

Binding and dynamic studies to explore the combinatorial  
recognition of H3R2K9me and 5-methyl cytosine oxidation  
derivatives by the reader domains of UHRF1

M. ANGEL NIVYA

A Dissertation Submitted to  
Indian Institute of Technology Hyderabad  
In Partial Fulfillment of the Requirements for  
The Degree of Doctor of Philosophy



भारतीय प्रौद्योगिकी संस्थान हैदराबाद  
Indian Institute of Technology Hyderabad

Department of Biotechnology

May, 2018

## Declaration

I declare that this written submission represents my ideas in my own words, and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

-----  
M. ANGEL NIVYA  
(Signature)

-----  
M. ANGEL NIVYA  
(Student Name)

-----  
BOI4RESCH11001  
(Roll No)

## Approval Sheet

This thesis entitled "Binding and dynamic studies to explore the combinatorial recognition of H3R2K9me and 5-methyl cytosine oxidation derivatives by the reader domains of UHRF1" by M. Angel Nivya is approved for the degree of Doctor of Philosophy from IIT Hyderabad.



Prof. Paiké Jayadeva Bhat

Dept of Biosciences and Bioengineering

IIT Bombay

Examiner

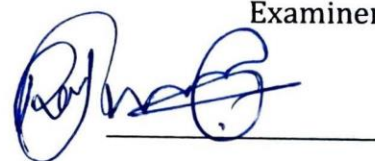


Dr. Basant Kumar Patel

Dept of Biotechnology

IIT Hyderabad

Examiner



Dr. Rajakumara Eerappa

Dept of Biotechnology

IIT Hyderabad

Adviser



Dr. Sunil Kumar Maity

Dept of Chemical Engineering

IIT Hyderabad

Chairman

## Acknowledgements

First and foremost, I would like to express my whole-hearted gratitude to my thesis supervisor, Dr. Rajakumara Eerappa, for his constant guidance, persistence and support in the joyful but challenging journey of my research. Apart from technical guidance, he has helped me to improve my interpersonal skills and make me a better individual. Any discussion with him has always been enriching and rejuvenating. If I should thank anyone the most for making me able to submit my thesis at IITH, it would be my supervisor; any word to express my gratitude towards him would fall short.

I would like to thank my doctoral committee members Dr. Basant Kumar Patel, Dr. Anamika Bhargava and Dr. Jyotsnendu Giri for their critical comments and valuable suggestions.

I would like to thank the former and present Head of Department of Biotechnology, IITH, for their kind cooperation and for providing necessary facilities and infrastructure in the department

I sincerely thank to Dr. Sanjeev Khosla, Centre for DNA Fingerprinting and Diagnosis (CDFD), for allowing me to work in his lab and as well as for his valuable suggestions, discussions and help during my research. I extend my thanks to CDFD director for giving permission and allowing us to access the CDFD facilities.

I am thankful to Dr. Rajan Sankaranarayanan, Centre for Cellular and Molecular Biology (CCMB), for providing us slots for carrying out ITC measurements, and Dr. P. Shobha Krupa Rani, for her kind help during the experiment.

Sincere thanks to all my colleagues and friends Mr. Naveen Kumar, Mr. Satish Mutyala, Mr. Abhishek Suman, Ms. Deeksha Waghela, Mr. Sreeragh Sivadas, Ms. Shivani Dixit and Ms. Vani for their co-operation, providing me a friendly

atmosphere during my work in lab and critical comments on my thesis. I would like to thank Mr. Abhishek Suman once again who helped me in carrying out molecular dynamics simulation studies.

I also thank CFD friends Dr. Prakruti, Dr. Imitiyaz, Ms. Archana, Ms. Thushara, Mr. Viplove, Mr. Rajeev, Mr. Ambay, Mrs. Lalitha, Mr. Hassan and Ms. Reka for their kind help during my research.

I thank all my IITH friends and batchmates, Ms. Nisha James, Ms. Shuba, Ms. V. Swarnalatha, Ms. Induvahi, Ms. M. Swarnalatha, Ms. Keerti, Dr. Manasa, Ms. Swathi, Ms. Karuna, Ms. Shiraj, Ms. Pragathi, Ms. Nazil, Ms. Shuba Jain, Ms. Sharanya, Ms. Pollomey, Ms. Sanjana, Ms. Keerthana, Ms. Vidhya, Ms. Aarthi, Ms. Shivangi, Ms. Monisha, Dr. Anu, Mr. Aravind, Dr. Archana, Dr. Vishwanath, Dr. Naveena, Dr. Guru, Dr. Narendar, Mr. Yogesh, Mr. Venkat, Mr. Manohar, Ms. Richa, Ms. Deepa, Ms. Neema, Ms. Srinithi, Ms. Narasimha, Mr. Himanshu, Mrs. Ashwini, Ms. Jayalakshmi, Ms. Dhanalakshmi, Ms. Shesha, Mrs. Lalitha, Dr. Rakesh, Mr. Suresh, Mr. Thomas, Mr. Pradeep, Mr. Joel, Mr. Koti, Mr. Yanthan, Mr. Sreenivas who made my life in IIT memorable and rejuvenating.

I am grateful to Indian Institute of Technology Hyderabad (IITH) for providing facilities and MHRD for financial support.

I am at a loss of words to express gratitude to my parents and my relatives. The way I grew up, the values I imbibed, the education I received and the person I am now is all due to them. I am grateful to my caring and loving sister who has always been an ardent listener and a strong support to me.

Above all, I thank to almighty god for His favor, love and blessings.

