM5B L326A, but in case of wild type KDM5B and rest two mutants (D331A, and V346A), we didn't see any signal dropping. Therefore, it can be concluded that they are retaining their secondary structure at room temperature (Figure 4.3).

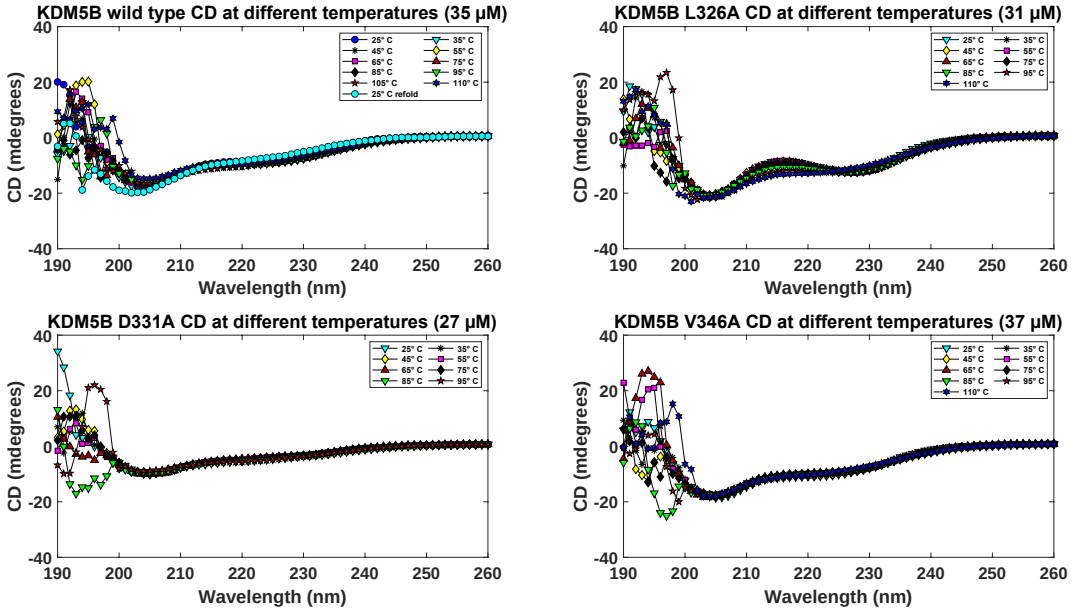


Figure 4.3: Circular Dichroism of PHD finger (KDM5B) and its mutants at different temperatures

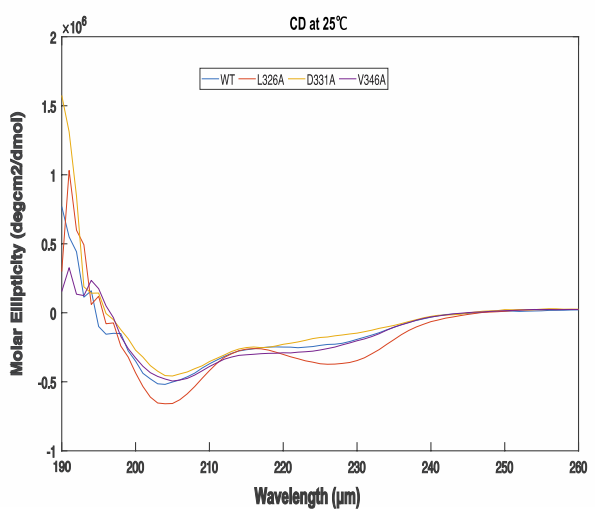


Figure 4.4: Circular Dichroism of PHD finger (KDM5B) at room temperatures. Compared Molar ellipticity of the PHD finger and its mutants.

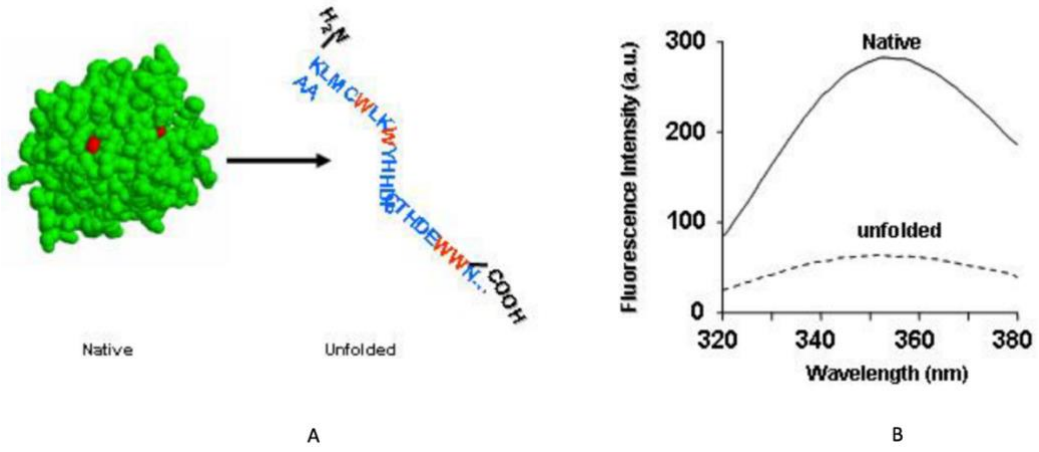


Figure 4.5: A. Protein unfolding as observed by Tryptophan fluorescence (Trp:W highlighted by Red). **B.** during folded state, Trp remains in the buried and hydrophobic environment and thus fluorescence intensity becomes very high, while during unfolded state, Trp gets exposed to the solvent, fluorescence intensity started decreasing (PC3267 2007).

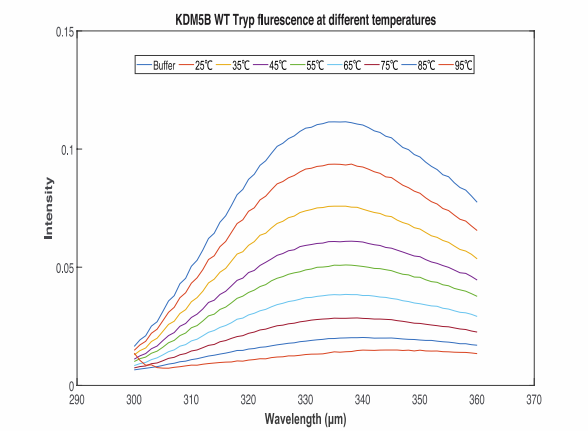


Figure 4.6: Fluorescence intensity of the Tryptophan of KDM5B-PHD at different temperatures.

Amino acid sequence of KDM5B has a single Tryptophan (W) residue. Wavelength of this tryptophan fluorescence emission peak is an important parameter to characterize the unfolded state of the protein (Duy C 2006) because fluorescence of tryptophan amino acid is highly sensitive to the surrounding climate. For example, if the tryptophan amino acid gets more access to the water or the solvent, which mostly happens in case of unfolding state, intensity of the emission starts decreasing (Figure 4.5). With the increasing temperature, proteins start unfolding and the buried Tryptophan starts getting exposed to water. But in case of KDM5B, peak for every temperature shows very less fluctuation from the previous one (Figure 4.6), which signifies that even though there are some unfolding noticed with the increasing temperature, but that is not significant enough compared to other proteins that remarkably unfolds with the increasing temperature. The peak ranges from ~308nm to 350nm represent non-polar environment inside the globular protein to the polar environment, the protein surface surrounded by water molecules respectively (Burstein 1976, Vivian 2001).

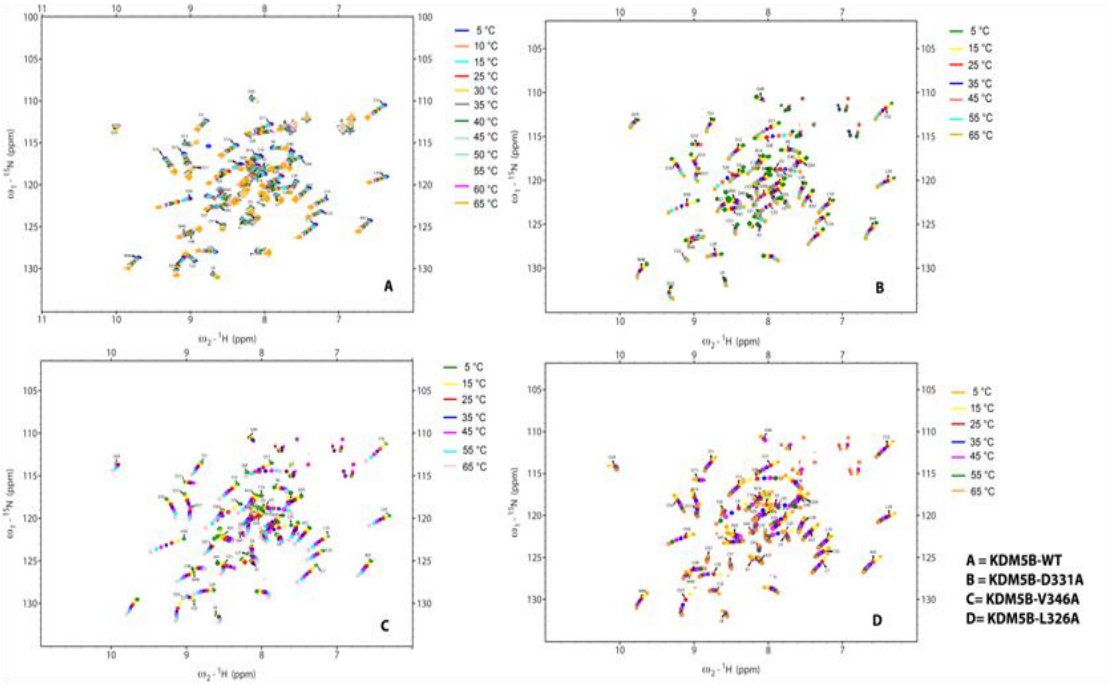


Figure 4.7: Comparison of the chemical shift as a function of the temperature (5-65) ºC.

A linear relationship as a function of temperature has a significant impact on the chemical shift of the amide proton (Baxter NJ 1997). Besides this, changes related to the nonlinear relationship has also been found to impact over the chemical shift of the amide proton, that contributes significantly in the conformational states of the high energy (Kumar A 2007). We have compared the chemical shifts through NMR-HSQC of the KDM5B and its mutants as a function of the temperature (5-65) ºC. We have found that the distances between the 15N labeled protons are similar in wild as well as mutants of the PHD finger (KDM5B). Also, the proton dispersion patterns are quite similar in the wild and mutants. This temperature-dependent NMR-HSQC spectra indicate that both wild and mutant KDM5B almost equally folded and not losing their three-dimensional molecular structures even at higher temperatures (65ºC) (Figure 4.7). Whereas, disordered proteins have much less spread out spectra compared to the spectra observed in properly folded proteins (Wolf-Watz M. 2004, Klechner IR. 2011).

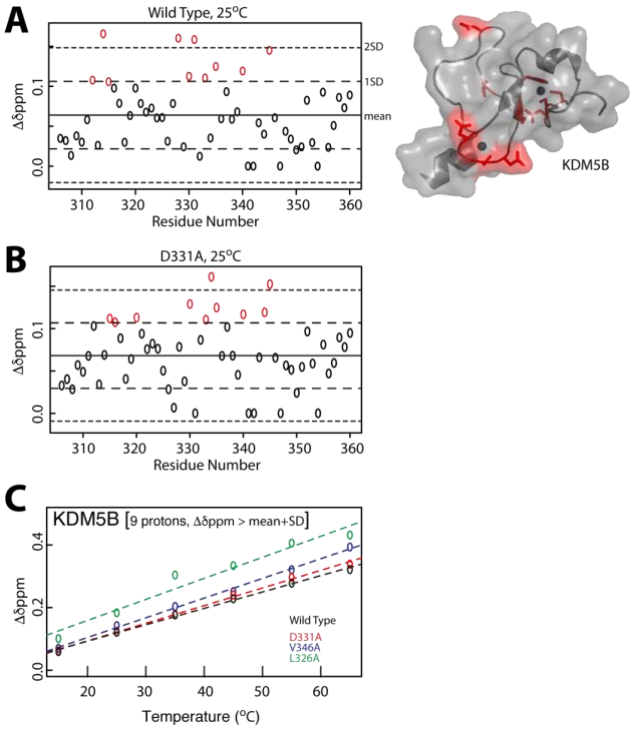


Figure 4.8: Statistical analysis of the melting comparison of Wild type KDM5B & its Mutants at different temperatures

---

We have performed statistical analysis of the melting behaviors of the PHD finger (KDM5B) and its mutants at room temperature (25 ºC) and at varying temperatures (20 to 60 ºC) respectively. At room temperature, we have found only 9 protons are showing Δδppm > mean+1sd (highlighted with red in the cartoon shown in Figure 4.8.A). We observed the same result for the mutants as well. Based on this finding of the 9 protons which are showing Δδppm > mean+1sd, when we plotted wild type KDM5B and three of its mutants (D331A, V346A, and L326A), we didn't notice a remarkable fluctuation in the wide temperature range of 20 to 60 ºC. Negligible fluctuation (~0.05 Δδppm) was observed in case of L326A (Figure 4.8.B,C).