M. Angel Nivya

Dedicated to My Family and Friends

## Abstract

UHRF1 is a multi-domain protein comprising of a tandem tudor domain (UHRF1 TTD), a PHD finger, and a SET and RING-associated (UHRF1 SRA) domain. It is involved in the maintenance of CG methylation, heterochromatin formation and DNA repair processes. Isothermal titration calorimetry binding studies of UHRF1 TTD with unmodified and methylated lysine histone peptides establishes that the UHRF1 TTD binds dimethylated lysine 9 on histone H3 (H3K9me<sub>2</sub>). Further, MD simulation and binding studies reveal that, together, UHRF1's TTD and PHD (UHRF1 TTD-PHD) preferentially recognizes dimethyllysine status. Importantly, it was found that in the binding pocket of UHRF1, the Asp145 which determines the preferential recognition of the dimethyl-ammonium group of H3K9me<sub>2</sub>. In contrast, PHD finger of the UHRF1 TTD-PHD has a negligible contribution to the binding affinity for recognition of H3K9me<sub>2</sub> by the UHRF1 TTD-PHD. Surprisingly, Lys4 methylation on H3 peptide has an insignificant effect on combinatorial recognition of R2 and K9me<sub>2</sub> on H3 (H3R2K9me<sub>2</sub>) by the UHRF1 TTD-PHD. Electrophoretic mobility shift assay and fluorescence polarization binding studies indicate that the UHRF1 SRA binds to all the oxidation products of 5-methyl cytosine (5mC) but exhibits lower affinity towards 5-carboxyl cytosine (5caC) and 5-formyl cytosine (5fC). Subtle variations of key residues at the binding pocket could determine status specific recognition of histone methyllysine by the reader domains. Thus, further studies are required to unravel the possibility of combinatorial recognition of H3R2K9me<sub>2</sub>, and 5hmCG DNA by the UHRF1 and role of combinatorial recognition on UHRF1 functions especially for CG methylation maintenance and heterochromatin formation.

## Nomenclature

5caC	-	5-carboxyl Cytosine
5fC	-	5-formyl Cytosine
5hmC	-	5-hydroxymethyl Cytosine
5mC	-	5-methyl Cytosine
6-FAM	-	6-carboxyfluorescein
Å	-	Angstroms
K <sub>A</sub>	-	Association constant
BER	-	Base Excision Repair
bp	-	Base pair
cDNA	-	Complementary DNA
CpG	-	Cytosine-Guanine
°C	-	Degree Celsius
DNA	-	Deoxyribonucleic acid
K <sub>D</sub>	-	Dissociation constant
DTT	-	Dithiothreitol
EMSA	-	Electrophoretic Mobility Shift Assay
ΔH	-	Enthalpy Change
ECRM	-	Epigenetic Code Replication Machinery
<i>E. coli</i>	-	<i>Escherichia coli</i>
EDTA	-	Ethylenediaminetetraacetic acid
fs	-	Femtoseconds
FP	-	Fluorescence Polarization
ΔG	-	Free energy of binding
H-bond	-	Hydrogen bond
IPTG	-	Isopropyl-1-thio-D-galactopyranoside
ITC	-	Isothermal Titration Calorimetry
KMT	-	Lysine methyl transferase



KDM	-	Lysine demethylase
kcal	-	Kilocalorie
kDa	-	Kilodalton
MgCl <sub>2</sub>	-	Magnesium chloride
µcal	-	Microcalorie
µg	-	Microgram
µl	-	Microliter
ml	-	Milliliter
mM	-	Millimolar
mP	-	Millipolarization
mg	-	Mlligram
N	-	Molar ratio
mol	-	Mole
MD	-	Molecular Dynamics
MM-GBSA	-	Molecular Mechanics-Generalized Born Surface Area
nm	-	Nanometer
nM	-	Nanomolar
nmol	-	Nanomole
ns	-	Nanosecond
PCH	-	Pericentric Heterochromatin
pM	-	Pico mole
ps	-	Pico second
PHD	-	Plant Homeo Domain
PTM	-	Post translational modification
RNA	-	Ribonucleic acid
RMSD	-	Root Mean Square Deviation
RMSF	-	Root Mean Square Fluctuation
sec	-	Second
SRA	-	Set and Ring Associated

SEC	-	Size Exclusion Chromatography
NaCl	-	Sodium chloride
TTD	-	Tandem Tudor Domain
t	-	Time
Tris-HCl	-	Tris Hydrochloride
UHRF1	-	Ubiquitin-like with PHD and RING Finger 1
U	-	Unit
vs	-	Versus
V	-	Volts
ZnCl <sub>2</sub>	-	Zinc Chloride

# Contents

Declaration.....	Error! Bookmark not defined.
Approval Sheet.....	ii
Acknowledgements.....	iii
Abstract.....	vii
Chapter 1.....	1
Review of Literature .....	1
1.1 Introduction.....	2
1.1.1 Nucleosomes and chromatin organization .....	3
1.1.2 Epigenetic modifications.....	5
1.1.3 Writers, eraser and readers.....	6
1.1.4 Classification of epigenetic reader .....	8
1.1.4.1 Chromatin architectural proteins.....	9
1.1.4.2 Chromatin remodelers.....	9
1.1.4.3 Chromatin modifiers.....	10
1.1.4.4 Adaptors.....	11
1.1.5 Lysine methylation .....	11
1.1.6 DNA methylation.....	14
1.1.7 Linking DNA methylation with histone mark.....	16
1.1.8 UHRF1.....	17
1.1.9 UHRF1 domain architecture.....	18
1.1.9.1 Ubiquitin like domain (UBQ) .....	18
1.1.9.2 Tandem Tudor Domain (TTD).....	19
1.1.9.3 Plant Homeodomain (PHD).....	19
1.1.9.4 SRA domain .....	19
1.1.9.5 RING domain .....	20

1.2	Aim of the thesis .....	21
Chapter 2 .....		22
Characterization of histone lysine methyl mark binding by TTD domain of UHRF122		
2.1	Introduction .....	23
2.1.1	Tudor domain .....	23
2.1.2	Tudor Domain of 53BP1 .....	24
2.1.3	Sequence comparison of UHRF1 TTD and 53BP1 TTD binding pocket ..	25
2.1.5	Role of flanking sequence on methyllysine recognition .....	27
2.1.6	Methylation status specific readout .....	28
2.1.7	Isothermal titration calorimetry .....	31
2.2	Materials and Methods .....	32
2.2.1	Sub-Cloning of UHRF1 TTD to generate hexahistidine-SUMO tagged construct .....	32
2.2.2	Expression and Purification of UHRF1 TTD .....	33
2.2.3	ITC measurements .....	34
2.3	Results .....	35
2.3.1	UHRF1 TTD selectively recognizes K9 methylation mark on histone H3	35
2.3.2	Recognition of H4K20me mark by the UHRF1 TTD .....	36
2.3.3	Methylation status specific readout of H3K9me by UHRF1 TTD .....	37
2.4	Discussion .....	38
Chapter 3 .....		40
Molecular Dynamic simulation and binding studies to explore the H3R2K9 methylation statuses recognition by the TTD-PHD of UHRF1 .....		40
3.1	Introduction .....	41
3.1.1	Combinatorial recognition of histone marks in single histone tail ( <i>cis</i> ) ....	42
3.1.2	Combinatorial recognition of PTMs in different histone tails ( <i>trans</i> ) .....	43

3.1.3	Multivalent histone engagement of UHRF1 .....	44
3.1.4	Fluorescence polarization.....	46
3.1.5	Molecular Dynamic Simulation.....	47
3.2	Material and Methods.....	49
3.2.1	Preparation of UHRF1 TTD-PHD-H3K9me2 complex for simulation.....	49
3.2.2	Preparation of UHRF1 TTD-PHD Asp145Glu and Asp145Ala mutants for simulations .....	49
3.2.3	Molecular dynamics (MD) simulations.....	49
3.2.4	Protein-peptide interaction analysis.....	50
3.2.5	Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) calculations.....	50
3.2.6	Sub-Cloning of UHRF1 TTD-PHD to generate hexahistidine-SUMO tagged construct.....	51
3.2.7	Expression and Purification of UHRF1 TTD-PHD.....	51
3.2.8	Generation of UHRF1 TTD-PHD Asp145Glu mutant.....	52
3.2.9	ITC measurements.....	52
3.2.10	Fluorescence Polarization (FP) measurements .....	53
3.3	Results .....	53
3.3.1	MD simulation studies on binding of H3R2K9me2 and H3R2K9me3 peptides to TTD-PHD UHRF1 cassette .....	53
3.3.2	UHRF1 TTD-PHD cassette preferentially binds H3R2K9me2 .....	57
3.3.3	UHRF1 TTD-PHD Asp145Glu mutant preferentially binds di-methyllysine 9 over tri-methyllysine 9.....	58
3.3.4	Effect of UHRF1 PHD domain on histone H3 lysine methyl mark recognition by the UHRF1 TTD .....	59
3.4	Discussion.....	60
3.4.1	Effect of negatively charged residue in the binding pocket on lower methyllysine status recognition.....	60

3.4.2	Comparison of dynamics of the dimethyllysine 9 and the trimethyllysine 9 recognition by the UHRF1 TTD .....	62
3.4.3	UHRF1 PHD doesn't contribute in recognition of H3K9me2 by UHRF1 TTD domain .....	63
Chapter 4	.....	65
	Molecular Dynamics (MD) simulation and binding studies to explore the effect of H3K4 methylation on H3R2K9me recognition by the TTD-PHD of UHRF1 .....	65
4.1	Introduction.....	66
4.2	Histone modifications crosstalk.....	67
4.3	Reader domain mediated histone crosstalk .....	67
4.4	Phospho/methyl switch .....	68
4.5	Roles of histone modification crosstalk.....	69
4.6	Conformation of N-terminal H3 peptide in the TTD-PHD of UHRF1 .....	69
4.7	Material and Methods.....	71
4.7.1	MD Simulation .....	71
4.7.1.1	Preparation of UHRF1 TTD-PHD for simulation .....	71
4.7.1.2	Molecular dynamics (MD) simulations.....	72
4.7.1.3	Protein-ligand interaction analysis .....	72
4.7.1.4	Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) .....	73
	Calculations.....	73
4.7.2	Sub-Cloning of UHRF1 TTD-PHD to generate hexahistidine SUMO tagged construct.....	73
4.7.3	Expression and Purification of UHRF1 TTD-PHD .....	74
4.7.4	ITC measurements.....	75
4.8	Results .....	76

4.8.1	MD simulation studies on binding of H3R2K4me2K9me2 and H3R2K4me2K9me3 peptides to UHRF1 TTD-PHD cassette .....	76
4.8.2	Dimethyllysine 4 on H3 has insignificant effect on H3R2K9me2 recognition by the UHRF1 TTD-PHD Cassette.....	79
4.9	Discussion.....	80
Chapter 5.....		81
Characterization of 5mC oxidation derivative recognition by SRA domain of UHRF1. ....		81
5.1.1	Oxidation of 5mC .....	82
5.1.2	Recognition of 5mC oxidation derivatives by the SRA domain.....	83
5.2	Material and methods .....	84
5.2.1	Sub-Cloning of UHRF1 SRA to generate hexahistidine-SUMO tagged construct .....	84
5.2.2	Expression and Purification of UHRF1 SRA .....	85
5.2.3	Qualitative analysis of SRA-DNA complex by SEC.....	86
5.2.4	Electrophoretic Mobility Shift Assay (EMSA) .....	86
5.2.5	Fluorescence Polarization (FP) measurements.....	87
5.3	Results .....	87
5.4	Discussion.....	90
6.1	Conclusion and future perspectives .....	91
References .....		92

## Chapter 1

### Review of Literature



## 1.1 Introduction

Though there has been a significant advancement in genomics, still, epigenetic regulation of genetic mechanisms remains one of the major challenges. In 1942, the term “epigenetics” was first coined by Conrad Waddington, defining it as “the branch of biology which studies the casual interactions between genes and their products which bring the phenotype into being” [1,2]. Today, more stringently, epigenetics is defined as the heritable chemical modifications, which, in turn can cause structural modifications of biomolecules including DNA, RNA and proteins, without any alteration in their primary sequence [3].