From the above experiments and consecutive results, we can conclude that PHD fingers are extremely stable and not losing its folding structure even at very high temperatures. We wanted to make sure whether mutations are causing any significant impact in their folding behavior which must be reflected in their function. From our previous study we came to know that upon making mutations, there are significant changes observed in their binding ability. Since PHD fingers are known for their reading ability of the histone proteins and highly important for chromatin signaling, so upon making mutations, their binding ability must be compromised to some extent. After proving that, we further wanted to make sure whether mutations are affecting the folding ability and the stability of the PHD finger or it is just restricted to binding ability. We found that instead of having mutations, PHD fingers are still folding properly and very stable at high temperature (up to 80 ºC), and also didn't unfold while being treated with 4M urea. But while we took off the zinc atom, we saw PHD fingers start to become unfolded. It clearly indicates that zinc atom plays a critical role in holding the structure of the PHD finger and also making it very stable and sturdy. Therefore, it's clear that mutations are not playing a significant role in folding and stability though they are crucial for the binding ability of the PHD fingers.

CHAPTER IV

Binding study of the PHD finger

Abstract

PHD fingers are structurally conserved modules that are found in many proteins that control the chromatin signaling. By controlling the chromatin signaling, they also play a crucial role in several cellular mechanisms like transcription, DNA repair, gene activation, etc. It has been found that the PHD fingers have a special capacity to read the histone peptide and it can be modulated by the interplay between different types of histone modifications. We have studied the effects of position-specific mutations in the binding affinity of the PHD finger, also we have studied the contribution of the coevolving residues (xCDxCDx motif) in the substrate binding. Compared to some RNA binding proteins where the charge of amino acids matters more than the complementary shape of the amino acids that play a crucial role for the binding affinity of the KDM5B-PHD finger. For example, if Arginine is replaced by Lysine, some RNA binding proteins will still bind to the RNA, as a positive charge of the amino acid plays the main role instead of the structure of the amino acid, but in case of KDM5B-PHD finger, we have observed that upon replacing the Arginine with Lysine, binding completely falls off. The binding was also disrupted while Lysine was replaced by Arginine. We wanted to understand the shape-wise sensitivity and contribution of each amino acids towards the binding affinity of the PHD finger. All these binding affinities have been probed by ITC and we have observed drastic conformational changes in the binding capacity of the PHD fingers upon altering the specific amino acids.

Introduction

Histone N-terminal tails are associated with numbers of post-translational modifications, and these are found to be regulating chromatin structure and chromatin signaling by getting involved in the series of processes like replication, transcription, and recombination. One of such modifications we are interested in, is histone acetylation. It occurs at the amino group of the lysine at the N-terminus of the H3 and H4. Lysine acetylation is catalyzed by histone acetyltransferases (HAT) and removed by histone deacetylases (HDACs). So far bromodomain was the only known acetyllysine reader, but recently numbers of other modules have been identified as histone acetyllysine recognition modules. PHD fingers are one of them, which are significantly known to probe the activation of the genes by reading the acetylated histones. From the previous studies, it has been found that the highly positively charged N-terminus part of the H3K14AC peptide is recognized by the negatively charged electrostatic potential of the protein interacting surface of the PHD finger (Li 2016). In this study, we have also tried to characterize the histone peptide, how each amino acid is contributing to the binding energetics, was one of our main concerns. We have purchased the synthetic histone H3 peptide with mutations at multiple positions and performed ITCs with the histone readers, mainly PHD fingers from different subgroups. Since nucleosome is a complex and tight structure composed mainly of the DNA and the histone proteins, therefore, we wanted to see the effect of the tiny methyl group in the packing density. Here our question was to address whether the addition or removal of the methyl group is contributing in the tight packing of the nucleosome. To address this question, we have purchased the commercial peptide with the mutation in the threonine to serine and performed ITCs with the PHD fingers to observe the changing attributes in the binding energetics.

Materials and methods

Protein Purification

Some of our constructs were in the pET28A vector. *E. coli* BL21 (DE3) cells were used for expressing the Histidine tagged recombinant proteins. Overnight grown LB culture of the construct was inoculated in 4 liters of TB media along with Kanamycin and let it grow at 37°C incubator shaker until OD reached to 0.6-1.0 at OD600. Then cells were induced with 1mM IPTG (500 µl of 1M in each 1litre TB culture) and transferred to 16°C incubator shaker (Biotek) for 18 hours for optimal growth as well as expression, following which cells were being collected and stored in -80°C until we got ready for purification. Initial few steps of Histidine tagged protein purification are same as GST tagged protein purification. We have used Ni-NTA (GE healthcare) beads for binding the histidine-tagged proteins and let it rock for overnight. Following morning proteins supposed to bind with the beads, and the beads with the protein were washed a couple of times with 1x PBS followed by 1M NaCl and again 1x PBS. We have used 1M NaCl buffer to get rid of the non-specifically bound proteins from the resins which are only supposed to bind with the target protein with high affinity. After thorough washing, proteins were eluted by 100mM, 150mM and 250mM imidazole elution buffer. Initially, three different concentrations of the elution buffer were being used to find out which one can give the best yield of eluted protein. After elution, for further purification, protein was being purified through HiLoad 26/600 superdex column using FPLC. We have confirmed the quality of the protein through 15% SDS PAGE gel. We have purified Histidine tagged proteins like BPTF wild type and mutant, KAT6A, DPF3, PHF13, DIDO wild, and mutant types using the above-mentioned method (Figure 5).

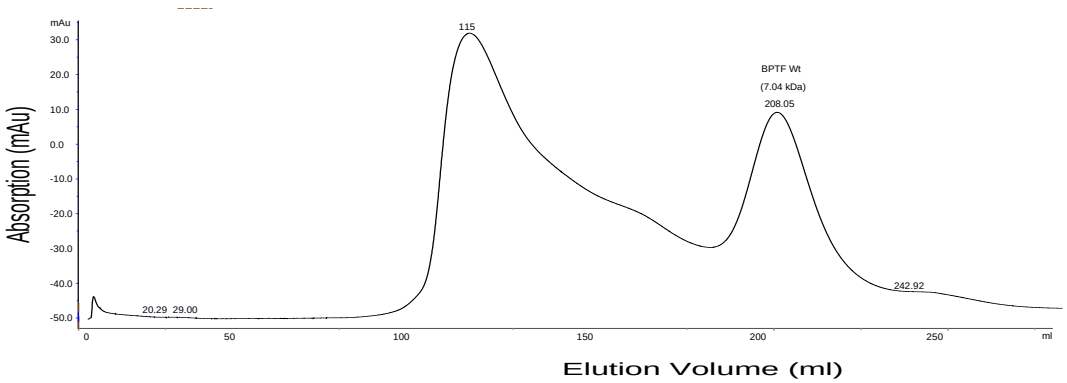


Figure 5.0: Chromatogram showing purification of wild type Histidine tagged BPTF through size exclusion chromatography (FPLC)

Synthetic peptides:

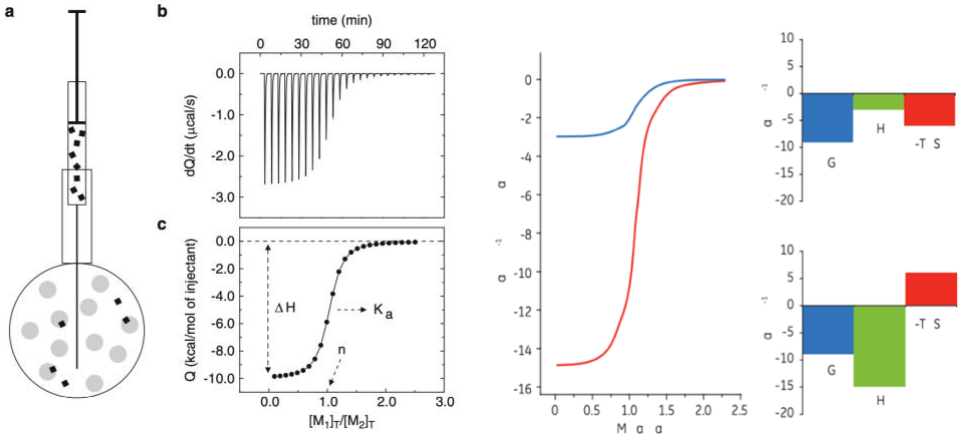
We have used 98% pure synthetic peptides for this section. We have used wild type Histone H3 1-11W and also mutated Histone H3 1-11W. For this project, we have purchased mutated histone peptides (Amino acid sequence: H3-AKTRQ, AKTKQ, ARTRQ) and H3-ARTKQ-wild type. We added W (Tryptophan) residue in the C-terminal of the synthetic peptide sequence for determining the peptide concentrations by absorbance using the computed molar extinction coefficient. We bought the peptides in lyophilized form from Genscript. We have prepared the working solution of the peptide (1mM) by dissolving the powder into 1xPBS buffer. We have used these histone peptides while doing ITCs to find out the reading activity of a PHD finger and its mutants.

Isothermal titration calorimetry (ITC):

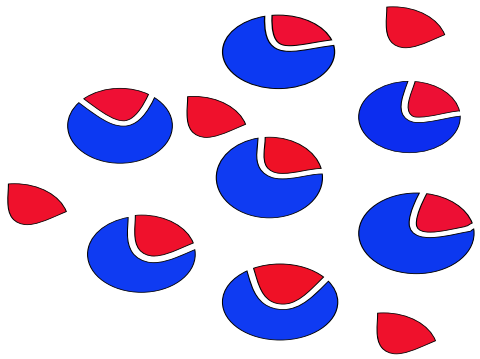
We have prepared the working solution of the peptide and the protein in the same 1xPBS buffer (25 mM phosphate, 150 mM sodium chloride at pH=7.6) for performing ITC. We have used MicroCal ITC200 (GE Healthcare) for studying binding energetics between protein (0.1-0.15 mM) and 98% pure commercial peptide (10-fold

higher than protein sample). We have loaded the protein sample in the sample cell and peptide in the syringe at 25°C. ITC was programmed to follow twenty 2µl- injections with an injection-interval of 3 minutes and a syringe stirring speed of 1000 rpm. Repeatedly the injections were performed and the resultant peaks getting smaller as the protein molecules getting saturated. Using the instrument software, each peak was being integrated at the end of the titration and being presented in the Wiseman plot. Stoichiometry (n), dissociation constant (K_D), enthalpy (ΔH), entropy (ΔS), and Gibb's free energy (ΔG) are calculated from each plot. Then association constant can be calculated from the K_D value ($K_A = 1/K_D$). At the room temperature, $\triangle G°$ of peptide binding was computed as $-RT\ln(K_A)$ where R, T and K_A are gas constant, temperature and association constant respectively or $\triangle G°$ can also be calculated by ΔH and ΔS values following the well-known equation: $\triangle G° = \triangle H - T\Delta S$. For the exothermic process, three cases can be possible: (1) $T\Delta S < 0$ when ΔG has a minimum negative value and the thermodynamic reaction is enthalpically driven. (2) $T\Delta S = 0$ when $\Delta G = \Delta H > 0$ and the reaction is thermodynamically driven, (3) $T\Delta S > 0$ when ΔG has a maximum negative value and the reaction can be enthalpically as well as entropically driven. In the molecular level, two fundamental processes can be concluded from this reaction: (1) trend of decreasing the energy (while binding occurs, $\Delta H < 0$ and (2) there is a thermal movement found (breaking multiple bonds, $\Delta S > 0$). A residue's energetic contribution can be calculated as $\triangle\triangle G° = \triangle G°$ mutant - $\triangle G°$ wildtype. Buffer and temperature conditions were maintained as same for all the titrations. In most of the cases, no binding or very negligible bindings were observed while mutants were being

tried to be titrated with the wild type histone H3 peptide. We have fitted the data using the MicroCal ITC Origin software latter. In these cases, △△G° (>> 1.5 kcal.mol-1) was assumed as large. For estimating the concentrations of protein, we have used Tryptophan residue for measuring the absorbance at 280 nm using the molar extinction coefficient of the PHD finger. Each set of binding study (titration) for each protein was repeated thrice to confirm the data, and to get the standard deviation, and mean value. ITC not only confirms whether there is any interaction possible between the ligand and the target molecules, it also confirms the nature of the noncovalent interaction. Polar interaction tends to contribute to the enthalpic component while hydrophobic interaction is more entropically favored (Figure 5.1).