Structural modification of chromatin leads to epigenetic changes in the gene expression that stem from altered accessibility of the transcription machinery to particular genes. The chromatin structural modulations signify the means of controlling access to genetic information. The packaging of nuclear DNA is one of the strategies for regulating the genetic information accessibility. In this approach, linear DNA is wrapped around histone particles to form the nucleosome, considered as a basic unit of chromatin. The strength of the interaction between DNA and histone proteins play a crucial role for DNA packaging, which dictates chromatin to be either in a condensed (silenced) or relaxed (active) state. The structural reorganization in the chromatin regulates gene activity and specific cellular phenotype [4,5].

Chromatin can be classified into euchromatin and heterochromatin, based on their functional perspectives. Typically, euchromatin represents actively transcribed genes in the genome. These euchromatin regions, usually, contain unmethylated CpG and hyper acetylated N-terminal lysine residues on H3 and H4 core histones; and are accessible to nucleases. On the other hand, heterochromatin represents transcriptionally inactive and highly condensed regions of the genome. Dinucleotide CpG methylation and significant hypoacetylation on histone usually occur in

heterochromatin and DNA is nearly inaccessible to the nucleases. Furthermore, heterochromatin is sub-classified into constitutive and facultative heterochromatin based on whether heterochromatin is formed in every cell type or particular cell, respectively [6]. The governing factors that affect the transition between euchromatin and heterochromatin are DNA methylation, RNA interference (RNAi), non-coding RNAs, histone post-translational modifications and DNA replication-independent incorporation of histone variants [7].

### 1.1.1 Nucleosomes and chromatin organization

The genomic DNA in eukaryotes is primarily packed into nucleosome which constitutes ‘core particle’, ‘linker DNA’ a 10-90 bp long nucleotide sequence which connects the two nucleosomes and ‘linker histone H1’. The core particle of nucleosome comprises 146 bp DNA which wraps around histone octamer, consisting of two copies of each of the four histones, H2A, H2B, H3, and H4 (Fig. 1.1A) [8,9]. The core histones share a structurally conserved motif, the histone fold. It comprises of three helices, two short helices ( $\alpha1$  and  $\alpha3$ ; 10-14 amino acids) and one longer helix ( $\alpha2$ ; 28 amino acids). These helices are connected by two short loops (L1 and L2). In solution, homo or heterodimer formation of histone is strongly influenced by the histone fold protein interaction (Fig. 1.1B and C) [10].

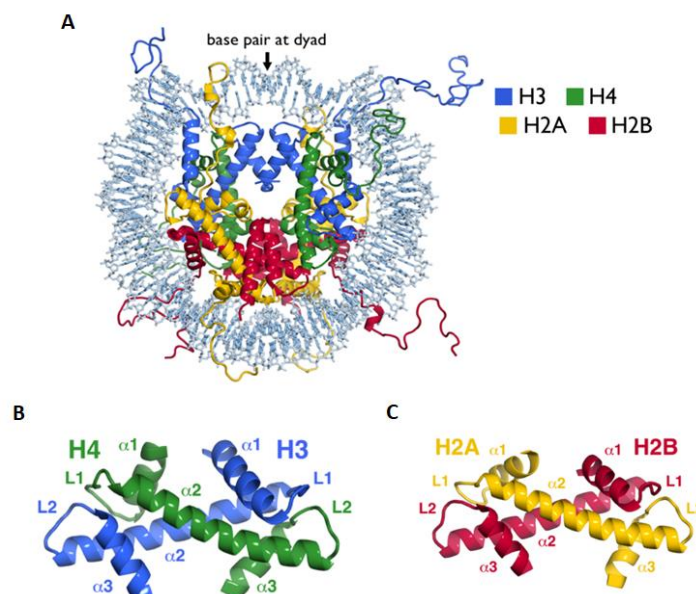


Figure 1.1: Structure of nucleosome core particle and histone fold heterodimers. (A) Nucleosome core particle structure (PDB ID: 1KX5). Histones and DNA are depicted in cartoon and sticks representations, respectively, and colored as indicated. (B) H3/H4 histone-fold heterodimer. (C) H2A/ H2B histone-fold heterodimer. [Adapted from McGinty and Tan, 2014]

The main functions of the nucleosome are: (i) to bring genome compaction through organization of the genomic DNA; (ii) provide scaffold for the binding of the chromatin associated proteins and displaying combinatorial array of post-translational modifications (PTMs) to regulate the stability of the nucleosome and compaction of chromatin; (iii) form high-order chromatin structures through self-assembly, leading to further compaction of the genome [11,12].

The primary chromatin structure, where DNA is wrapped around nucleosome, 11 nm thick, is called bead-on-a string form. The individual nucleosome interactions is a driving force for the folding of a nucleosomal array (the primary structure of chromatin) into the 30 nm fiber, solenoid (a secondary structure) and into large-scale configurations, chromatin (tertiary structures) that build an entire chromosome [13] (Fig. 1.2) The chromatin fiber within the nucleus is in dynamic state and also flexible over longer lengths. It usually undergoes several sorts of remodeling activities such as histone modifications, sliding and reduction of nucleosomes, and insertion of histone variants [14].