A B



Blue = Protein molecule
Red = Peptide molecule (Ligand)

C

Figure 5.1: A.a) Schematic diagram of how ITC works (mechanism) b & c shows Standard curve of ITC. B. ITC shows different nature of binding mechanism. Blue curve signifies entropically favored interaction while red curve shows enthalpically favored interaction. C. Cartoon shows how proteins (macromolecules) bind with the peptides (Ligand).

Results and Discussions:

Significant changes in the cellular epigenetic state are expected from the expression and incorporation of the histone variants (Banaszynski LA 2010, Maze I 2014). Histones are known as DNA packaging agents. Apart from that, its another crucial function is to confer variations in the chromatin structure. And this is extremely important in order to determine the dynamic patterns of the transcriptional regulation that can control gene expression. Mutation can happen either on histone variants or on the histone readers. In both ways, the gene regulation can be affected. In this study, we have looked at both aspects. We wanted to see whether histone H3 peptides are still recognized by the readers even after having subsequent mutations on it. Histone tails are highly positively charged. So, there is a significant energetic contribution observed while readers supposed to be interacting with the mutated histone peptides. Thus, we are interested in further study of the characterization of the histone peptides, rich with Lys and Arg residues. We have used ITC for measuring the heat generated or absorbed while there is an interaction. Changes of heat are very small (submillionths of a degree) and were being measured by an extremely sensitive instrument of ITC. We have used four synthetic peptides for this study: (1) wild type histone H3 residues 1−11 (H3−1−11W-ARTKQ), (2) histone H3 residues 1−11 (AKTRQ), (3) histone H3 residues 1−11 (AKTKQ), (4) histone H3T3S residues 1−11 (H3T3S-1−11W). These

peptides were being used for the binding studies with the wild type histone reader PHD finger (KDM5B) to determine whether mutated histone H3 peptides are still getting recognized by the readers or not. And we observed (12kcal/mol) heat to be released while bonding formed between both the wild types of the protein and the peptide. Upon using the mutated peptide (ARTKQ- AKTKQ) where Arg was mutated to Lys, we observed a drastic change in the binding. We observed high value of K_D, which means a very less amount of heat (2kcal/mol) was released. But while mutated peptide (ARTKQ- AKTRQ) was being used as a ligand, we found there was a complete disruption of the binding (no heat was released at all). It clearly indicates that even though both Arg and Lys are positively charged amino acids, and most exposed to the surface of the protein, but Arg contributes more to the structure of the protein and it is very much position-specific and nature has perfectly and minutely designed it, upon altering the position-specific amino acid residue, binding disrupts (Figure 5.3). Histone H3 has few sequences features due to which it adopts helical conformation. These features are 1) the N-cap H3T3-OH (Aurora R 1998) 2) K9 positive charge for accommodating the helix dipole (Shoemaker KR 1987) and 3) H3G11, a helix that terminates with Gly (Gunasekaran K 1998). Histone H3 with N-cap H3T3-OH is found in the peptide complexes of many proteins like KDM5B, UHRF1(Arita K 2012, Bortoluzzi A 2017), KAT6A, DPF3(Li W 2016), MORF (Dreveny 2014, Xiong X 2016, Klein BJ 2017), and BAZ2A(Tallant C 2015).

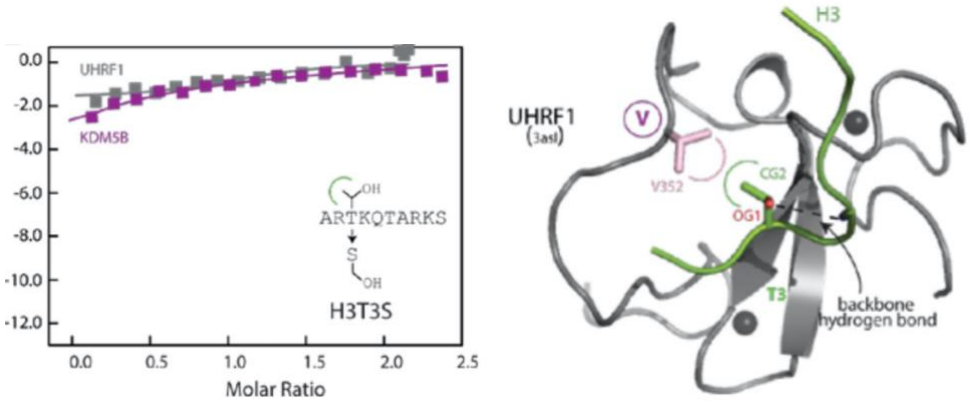


Figure 5.2: A. ITC binding curve between KDM4B (purple) or UHRF1 (gray) vs H3T3S. B. The cartoon shows the nonpolar contact between the V position of the UHRF1 (3asl) and the peptide H3T3 complex.

To determine the Van der Wal's interaction between the side chain of V position and CH3– group of the histone H3 peptide Thr3 residue (H3-Thr3), this Thr3 group has been replaced by Ser. Through ITC, we observed the significant changes in the binding, this substitution disrupted in the peptide-binding (Figure 5.2) and this is continuously been found the same while there is a nonpolar contribution with the V position. The purpose of this experiment was to show the significance of the tiny methyl group in the tight packing between the reader and the Histone-H3, Histone-H3T3S. By using the mutation of Threonine to Serine of the synthetic Histone-H3 peptide, we wanted to see what happen in the interface zone while a methyl group is missing.

Besides looking at the functional aspects of the mutated proteins, we also paid attention to the binding behavior of the mutated Histone H3 peptide because it plays a crucial role in chromatin deregulation. Mutation in the histone H3 peptide is also identified in several types of cancers like pediatric high-grade glioma, leukemia (Behjati 2013), giant cell tumors of bone, chondrosarcomas (Lowe RB. 2019), etc. We

have performed a series of ITC for Histone reader (wild type PHD finger-KDM5B) Vs wild as well as mutated Histone H3 peptide. From the ITC result, it has been found that wild type histone H3 peptide was properly read and bound with the wild type histone reader (PHD finger-KDM5B) as the binding interaction released 12kcal/mol heat that forms a sigmoidal curve. Whereas, binding completely disrupts when Arginine was mutated to Lysine and Lysine to Arginine in the peptide sequence. Slight heat was observed while only Arginine was mutated to Lysine and other amino acids of the peptide sequence were left intact (Figure 5.4).

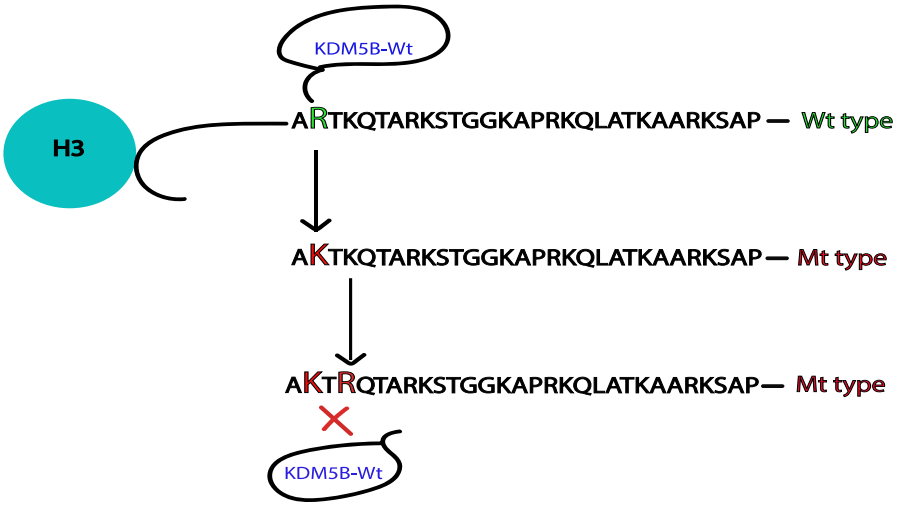


Figure 5.3: Diagram of the Arginine specific mutations on the Histone H3 peptide

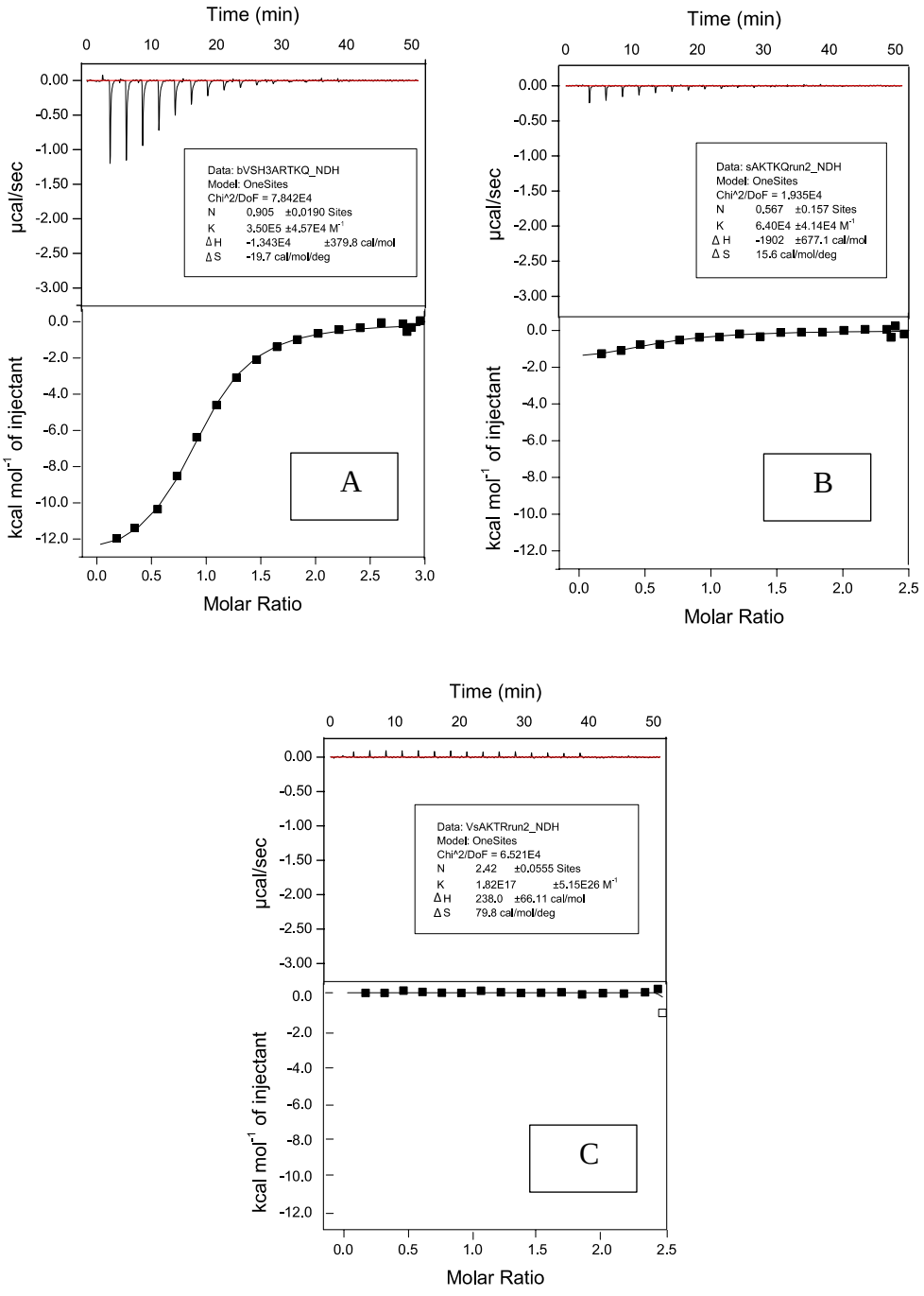
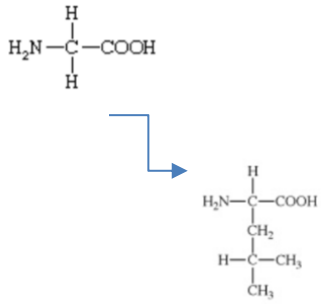
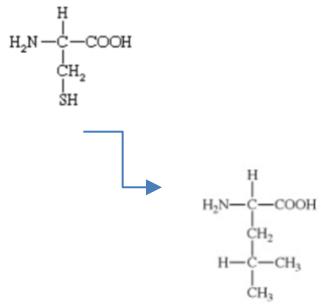
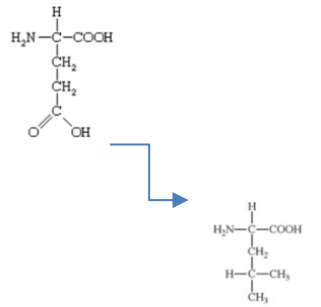
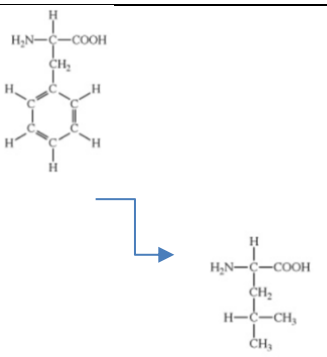


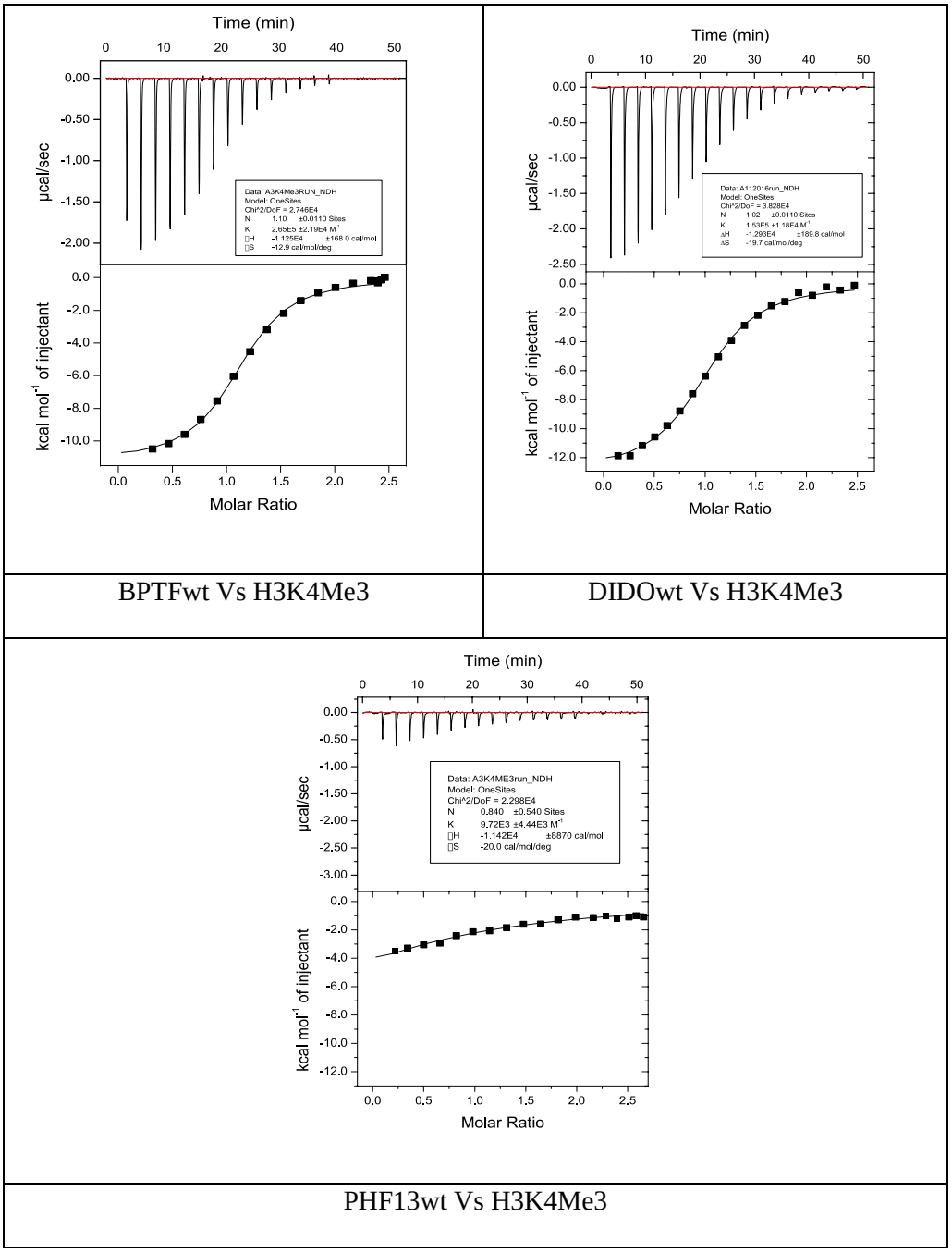
Figure 5.4: ITC binding studies: Compared overview of the binding interaction between wild type PHD finger (KDM5B) Vs wild and mutated Histone H3 peptides respectively. A. KDM5Bwt Vs H3ARTKQ; B. KDM5B wt Vs H3AKTKQ; C. KDM5Bwt Vs H3AKTRQ.

Compared overview of the Binding studies:

Table 1.8: List of mutations that contribute significantly in the binding studies of the PHD fingers

Protein Name	**Mutation**	**Chemical transition**	**Ligand**
BPTF	Glyvine (G) to Leucine (L)		Histone-H3K4Me3
DIDO	Cystein (C) to Leucine (L)		Histone-H3K4Me3
PHF13	Glutamic acid (E) to Leucine (L)		Histone-H3K4Me3
KAT6A DPF3 KDM5D	Phenylalanine (F) to Leucine (L)		Histone-H3K14AC

In the histone reader re-engineering, the side chain plays an important role in the packing density. We have selected nucleosome re-modeling factors like BPTF, DIDO, and PHF13 to see how mutations in these proteins are affecting their reading ability of the H3K4Me3 (Table 1.8). A strong heat release signal was noticed while ITC was performed with wild type BPTF/DIDO/PHF13 Vs H3K4Me3 peptide (Figure 5.5). But the binding affinity dropped (~4kcal/mol) while Glycine was mutated to Leucine. Usually a methyl group (-Ch3) plays important role in tight packing because of its stacking ability and polarizability (Sowers LC 1987, Wang S 1995) and also known for causing more stability because of its hydrophobic nature (Rossi M. 2013). But in this case, we have observed that upon adding methyl group as a side chain of leucine, the binding affinity of BPTF has been compromised. Contribution of Methyl group (-Ch3) in compromising the binding affinity has been observed again in the case of DIDO while cysteine residue was replaced by leucine. Negatively charged Glutamic acid is more important in tight packing than the methyl group, as we can see with the mutation of E to L leads to complete disruption of the binding. Complete disruption of the interface due to the absence of methyl group (Figure 5.6).



BPTFwt Vs H3K4Me3 | DIDOwt Vs H3K4Me3

PHF13wt Vs H3K4Me3

Figure 5.5: ITC binding studies showing binding affinity between H3K4Me3 Vs wild type BPTF/DIDO/PHF13.

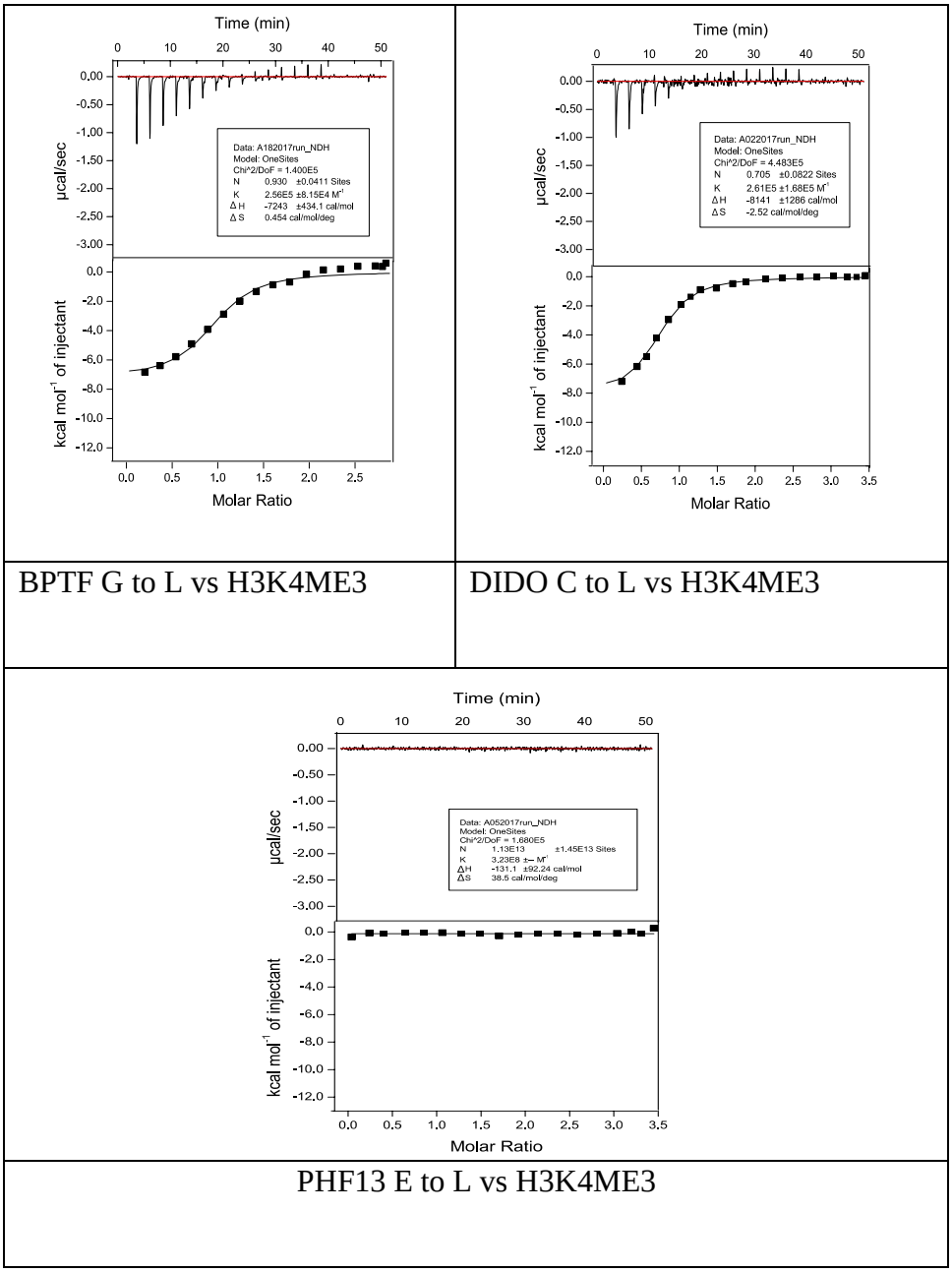
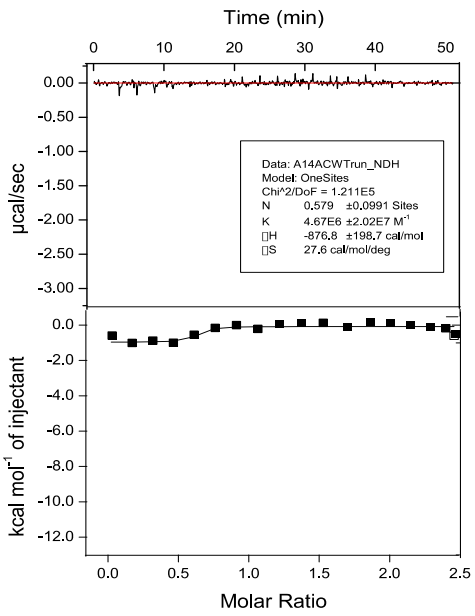