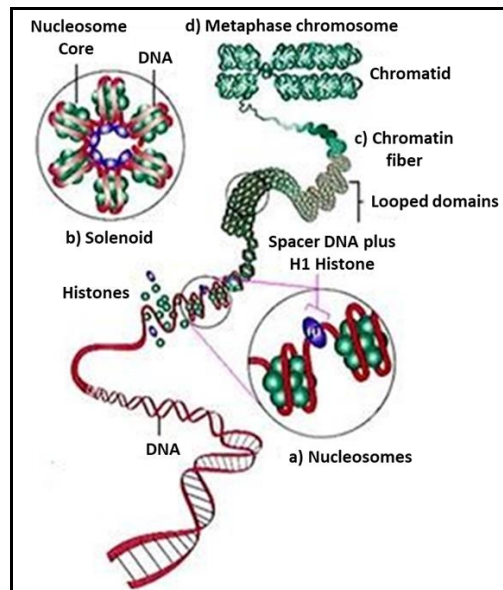


Figure 1.2: Hierarchical organization of chromatin.

[Adapted from Klug and Cummings, 1997]

### 1.1.2 Epigenetic modifications

The function of chromatin is largely regulated by the post-translational modifications (PTMs) of histones [7,15]. N-terminal tails of histone proteins usually undergo various covalent modifications including methylations, phosphorylation and acetylation (Fig. 1.3) [16]. Such covalent modifications directly regulate the structure of chromatin and often serve as binding sites for the non-histone proteins recruited to the chromatin [17].

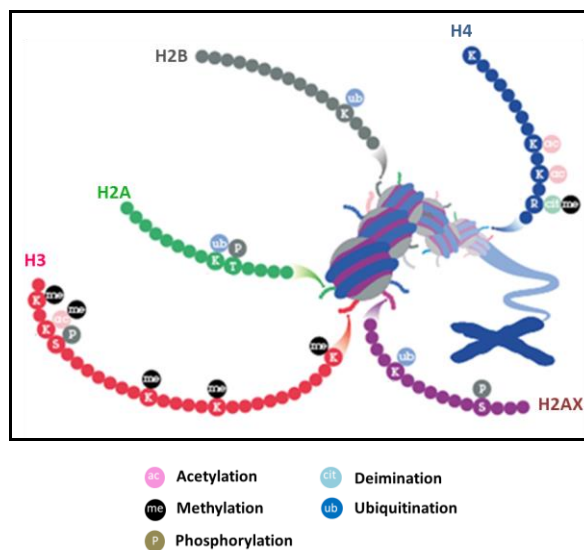


Figure 1.3: Post-translational modifications of N-terminal histone tails.

### 1.1.3 Writers, eraser and readers

PTMs of histones offer a controlled mechanism for regulating chromatin structure and dynamics. Thus, the direct interactions between histones and DNA are strongly influenced by the PTMs [18,19], hence affects the DNA-associated processes. Proteins involved in epigenetic modifications are classified as writers, readers and erasers (Fig. 1.4). The incorporation of specific PTM is performed by “writer” proteins. For example, acetyltransferases, methyltransferases and kinases add the chemical groups acetyl, methyl and phosphoryl, respectively, onto histone tails. Such covalent modifications act as docking sites for “reader” domains such as bromodomain, tudor domain, PHD finger and chromodomain. Whereas, “eraser” proteins remove the chemical groups added by the writers, for example, deacetylase, demethylase, and phosphatase [20,21].

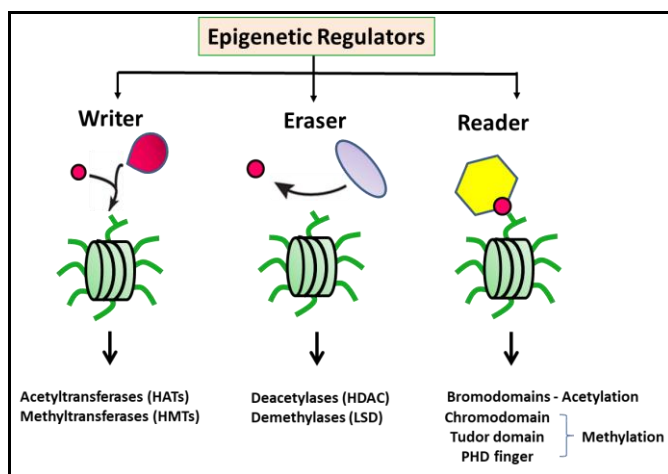


Figure 1.4: Epigenetic regulator proteins of histone marks.

Gene expression is influenced by the accessibility of DNA rendered by nucleosome structure and chromatin conformation, histone PTMs, along with linker histones, histone chaperons and ATP-driven nucleosome remodelers. In another mechanism, effector proteins specifically bind to PTMs on histone through their reader domain then regulate gene expression. Chromatin associated reader domains are often

present in multi-protein complexes. These domains bind to the epigenetic marks and affects chromatin association (Fig. 1.5), which eventually led to the specific biological outcomes. Therefore, histone modifications regulate the expression of DNA encoded genetic information [22].