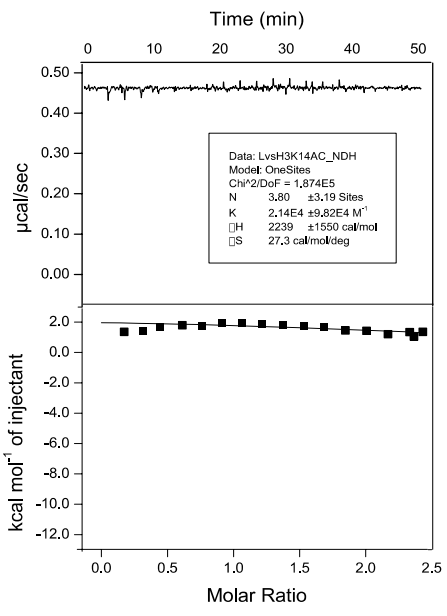
Figure 5.6: ITC binding studies showing contribution of the key factors in the tight packing

Since we are already trying to characterize and understand the mechanism of the PHD fingers and chromatin signaling, therefore, we found interest to dig into the detailed mechanism of the histone acetylation and how do mutations in the readers can affect the recognition of the acetylated histones. We have started investigating this phenomenon with the three proteins of the PHD finger subtype PHD_nW:

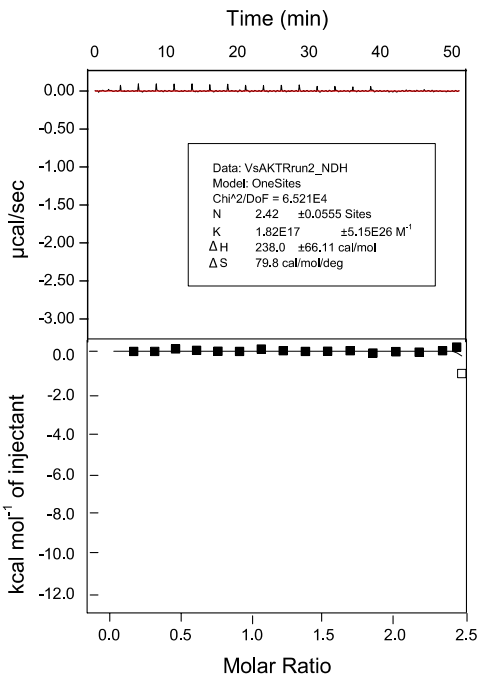
KAT6A, DPF3, and KDM5D respectively, where we have replaced the phenylalanine (F) with leucine (L) as phenylalanine contains aromatic ring which might be contributing in the tight packing. We wanted to see even after making mutations in the protein part, whether they can still bind to the acetylated histone. We found that upon making a mutation of phenylalanine to leucine, these proteins were unable to bind to the acetylated histone peptide with the same strength as we have found their binding affinity when there was no alteration in their amino acid sequences. Without the aromatic ring, binding completely falls off (Figure 5.7).



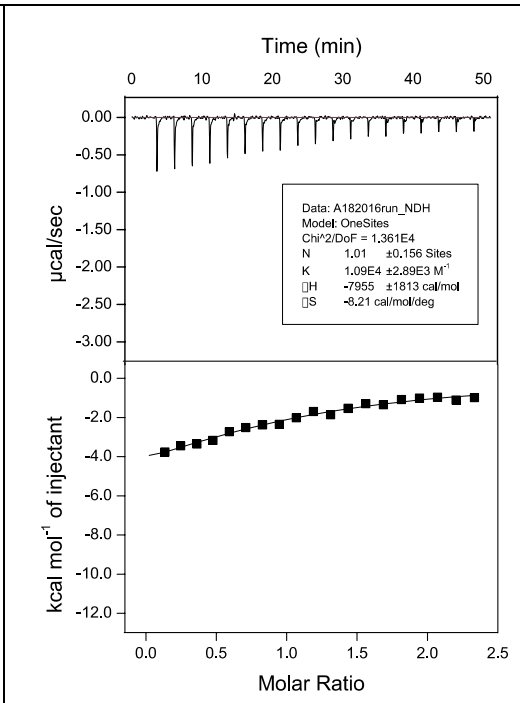
A: ITC: KAT6A F to L vs H3K14AC



B: ITC: KDM5D F to L vs H3K14AC



C: ITC: DPF3 F to L vs H3K14AC



D: ITC: KAT6A wt vs H3K14AC

Figure 5.7: ITC binding studies: A-C: There are no binding when ITCs were performed for F to L mutated PHD fingers vs H3K14AC. We have observed 4 kcal/mol heat was generated while H3K14AC was recognized by the wild type KAT6A.

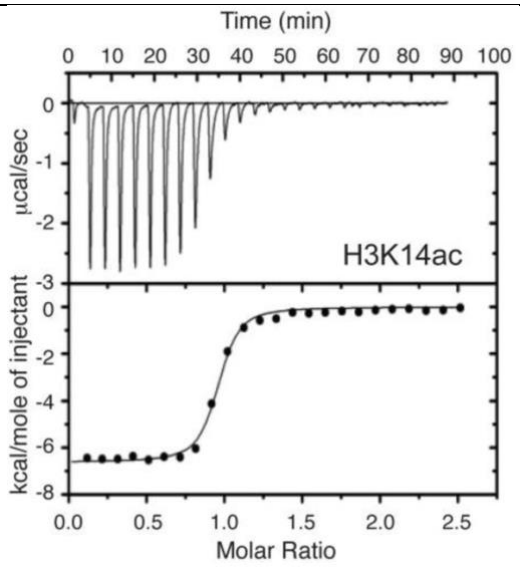


Figure 5.8: ITC showing interaction between DPF3 Wt-PHD finger and H3K14AC (Zeng L. 2010).

Previous researchers have found that 7kcal/mol heat was generated while H3K14AC was recognized by the wild type DPF3 (Zeng L. 2010) (Figure 5.8). But when we made F to L mutation, we didn't notice any binding. Therefore, we can conclude that the aromatic ring of the phenylalanine was contributing to the tight packing, upon replacing it with leucine, binding was completely disrupted.

CHAPTER V

Case study of the weak interaction using QM-PDZ domain as a model

Abstract

Interactions in which aromatic functional groups are involved, are found to be an important component in the weak interaction and weak interaction is critical in the regulation of the protein function. Thus, it's getting more attention in the field of chemistry, such as for the designing of the catalysts. It has been also found that weak interaction is also important for the folding and stability of the protein. One of the most prominent weak interactions is anion-quadrupole interaction in which interaction takes place between anions and the aromatic rings of the proteins. Certain anions like aspartate, and glutamate were being chosen to see the pattern of interaction with the amino acids like Tyr, Phe, and Trp that have the aromatic ring in their structures. We have studied the binding energetics of the weak interaction of the anion-quadrupole interaction because this is related to many disorders. Compared to hydrogen bonding and salt bridges, we wanted to measure the strength of the anion-quadrupole interaction in the interphase binding, which has been less studied so far. We have created a library

of proteins out of which we have cloned QM protein and made mutations on it to see the differences in the binding energetics by ITC. We didn't notice many changes in the binding energetics of the wild and mutant proteins. It suggests that sometimes amino acids readjust to regain the lost energy. Therefore, we felt the necessity of the structural analysis further and crystallization can help us to better understand that.

Introduction

Noncovalent interactions like hydrogen bonds, ionic interaction, hydrophobic interaction, and dispersion forces play a significant role in the biological structure, function, and stability (Dougherty 1996, Brandl 2001, Jiang 2002, Ringer 2007). Noncovalent interactions can be found in both inter and intramolecular interactions. Recent studies have shown that there is another type of noncovalent interaction that also plays a crucial role in the biological system (Sinnokrot 2006, Bartlett 2010, Sippel 2015, Kapoor K. 2016). This interaction is found to occur between the positively charged ring edge of aromatic groups and negatively charged amino acid sidechains (Figure 5.9). The edgewise positive charges of these groups are found to be associated with the quadrupole moment of the aromatic rings. Thus it is termed as anion-quadrupole interaction (Jackson 2007). From the previous study, it has been found that this kind of interaction can be stabilizing and comparatively more infrequent while involving with the isolated anions (Alkorta 2002, Mascal 2002, Quinonero 2002). This type of interaction can provide at least ~0.5 kcal/mol additional stability to the helix (Olson 2001, Shi 2002).

We have studied anion-quadrupole interaction as a reference of the weak interaction that plays an important role in protein structure and stability. It helps in macromolecular folding, assembly, and recognition. Thus, it has been found to be associated with several human disorders. Lot of disease mutations that disrupt anion-

quadrupole, have been mapped. But to compare with all diseases, we also made a reference set of proteins. So, in the reference set, we wanted to see how the anion-quadrupole behaves if there is a mutation in the protein or the peptide part. We wanted to compare the interactions of these proteins with the proteins where there are mutations denoted. Some anion-quadrupoles are weak by nature and don't have much effects while some have lot of effects, and that's why they cause the disease. Proteins Causing disease means they are losing their structure as well as function. For the convenience of the experiment, we categorized the proteins into two sets: one that causes disease and another set where we make mutations at the interface, so that using ITC we can quickly check the interaction.

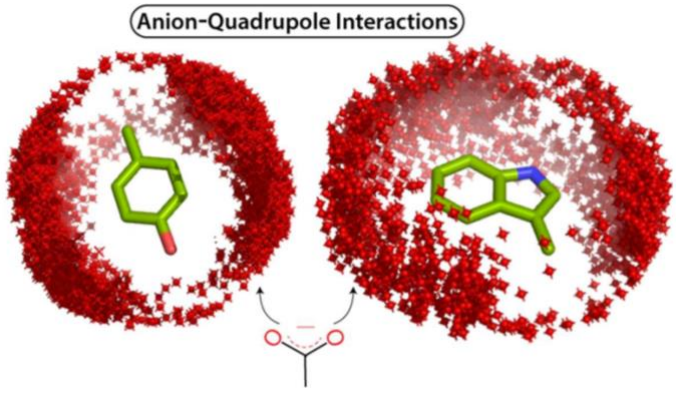


Figure 5.9: Anion-Quadrupole interaction between anions and Phenylalanine of the aromatic ring of a protein (Chakravarty S 2018)

Therefore, we have started working on thermodynamic quantification of the anion-quadrupole interaction. As a part of this project, we have started the same with QM protein interacting with PDZ domain in order to find out the pattern of the interaction between the Phenylalanine of the QM protein with the anions of the PDZ domain.

Material and Methods

DNA Construct:

Synthetic DNA of QM wild type (Table 1.9) was bought from Genscript and was cloned into Ampicilin resistant pGEX4T3 vector using the primers bought from IDT technologies. Then E912Q, E912D, and E912A mutations were created in the QM using the respective mutated primers (Table 2.0) by site directed mutagenesis. Both cloning and site directed mutagenesis were done following the protocol discussed in the chapter II.

Table 1.9: Nucleotide sequence and amino acid sequence of the QM protein and its corresponding primers.

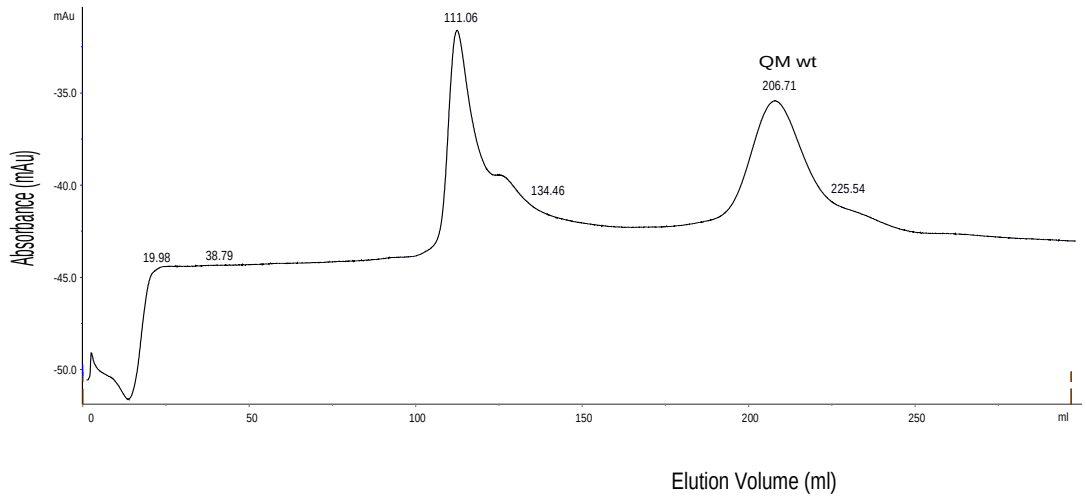
QM mRNA Sequence	AAAGTCACTCACAGCATCCACATTGAGAAGTC AGATACAGCTGCTGATACTTACGGGTTTTCACT TTCTTCTGTGGAAGAAGATGGTATTCGAAGGC TGTACGTGAATAGTGTGAAGGAAACCGGTTTA GCTTCCAAGAAAGGCCTGAAAGCAGGAGATGA GATTCTTGAGATCAATAATCGTGCTGCTGACGC CCTGAACTCTTCTATGATGGAAGATTTCTTCTC ACAACCCTCGGTGGGCCTCCTGGTGAGGACCT ACCCCGAGCTG
QM Protein Sequence	KVTHSIHIEKSDTAADTYGFSLSSVEEDGIRRLYV NSVKETGLASKKGLKAGDEILEINNRAADALNSS MMEDFFSQPSVGLLVRTYPEL
Primer	Primer sequence
Forward Primer (5’ to 3’)	CGCGGATCCAAAGTCACTCACAGCATCC
Reverse Primer (5’to 3’)	CCGCTCGAGTTACAGCTCGGGGTAGGTCC

Table 2.0: Protein sequence of the mutants created on QM and its corresponding mutant primers

Mutant	**Forward Primer (5'- 3')**	**Reverse Primer (5'-3')**	**Protein Sequence**
E912Q	TGAACTCTTCTA TGATGCAAGATT TCTTCTCACAAC C	GGTTGTGAGAAGA AATCTTGCATCAT AGAAGAGTTCA	KVTHSIHIEKSDTAAD TYGFSLSSVEEDGIRR LYVNSVKETGLASKK GLKAGDEILEINNRAA DALNSSMMQDFFSQP SVGLLVRTYPEL
E912A	TGAACTCTTCTA TGATGGCAGATT TCTTCTCACAAC C	GGTTGTGAGAAGA AATCTGCCATCAT AGAAGAGTTCA	KVTHSIHIEKSDTAAD TYGFSLSSVEEDGIRR LYVNSVKETGLASKK GLKAGDEILEINNRAA DALNSSMMADFFSQP SVGLLVRTYPEL
E912D	TGAACTCTTCTA TGATGGATGATT TCTTCTCACAAC C	GGTTGTGAGAAGA AATCATCCATCAT AGAAGAGTTCA	KVTHSIHIEKSDTAAD TYGFSLSSVEEDGIRR LYVNSVKETGLASKK GLKAGDEILEINNRAA DALNSSMMDDFFSQP SVGLLVRTYPEL

Purification of Protein

In this project we have synthesized, and purified GST tagged wild and mutant type QM protein following the column digestion method. All the steps till binding to the beads are same for this protein as mentioned in chapter II. After at least 3-4 hours of binding, we washed the beads-protein couple of times with 1xPBS followed by 1M NaCl wash and finally couple of times wash with 1xPBS. Then we have continued purification with the column digestion method using Thrombin protease for overnight (7-8 hours). After overnight thrombin digestion, protein supposed to leave the GST protein which remains bound to the beads. Then the protein sample was run through FPLC and purified (Figure 6). We have purified QM wild type and mutants following the above method (Figure 6.1). After that it was further concentrated to 0.15mM which was used further for performing ITC for thermodynamic study. Protein quality was determined by SDS PAGE gel (15%).



`Figure 6: Chromatogram showing purification of wild type GST-QM wild type through size exclusion chromatography (FPLC): purified by column digestion method

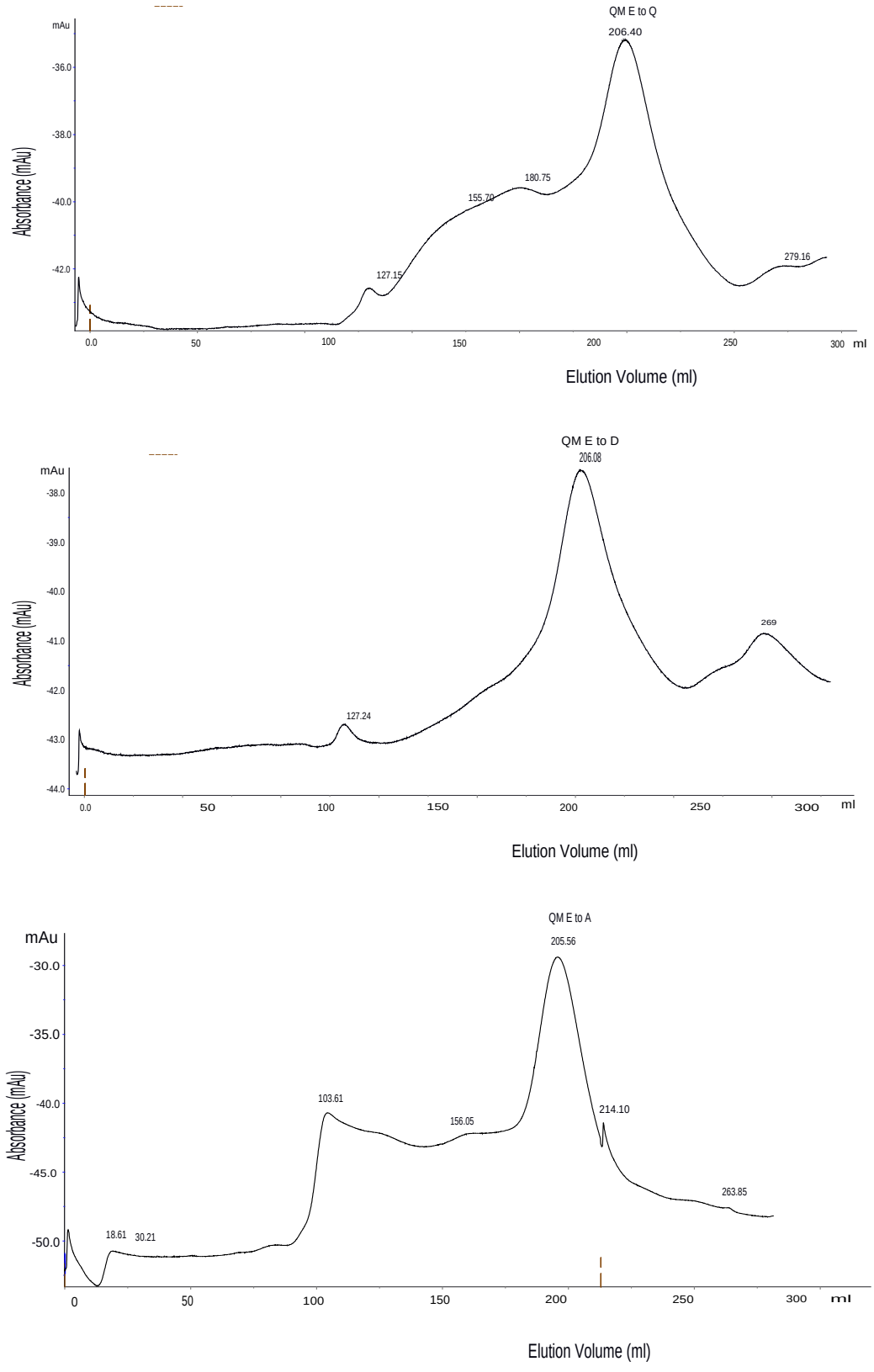


Figure 6.1: Chromatogram showing FPLC profiles of QM mutants (E to Q, E to D and E to A) purification

Synthetic peptides:

We have bought commercial PDZ domain (lyophilized) from Genscript and diluted it to 1mM working solution using 1xPBS buffer. We have used this PDZ domain while doing titration using ITC for determining the binding activity of the wild type QM protein as well as its mutants E912A, E912D, and E912Q.

Result and Discussion:

We have performed ITC for thermodynamic quantification of the anion quadrupole interaction. Anion quadrupole interaction is an edgewise interaction between anions and phenylalanine (Phe) aromatic rings in proteins (Chakravarty S 2018). Though it was an example of weak interaction, but it has been found to be associated with many disorders as weak interaction is highly important for regulating protein function in several ways. We have started this experiment with QM protein. We have cloned it in pGEX4T3 vector, expressed it and finally purified it. We also made mutations on it (E to A, E to Q, and E to D). Finally, we have performed ITC to observe the interaction between the proteins (phenylalanine (Phe) aromatic rings) and anions of PDZ domain. From ITC, though we didn't see any remarkable thermodynamic changes in wild as well as mutants of QM protein while they were interacting with PDZ domain. Almost equal amount of heat was generated in either case of wild and mutants. We have other proteins in a row which also have phenylalanine (Phe) aromatic rings, with which we are planning to see thermodynamic changes through ITC while interacting with anions of PDZ domain (Figure 6.2).

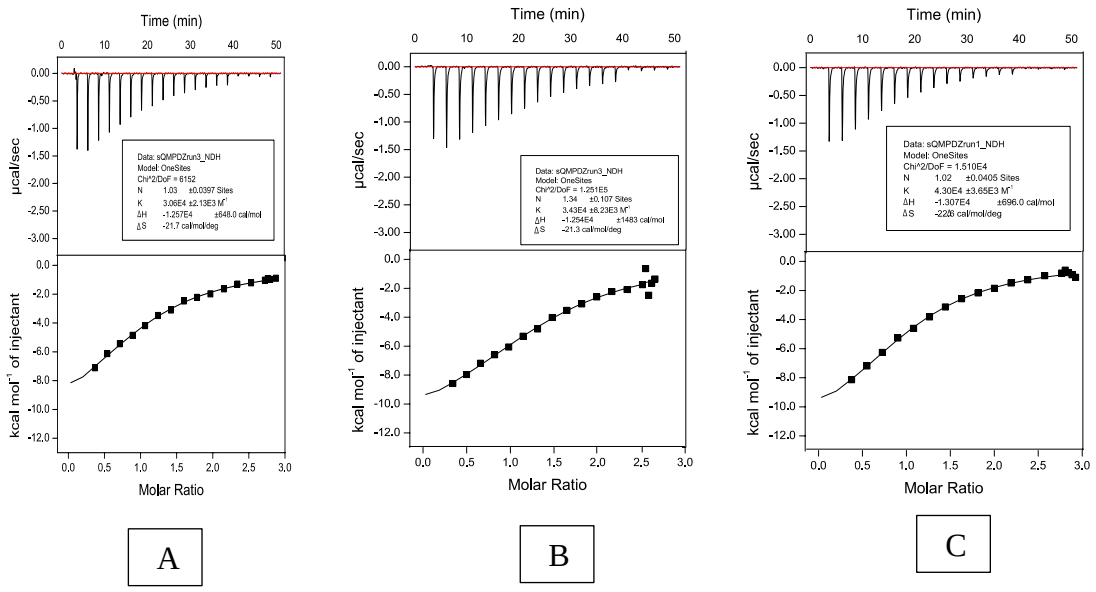


Figure 6.2: Thermodynamic quantification of anion-quadrupole interaction through Isothermal titration calorimetry (ITC) A. ITC showing binding interactions between QM wt Vs PDZ domain. B. ITC showing binding interactions between QM E to A Vs PDZ domain. C. ITC showing binding interactions between QM E to Q Vs PDZ domain.

Bibliography

Alexandrescu A. T., E. P. A., Pitkeathly M., Baum J., Dobson C. M. (1993). "Structure and dynamics of the acid-denatured molten globule state of .alpha.-lactalbumin: a two-dimensional NMR study." Biochemistry **32**(7): 1707-1718.

Alkorta, I., Rozas, I., and Elguero, J. (2002). "Interaction of anions with perfluoro aromatic compounds." Journal of the American Chemical Society **124**(29): 8593-8598.

Allis, C. D., and Jenuwein, T.` (2016). "The molecular hallmarks of epigenetic control." Nat Rev Genet **17**: 487-500.

Arita K, I. S., Oda T, Unoki M, Sugita K, Sekiyama N, Kuwata K, Hamamoto R, Tochio H, Sato M, Ariyoshi M, Shirakawa M. (2012). "Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1." Proceedings of the National Academy of Science of the United States of America **109**(32).