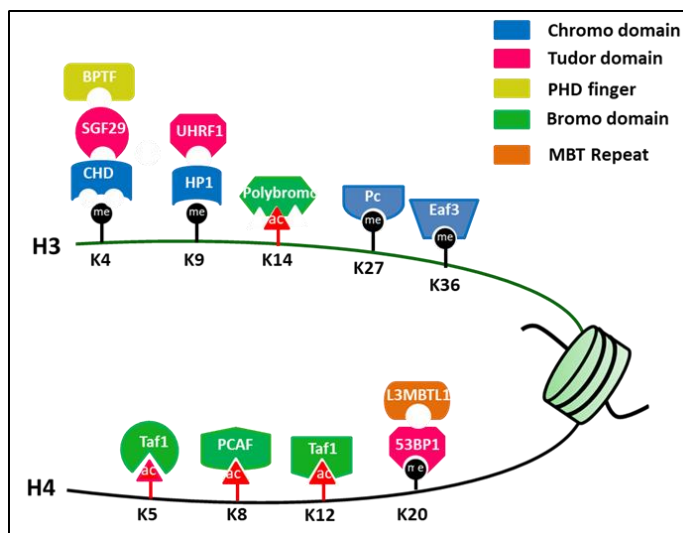


Figure 1.5: Recognition of histone modifications by reader domains of respective proteins.

It has been well-established that the recognition of specific histone PTMs is facilitated by conserved binding modules through a variety of intermolecular interactions. For instance, two most commonly recognized PTMs are methylation and acetylation on lysine (Table 1.1). Bromodomains bind to acetylated lysine (Kac), through a deep hydrophobic cavity formed by four conserved helical bundles, and exhibit affinity for single or multiple Kac. PHD fingers and royal superfamily domains (chromo, tudor, plant Agenet, MBT and PWWP), on the other hand, form an aromatic cage which can accommodate the methylated lysine through cation- $\pi$  or CH- $\pi$  interactions. Size and other binding pocket residues of the aromatic cage determines the binding specificity for different methylation status of lysine (mono-, di- or tri-methylation) [23]. Readers also differentiate same type of modification on a residue at different position on histones, through sequence specific interactions, by interacting with the flanking sequence of the altered amino acid [24].

Table 1.1: Readers of histone PTM.

Recognition of	Reader domain	Histone PTM
Methyllysine	ADD	H3K9me3
	Ankyrin	H3K9me2, H3K9me1
	BAH	H4K20me2
	Chromo-barrel	H3K36me3, H3K36me2, H3K36me1, H3K4me1
	Chromodomain	H3K9me3, H3K9me2, H3K27me3, H3K27me2
	DCD	H3K4me3, H3K4me2, H3K4me1
	MBT	H3K4me1, H3K4me2, H4K20me1, H4K20me2
	PHD	H3K4me3, H3K4me2, H4K9me3
	PWWP	H3K36me3, H4K20me1, H4K20me3, H4K79me3
	TTD	H3K4me3, H3K9me3, H4K20me2
	Tudor	H3K36me3
	WD40	H3K27me3, H3K9me3
	Zf-CW	H3K4me3
Methylarginine	ADD	H4R3me2s
	Tudor	H3Rme2, H4Rme2
	WD40	H3R2me2
Acetyllysine	Bromodomain	H3K36ac, H4Kac, H2AKac, H2BKac
	DBD	H3K18acK23ac, H4K5acK8ac
Phosphoserine (or)	14-3-3	H3S10ph, H3S28ph
	BIR	H3T3ph
Phosphothreonine	Tandem BRCT	H2AXS139ph
Unmodified histone	ADD	H3un
	PHD	H3un

#### 1.1.4 Classification of epigenetic reader

Reader domain containing proteins are broadly categorized into four groups, (Fig. 1.6) namely

1. Chromatin architectural proteins
2. Chromatin remodelers
3. Chromatin modifiers
4. Adaptors

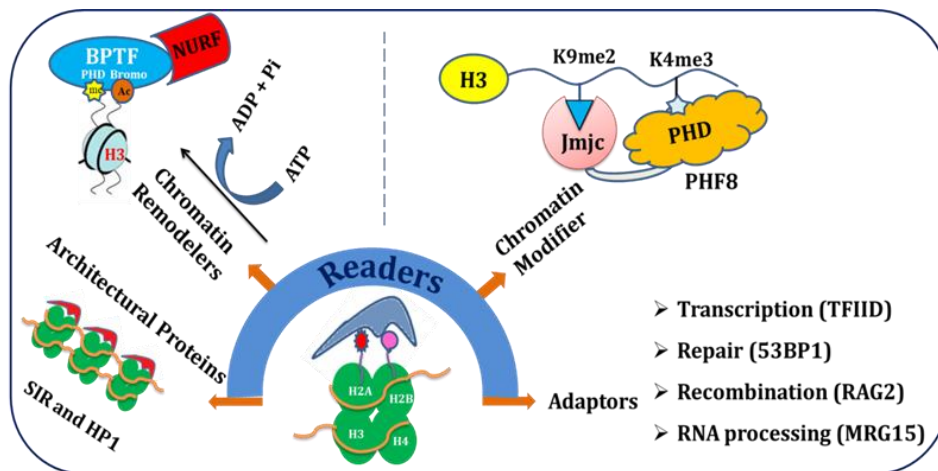


Figure 1.6: Functional outcomes of reader domain on histone modification readout.

#### 1.1.4.1 Chromatin architectural proteins

Chromatin architectural protein complexes recognize multiple nucleosomes concurrently and have the capacity to compact chromatin or to make underlying DNA inaccessible by providing physical shields. Such proteins frequently undergo self-propagation and oligomerization resulting in the spreading of protein over a large region of a chromatin [24]. Reader domains present on chromatin architectural proteins are required for modification specific binding and compaction of chromatin. For example, HP1 proteins which are involved in the establishment and maintenance of chromatin higher-order structures [25], contains domains, such as chromodomain (CD) and chromo shadow domain (CSD). CD binds to H3K9me2/3 mark, while CSD is required for creating a dimeric interface for recruiting specific ligands to connect the DNA-interacting hinge region [26].