Arnold JT DSS, P. M. (1957). "Chemical effects on nuclear-induction signals from organic compounds." Journal of Chemical Physiology **19**: 507.

Asakura T, I. M., Demura M, Williamson MP. (1999). "Structural analysis of silk with C-13 NMR chemical shift contour plots." International Journal of Biological Macromolecule **24**: 167–171.

Asakura T, T. K., Demura M, Williamson MP. (1995). "The Relationship Between Amide Proton Chemical Shifts and Secondary Structure in Proteins." Journal of Biomolecular NMR **6**: 227–236.

Aurora R, R. G. (1998). "Helix capping." Protein Science **7**(1): 21-38.

Bai Y., M. J. S., Mayne L., Englander S. W. (1993). "Primary structure effects on peptide group hydrogen exchange." Proteins: Structure, Function, and Bioinformatics **17**(1): 75-86.

Baldwin J.A., K. E. L. (2009). "NMR spectroscopy brings invisible protein states into focus." Nature Chemical Biology **5**: 808-814.

Banaszynski LA, A. C., Lewis PW. (2010). "Histone variants in metazoan development." Developmental Cell **19**(5): 662-674.

Bartlett, G. J., Choudhary, A., Raines, R. T., and Woolfson, D. N. (2010). "n– > pi* interactions in proteins." Nature Chemical Biology **6**(8): 615-620.

Baxter NJ, W. M. (1997). "Temperature dependence of 1H chemical shifts in proteins." Journal of Biomolecular NMR **9**: 359-369.

Behjati, S. T., P.S.; Presneau, N.; Scheipl, S.; Pillay, N.; Van Loo, P.; Wedge, D.C.; Cooke, S.L.; Gundem, G.; Davies, H.; (2013). "Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor

of bone." Nat Genet **45**(12): 1479-1482.

Bienz, M. (2006). "The PHD finger, a nuclear protein-interaction domain." Trends of Biochemical Science **31**: 35-40.

Boamah D, L. T., Poppinga AF, Basu S, Rahman S, Essel F, and Chakravarty S (2018). "Characteristics of a PHD Finger Subtype." Biochemistry **57**: 525-539.

Bortoluzzi A, A. A., Lucas X, Blank M, Ciulli A. (2017). "Structural basis of molecular recognition of helical histone H3 tail by PHD finger domains. ." Biochemical Journal **474**(10): 1633-1651.

Brandl, M., Weiss, M. S., Jabs, A., Suhnel, J., and Hilgenfeld, R. (2001). "**C-H···pi-interactions in proteins.**" Journal of Molecular Biology **307**(1): 357-377.

Bullock AN, H. J., Fersht AR (2000). "Quantitative analysis of residual folding and DNA binding in mutant p53 core domain: Definition of mutant states for rescue in cancer therapy." Oncogene **19**(10): 1245–1256.

Burstein, E. A. (1976). "Luminescence of protein chromophores: model studies. In Advances in Science and Technology, Ser." Biophysical Journal **6**.

Capili, A. D., Schultz, D. C., Rauscher, I. F., and Borden, K. L (2001). "Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains." The EMBO Journal **20**(1): 165-177.

Cavagnero S., N. C., Schwarzinger S., Dyson H. J., Wright P. E. (2001). "Conformational and Dynamic Characterization of the Molten Globule State of an Apomyoglobin Mutant with an Altered Folding Pathway." Biochemistry **40**(48): 14459-14467.

Chakravarty S, E. F., Lin T, and Zeigler S (2015). "Histone Peptide Recognition by KDM5B-PHD1: A Case Study." Biochemistry **54**: 5766−5780.

Chakravarty S, U. A., Moore B, Shore J, Alshamrani M (2018). "A Comprehensive Analysis of Anion-Quadrupole Interactions in Protein Structures." Biochemistry **57**(12): 1852-1867.

Chamberlain A. K., M. S. (1998). "Molten Globule Unfolding Monitored by Hydrogen Exchange in Urea." Biochemistry **37**(7): 1736-1742.

Chatterjee, K., Tripathi, V., Das, R (2019). "A conserved and buried edge-to-face aromatic interaction in small ubiquitin-like modifier (SUMO) has a role in SUMO stability and function." Journal of Biological Chemistry **294**(17): 6772–6784.

Cornilescu G, D. F., Bax A. (1999). "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." Journal of Biomolecular NMR **13**: 289–302.

Cox HE, a. M. L. (2000). "Zinc-dependent protein folding." Current Opinion in Chemical Biology **4**(2): 162-165.

Croce, C. M. (2008). "Oncogenes and cancer." The New England Journal of Medicine **358**(5): 502-511.

Dalgarno DC, L. B., Williams RJ. (1983). "Structural information from NMR secondary chemical shifts of peptide alpha C-H protons in proteins." Biosci Rep **3**: 443–452.

Dill KA, S. D. (1991). "Denatured states of proteins." Annual review of Biochemistry **60**(1): 795-825.

Dobson, C. M., Sali, A. & Karplus, M. (1998). " Protein folding: a perspective from theory and experiment." Angew Chem Int Ed Engl. **37**(7): 868-893.

Dougherty, D. A. (1996). "Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp." Science **271**(5246): 163-168.

Dreveny, I., Deeves, S. E., Fulton, J., Yue, B., Messmer, M., Bhattacharya, A., Collins, H. M., and Heery, D. M. (2014). "The double PHD finger domain of MOZ/MYST3 induces alpha-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification." Nucleic Acid Research **42**: 822-835.

Duy C, a. F. J. (2006). "How Aggregation and Conformational Scrambling of Unfolded States Govern Fluorescence Emission Spectra." Biophysical Journal **90**: 3704-3711.

Dyson H. J., W. P. E. (1998). "Equilibrium NMR studies of unfolded and partially folded proteins." Nature Structural Biology **5**: 499-503.

Dyson JH., W. E. (2004). "Unfolded Proteins and Protein Folding Studied by NMR." Chemical Review **104**: 3607-3622.

E., O. (1995). "Chemical Shifts and Three-Dimensional Protein Structures." Journal of Biomolecular NMR **5**: 17– 225.

Farrow N. A., M. R., Singer A. U., Pascal S. M., Kay C. M., Gish G., Shoelson S. E., Pawson T., Forman-Kay I. D., Kay L. E. (1994). "Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation." Biochemistry(19): 5984–6003.

Friedler A, H. L., Veprintsev DB, Freund SM, Rippin TM (2002). "A peptide that binds and stabilizes p53 core domain: Chaperone strategy for rescue of oncogenic mutants." Proceedings of the National Academy of Science of the United States of America **99**(2): 937-942.

Greenfield, N. J. (2006). "Using circular dichroism spectra to estimate protein secondary structure." Nature Protocol **1**(6): 2876-2890.

Greenfield NJ, F. G. (1969). "Computed circular dichroism spectra for the evaluation of protein conformation." Biochemistry **8**(10): 4108-4116.

Greenman, C., Stephens,P., Smith,R., Dalgliesh,G.L., Hunter,C., and G. Bignell, Davies,H., Teague,J., Butler,A., Stevens,C (2007). "Patterns of somatic mutation in human cancer genomes." Nature **446**: 153-158.

Gunasekaran K, N. H., Ramakrishnan C, Balaram P (1998). "Stereochemical punctuation marks in protein structures: glycine and proline containing helix stop signals." Journal of Molecular Biology **275**(5): 917-932.

He, F., Umehara, T., Saito, K., Harada, T., Watanabe, S., Yabuki, T., Kigawa, T., Takahashi, M., Kuwasako, K., Tsuda, K., Matsuda, T., Aoki, M., Seki, E., Kobayashi, N., Guntert, P., Yokoyama, S., and Muto, Y. (2010). "Structural insight into the zinc finger CW domain as a histone modification reader." Structure **18**(9): 1127-1139.

Hecht, M. H., Sturtevant, J. M., and Sauer, R. T. (1984). "Effect of single amino acid replacements on the thermal stability of the NH2-terminal domain of phage lambda repressor." Proceedings of the National Academy of Science of the United States of America **81**(18): 5685-5689.

Hellman M, T. H., Rappu P, Heino J, Permi P (2011). "Characterization of Intrinsically Disordered Prostate Associated Gene (PAGE5) at Single

Residue Resolution by NMR Spectroscopy." PLoS ONE **6**(11): 26633.

Henzler-Wildman K., K. D. (2007). "Dynamic personalities of proteins." Nature **450**: 964-972.

Hoelen H, K. B., Schmidt A, Richardson J, Charitou P. (2010). "The primary folding defect and rescue of DeltaF508 CFTR emerge during translation of the mutant domain." PLoS ONE **5**(11): e15458.

Hoffman CR, H. J., and Klevit ER (1993). "Structures of DNA-binding mutant zinc finger domains: Implications for DNA binding " Protein Science **2**: 951-965.

Holzwarth G, D. P. (1965). "The ultraviolet circular dichroism of polypeptides." Journal of the American Chemical Society **87**(2): 218-228.

Jackson, M. R., Beahm, R., Duvvuru, S., Narasimhan, C., Wu, J., Wang, H.-N., … Howell, E. E. (2007). "A Preference for Edgewise Interactions between Aromatic Rings and Carboxylate Anions: The Biological Relevance of Anion−Quadrupole Interactions." The Journal of Physical Chemistry B **111**(28): 8242-8249.

Jiang, L., and Lai, L (2002). "CH···O hydrogen bonds at proteinprotein interfaces." Journal of Biological Chemistry **277**: 37732-37740.

Jones, S., Hruban,R.H., Kamiyama,M., Borges,M., Zhang,X., Parsons,D.W., Lin,J.C., Palmisano,E., Brune,K., Jaffee,E.M. (2009). "Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene." Science **324**: 217.

Kapoor K., D. M., Upadhyay A., Bucci JC., Saxton AM., Hinde RJ., Howell EE., Baudry J. (2016). "Highly Dynamic Anion-Quadrupole Networks in Proteins." Biochemistry **55**(43): 6056-6069.

Karpen, M. E., de Haseth, P. L., and Neet, K. E. (1992). "Differences in the amino acid distributions of 3(10)-helices and alpha-helices." Protein Science : A Publication of the Protein Society. **1**(10): 1333-1342.

Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G. A., Li, Z., Vedadi, M., Schapira, M., Min, J., and Arrowsmith, C. H. (2011). "Recognition and specificity determinants of the human cbx chromodomains." Journal of Biological Chemistry **286**(1): 521-529.

Klechner IR., F. M. (2011). "An introduction to NMR-based approaches for measuring protein dynamics." Biochem Biophys Acta **1814**(8): 942-968.

Klein BJ, S. J., Wang X, Ahn J, Andrews FH, Zhang Y, Cote J, Shi X, Garcia BA, Kutateladze TG. (2017). "Recognition of Histone H3K14 Acylation by MORF. ." Structure **25**(4): 650-654.

Kumar A, S. S., Hosur RV. (2007). "NMR characterization of the energy landscape of SUMO-1 in the native-state ensemble." Journal of Molecular Biology **367**(5): 1480-1493.

Kunkel, T. A. (1985). Proceedings of the National Academy of Sciences U.S.A. **82**(2): 488-492.

L., S. (1995). "Chemical Shifts in Proteins Come of Age." Prog Nucl Magn Reson Spectros **27**: 325-443.

Larry R. Masterson, L. S., Marco Tonelli, Alessandro Mascioni, Michael M. Mueller, and Gianluigi Veglia (2009). "Backbone NMR resonance assignment of the catalytic subunit of cAMP-dependent protein kinase A in complex with AMP-PNP." Biomol NMR Assign **3**(1): 115-117.

Lee W, T. M., Markley JL (2015). "NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy." Bioinformatics **31**: 1325-1327.

Li R. H., W. C. (1999). "The hydrogen exchange core and protein folding." Protein Science **8**(8): 1571-1590.

Li, W., Jaroszewski, L., and Godzik, A. (2002). "Tolerating some redundancy significantly speeds up clustering of large protein databases." Bioinformatics **18**(1): 77-82.

Li W, Z. A., Tempel W, Loppnau P, Liu Y. (2016). "Crystal structure of DPF3b in complex with an acetylated histone peptide." Journal of Structural Biology **195**(3): 365-372.

Li, W., Zhao, A., Tempel, W., Loppnau, P., and Liu, Y. (2016). "Crystal structure of DPF3b in complex with an acetylated histone peptide." Journal of Structural Biology **195**(3): 365-372.

Lin S, G. U., and Kobilka KB. (1996). "Ligand Stabilization of the B2 Adrenergic Receptor: Effect of DTT on Receptor

Conformation Monitored by Circular Dichroism and Fluorescence Spectroscopy." Biochemistry **35**(46).

Low LY, H. H., Robinson CV, O'Brien R, Grossmann JG, Ladbury JE, Luisi B (2002). "Metal-dependent folding and stability of nuclear hormone receptor DNA-binding domains." Journal of Molecular Biology **319**(1): 87-106.

Lowe RB., M. A., Hamey JJ., Wilkins RM., Partridge J. (2019). "Histone H3 Mutations: An Updated View of Their Role in Chromatin Deregulation and Cancer." Cancers (Basel) **11**(5).

Luger K, M. A., Richmond RK, Sargent DF, Richmond TJ (1997). "Crystal structure of the nucleosome core particle at 2.8-Å resolution. ." Nature Methods **389**(6648): 251–260.

M.C., D. (2003). "Protein folding and misfolding." Nature **426**(6968): 884-890.

Mascal, M., Armstrong, A., and Bartberger, M. D. (2002). "Anionaromatic bonding: a case for anion recognition by pi-acidic rings." Journal of the American Chemical Society **124**(22): 6274-6276.

Mayer S, R. d. S., Ang HC, Joerger AC, Fersht AR (2007). "Correlation of levels of folded recombinant p53 in escherichia coli with thermodynamic stability in vitro." Journal of Molecular Biology **372**(1): 268–276.

Maze I, N. K., Soshnev AA, Allis CD. (2014). "Every amino acid matters: Essential contributions of histone variants to mammalian development and disease." Nature Reviews Genetics **15**: 259-271.

Minde DP, M. M., Ru¨diger SGD (2012). "Determining Biophysical Protein Stability in Lysates by a Fast Proteolysis Assay, FASTpp." PLoS ONE **7**(10): 0046147.

Mittag T, F.-K. J. (2007). "Atomic-level characterization of disordered protein ensembles." Current opinion in structural biology **17**(1): 3-14.

Moskowitz MS, C. F., Sternen LD, Cheng E, Gibson LR, Marshall GS & Cutting RG (2008). "Clinical practice and genetic counseling for cystic fibrosis and CFTR-related disorders." Genetics in Medicine **10**: 851-868.

MP, W. (1990). "Secondary-Structure Dependent Chemical Shifts in Proteins." Biopolymers **29**: 1428-1431.

Murata, H. C., C. S.; Koepsel, R. R.; Russell, A. J. (2013). "Polymer-Based Protein

Engineering Can Rationally Tune Enzyme Activity, pH-Dependence, and Stability. ." Biomacromolecules **14**: 1919-1926.

Musselman AC, a. K. T. (2009). "PHD Fingers: Epigenetic Effectors and Potential Drug Targets." Molecular Intervention **9**(6): 314-323.

Musselman CA, L. M., Côté J, and Kutateladze TG (2012). "Perceiving the epigenetic landscape through histone readers." **19**(12): 1218-1227.

Olson, C. A., Shi, Z., and Kallenbach, N. R. (2001). "Polar

interactions with aromatic side chains in alpha-helical peptides: Ch···OH-bonding and cation-pi interactions." Journal of the American Chemical Society **123**(26): 6451-6452.

Orekhov V. Y., P. K. V., Korzhnev D. M., Arseniev A. S. (1995). "Backbone dynamics of (1-71)- and (1-36)bacterioopsin studied by two-dimensional (1)H- (15)N NMR spectroscopy." Journal of Biomolecular NMR **6**(2): 113-122.

Osapay K, C. D. (1991). "A New Analysis of Proton Chemical Shifts in Proteins." Journal of American Chemical Society **113**: :9436–9444.

Park C, M. S. (2005). "Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding." Nature Methods **2**(3): 207-212.

Pastore A, S. V. (1990). "The Relationship Between Chemical Shift and Secondary Structure in Proteins." **90**: 165–176.

PC3267 (2007). "Fluorescence Spectroscopy: A tool for Protein folding/unfolding Study." Biophysics.

Proctor WG, Y. F. (1950). "The dependence of a nuclear magnetic resonance frequency upon chemical compound." Phy Review **77**: 717.

Qiu Y, L. L., Zhao C, Han C, Li F, Zhang J, Wang Y, Li G, Mei Y, Wu M, Wu J, and Shi Y (2012). "Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription." Genes & Development, American Association of Cancer Research **26**: 1376-1391.

Quinonero, D., Garau, C., Rotger, C., Frontera, A., Ballester, P., Costa, A., and Deya, P. M (2002). "Anion-pi Interactions: do they exist?" Angew International Edition Chemie. **41**(18): 3389-3392.

RainerJaenicke (2000). "Stability and stabilization of globular proteins in solution." Journal of Biotechnology **79**(3): 193-203.

Rajakumara, E., Wang, Z., Ma, H., Hu, L., Chen, H., Lin, Y., Guo, R., Wu, F., Li, H., Lan, F., Shi, Y. G., Xu, Y., Patel, D. J., and Shi, Y. (2011). "PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression." Molecular Cell **43**: 275-284.

Reddy, T., & Rainey, J. K. (2010). "Interpretation of biomolecular NMR spin relaxation parameters." Biochemistry and Cell Biology **88**(2): 131-142.

Ringer, A. L., Senenko, A., and Sherrill, C. D. (2007). "Models of

S/pi interactions in protein structures: comparison of the H2S benzene complex with PDB data." Protein Science **16**(10): 2216-2223.

Rossi M., T. A., Rempe BS., Varma S., (2013). "Role of methyl-induced polarization in ion binding." Proceedings of the National Academy of Sciences of the United States of America. **110**(32): 12978-12983.

Rü diger S, F. S., Veprintsev DB, Fersht AR (2002). "CRINEPT-TROSY NMR reveals p53 core domain bound in an unfolded form to the chaperone Hsp90." Proceedings of the National Academy of Sciences of the United States of America **99**(17): 11085-11090.

Sanchez, R., and Zhou, M. M. (2011). "The PHD finger: a versatile epigenome reader." Trends Biochem Science **36**(7): 364-372.

Schang AL, S.-D. D., Mezger V (2018). "The impact of epigenomic next-generation sequencingapproaches on our understanding of neuropsychiatricdisorders." Clinical Genetics **93**: 467-480.

Schindler, U., Beckmann, H., and Cashmore, A. R (1993). "HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region." Plant Journal **4**: 137-150.

Schuetz A, A.-H. A., Martín F, Loppnau F, Vedadi M, Bochkarev A, Plotnikov AN, Arrowsmith CH, and Min J. (2006). "Structural basis for molecular recognition and presentation of histone H3 By WDR5." The EMBO Journal **25**(18): 4245–4252.

Schulman B. A., K. P. S., Dobson C. M., Redfield C. (1997). "A residue-specific NMR view of the non-cooperative unfolding of a molten globule." Nature Structural Biology **4**: 630-634.

Schwartz M. P., H. S., Matouschek A. (1999). "The Structure of Precursor Proteins during Import into Mitochondria." Journal of Biological Chemistry **274**(18): 12759–12764.

Sharma D, R. K. (2000). "C-13 NMR chemical shifts can predict disulfide bond formation." Journal of Biomolecular NMR **18**: 165–171.

Sherr, C. J. (2004). " Principles of tumor suppression." Cell **116**(2): 235-246.

Shi, Z., Olson, C. A., and Kallenbach, N. R. (2002). "Cation-pi

interaction in model alpha-helical peptides." Journal of the American Chemical Society **124**(13): 3284-3291.

Shoemaker KR, K. P., York EJ, Stewart JM, Baldwin RL. (1987). "Tests of the helix dipole model for stabilization of alpha-helices. ." Nature **326**(6113): 563-567.

Sinnokrot, M. O., and Sherrill, C. D (2006). "High-accuracy quantum mechanical studies of pi-pi interactions in benzene dimers. ." The Journal of Physical Chemistry A **110**(37).

Sippel, K. H., and Quiocho, F. A. (2015). "Ion-dipole interactions and their functions in proteins." Protein Science **24**(7): 1040-1046.

Smith, E., and Shilatifard, A. (2010). "The Chromatin Signaling Pathway: Diverse Mechanisms of Recruitment of Histone-Modifying Enzymes and Varied Biological Outcomes." Molecular Cell **40**: 689-701.

Sosnick TR, B. D. (2011). "The folding of single domain proteins--have we reached a consensus?" Current opinion in structural biology **21**(1): 12-24.

Sowers LC, S. B., Sedwick WD. (1987). "Base stacking and molecular polarizability: Effect of a methyl group in the 5-position of pyrimidines." Biochemical and Biophysical Research Communications. **148**(2): 790-794.

Tallant C, V. E., Fedorov O, Overvoorde L, Ferguson FM, Filippakopoulos P, Svergun DI, Knapp S, Ciulli A. (2015). "Molecular basis of histone tail recognition by human TIP5 PHD finger and bromodomain of the chromatin remodeling complex NoRC." Structure **23**(1): 80-92.

Tencer AH, G. J., Klein BJ, Khan A, Zhang Y, Strahl BD, Wely KHM, and Kutateladze TG (2017). "A unique pH-dependent recognition of methylated histone H3K4 by PPS and DIDO." Structure **25**(10): 1530-1539.

Velazquez-Campoy, A., Leavitt, S. A., & Freire, E (2015). "Characterization of Protein-Protein Interactions by Isothermal Titration Calorimetry." Methods in Molecular Biology **1278**: 183-204.

Venyaminov S, B. I., Shen ZM, Wu CS, Yang JT (1993). "Circular dichroic analysis of denatured proteins: inclusion of denatured proteins in the reference set." Anal Biochemistry **214**(1): 17-24.

Vivian, J. T., and P. R. Callis (2001). " Mechanisms of tryptophan fluorescence shifts in proteins." Biophysical Journal **80**(5): 2093-2109.

Wang S, K. E. (1995). "Origins of the large differences in stability of DNA and RNA helices: C-5 methyl and 2'-hydroxyl effects." Biochemistry **34**(12): 4125-4132.

WC, J. (1999). "Analyzing protein circular dichroism spectra for accurate secondary structures." Proteins **35**(3): 307-312.

Wijesinha-Bettoni R., D. C. M., Redfield C. (2001). "Comparison of the denaturant-induced unfolding of the bovine and human α-lactalbumin molten globules." Journal of Molecular Biology **312**(1): 261-273.

Wishart DS, B. C., Yao J, Abilgaard F, Dyson HJ. (1995). "1H, 13C and 15N chemical shift referencing in biomolecular NMR." Journal of Biomolecular NMR **6**(2): 135-140.

Wishart DS, S. B., Richards FM (1991). "Simple Techniques For the Quantification of Protein Secondary Structure By H-1 NMR Spectroscopy." FEBD Lett **293**: 72-80.

Wolf-Watz M., T. V., Henzler-Wildman K., Hadjipavlou G., Eisenmesser E and Kern D., (2004). "Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. ." Nature Structural & Molecular Biology **11**(10): 945-949.

Wong K. B., F. S. M., Fersht A. R. (1996). "Cold Denaturation of Barstar:1H,15N and13C NMR Assignment and Characterisation of Residual Structure." Journal of Molecular Biology **259**(4): 805-818.

Xiong X, P. T., Yang S, Zhao S, Yan P, Zhang W, Xie W, Li Y, Zhao Y, Allis CD, Li H. (2016). "Selective recognition of histone crotonylation by double PHD fingers of MOZ and DPF2." Nature Chemical Biology **12**(12): 1111-1118.

Yun M, W. J., Workman JL, and Li B (2011). "Readers of histone modifications." Cell Research **21**(4): 564–578.

Zeng L., Z. Q., Li S., Plotnikov A. N., Walsh M. J., and Zhou M. M. (2010). "Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b." Nature **466**(7303): 258-262.

Zhuravleva, A., & Korzhnev, D. M (2017). "Protein folding by NMR." Progress in Nuclear Magnetic Resonance Spectroscopy **100**: 52-77.

Zou Q, H.-R. S., Murphy KP. (1998). "Urea effects on protein stability: hydrogen bonding and the hydrophobic effect." Proteins **31**(2): 107-115.