#### 1.1.4.2 Chromatin remodelers

Chromatin remodelers are multi-protein complexes which have ATPase subunit of the Snf2 subfamily, that utilize the energy generated during ATP hydrolysis for mobilization/sliding of nucleosomes, that in turn modulate the chromatin structure. Remodelers include proteins that ensure the proper density and spacing of nucleosomes and also contribute to gene repression. Some remodelers facilitate the



binding of site-specific transcription factors to DNA by cooperating with transcription factors and enzymes that modify histones [27].

Remodeling complexes possess reader domain to identify chromatin signals. Reader domains in combination with other domains are indispensable for the specific targeting and regulating the enzyme activity [28]. For example, NoRC (Nucleolar Remodeling Complex) becomes functional when the bromodomain of Tip5 (a subunit of NoRC) binds to the acetyllysine 16 of histone 4 (H4K16ac). Even a point mutation in bromodomain, impedes the association of NoRC with chromatin, precludes heterochromatin formation, and hinders the transcriptional repression [28]. Similarly, the recognition of H3K4me3 by PHD and Kac by bromodomain from BPTF (sub-unit of NURF), respectively, plays an important role in localization of NURF (NUcleosome Remodeling Factor) to chromatin [24].

#### 1.1.4.3 Chromatin modifiers

There are some reader domains containing proteins that do not modulate chromatin architecture directly but recruit secondary chromatin modifiers or revert the existing modification.

Several proteins are which involved in either addition or removal of the PTMs, like histone lysine methyltransferases (HKMTs), histone deacetylases (HDACs), histone acetyltransferases (HATs), lysine demethylases (KDMs), E3 ubiquitin ligase and deubiquitylases (DUBs), are multi-subunit complexes. They target the globular domains or N-terminal tails of core histones, particularly specific residues, for modifications. For instance, Sin3S complex is recruited to the methyllysine of histones via Sin3 tandem bromodomain. Co-repressor of Sin3S complex contains histone deacetylases such as HDAC1, HDAC2 and HDAC3, and thus the recruitment of this complex lead to the histone deacetylation, i.e., secondary histone modification [29].

#### 1.1.4.4 Adaptors

The primary function of these reader domain-containing adaptor proteins is to recruit factors that are associated with DNA metabolic processes such as transcription, DNA recombination, DNA replication, DNA damage repair, and RNA processing. For example, BRCT (BRCA1 N-terminus) domain of MDC1, an important mediator of the DNA damage response, interacts with the phosphorylated serine residue on histone H2AX, and act as an adaptor to recruit the histone ubiquitin ligase, RNF8, to the site of double-strand break-flanking chromatin.

#### 1.1.5 Lysine methylation

Histone lysine methylation plays a vital role in the regulation of cell cycle, nuclear architecture, genome stability and gene expression [30]. The most interesting feature in histone lysine methylation is the effect of methylation at two different close positions that displays completely distinct functions. For instance, methylation on H3K4 represents transcriptionally active genes, while H3K9 methylation is associated with gene silencing [31]. Genomic stability and maintenance of the cell are strongly by lysine methyltransferases (KMTs) and lysine demethylases (KDMs). H3K9 methylation is regulated by SUV39H1 while demethylation is carried out by JHDM2A. Until now, it was found that, KMTs methylate five lysine residues, K4, K9, K27, K36 and K79 on H3 and K20 residue on the H4 tail (Table 1.2) [32].

Table 1.2: Canonical sites of lysine methylation.

PTM	Position		Reader domain	Protein	Functions
Lysine methylation	H3	K4me0	PHD	WD40	Autoimmune regulator
		K4me	Chromo	CHD	ATPase
			PHD	RAG2	Recombination
		K9	Chromo	HP1	Heterochromatin
		K27	WD40	EED	PRC mediated

					repression
		K36	Chromo	EAF3	Histone deacetylation
		K79	Tudor	53BP1	DSB response
	H4	K20	Tudor	53BP1	DNA damage repair

Lysine is methylated at its  $\epsilon$ -amino group and it can undergo mono, di and tri methylation (Fig. 1.7) known as methylation status or degree of methylation. The position and degree of methylation endows methylated histones to confer active or repressive transcription, however, other histone modification just specify the state of the active or repressed chromatin. Usually, active transcriptions are marked by methylations of K4, K36 and K79 on H3, whereas silenced chromatin states are represented by methylations of K9 and K27 on H3, and K20 on H4.

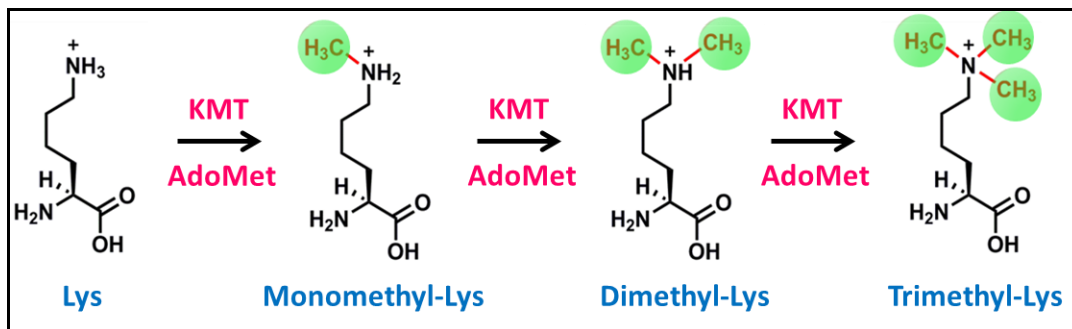


Figure 1.7: Modification status of histone lysine residues. (AdoMet stands for S-adenosylmethionine)

In order to regulate the specific gene expression, HKMTs further communicate/interact with methylated DNA as well as other modified histones. For instance, ubiquitination of H2B in yeast is a prerequisite for stimulation of methylation of H3K4 and H3K79. Besides, in embryonic stem cells (ESCs), bivalent modifications of histones, H3K4 and H3K27 methylations, play a critical role in transforming poised state of gene expression to active or inactive states [33]. Mostly, histone lysine methylations are believed to cause regulatory functions when they are specifically recognized by the effector proteins. These ‘reader’ proteins contain methyllysine-binding motifs, including chromo, PHD, tudor, BAH, MBT, ADD, PWWP, Ankyrin repeat, zn-CW domains and WD40, and also have the ability to

distinguish surrounding amino-acid residues and methyllysine statuses. (Table 1.2) [33].

Mono and dimethyllysine binders have a, relatively, small hole-type cavity that can only accommodate mono or dimethyllysine, but not the bulky trimethyllysine. On the other hand, di and trimethylated lysine binders use wider surface groove as the binding pocket for accommodating the di- or trimethylation moiety. The binding cavity for methyllysine recognition is formed by either only aromatic residues (maximum 4) or aromatic residues (maximum 2) in combination with other residues for methyllysine status specific recognition. (Fig. 1.8) [24].

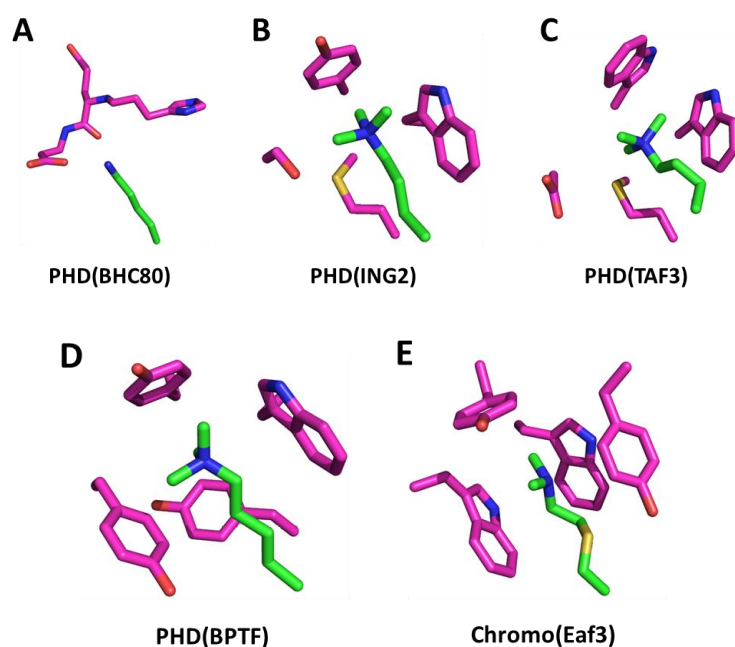


Figure 1.8: Comparison of binding surfaces from different methyllysine readers. Lysines are labeled in green and pocket-forming residues are labeled in pink. (A) No Aromatic cage; (B, C) Half aromatic and (D, E) Fully Aromatic cage. PDB IDs: (A) 2PUY, (B) 2G6Q, (C) 2K17, (D) 2F6J and (E) 2K3Y

Di- and trimethylated ammonium groups are held in the binding pocket composed of aromatic residues by cation- $\pi$ , CH- $\pi$  and van der Waals interactions. Among aromatic residues, Tryptophan can form stronger cation- $\pi$  interaction compared to

tyrosine and phenylalanine. Some readers can recognize the unmethylated lysine (Kme0) through intermolecular hydrogen bonds [24].

### 1.1.6 DNA methylation

DNA methylation is a stable heritable epigenetic mark. DNA methyltransferases (DNMTs) add the methyl group (CH<sub>3</sub>) onto the fifth carbon of cytosine ring of DNA, (Fig. 1.9) [34]. De novo methyltransferases DNMT3a, DNMT3b and their regulatory unit DNMT3l establish cytosine methylation during early embryonic and germ cell development. DNMT1 maintains the methylation patterns during cell generations [35]. Together, they ensure the formation of proper epigenetic landscape in an early development process, and DNMT1 is responsible for transmission of cytosine methylation through cell division.

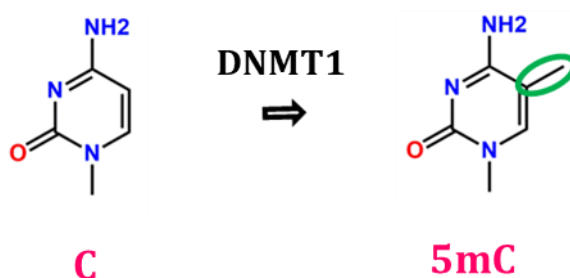


Figure 1.9: Methylation of Cytosine at C5.

Cytosine methylation (5mC) found in either symmetrical (CG or CHG) or asymmetrical (CHH, where H is A, T, or C) sequence context in plants. In mammals, it occurs mostly CG sequence context. Majority of DNA methylation (> 98%) occurs in CpG dinucleotide in somatic cells. However, minor non-CpG methylation in embryonic stem cells (ESCs), the removal and re-establishment of DNA methylation during zygote formation and in the embryo respectively at the time of implantation.

DNA methylation plays a critical role in several important processes including gene expression, X-chromosome inactivation, cellular differentiation, genomic imprinting

and, transposon silencing [34,36]. 5mC can undergo sequential oxidation to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). These oxidation reactions are catalyzed by Ten-eleven translocation (Tet) family dioxygenases. Sequential oxidation of 5mC is required for base excision repair (BER) mediated 5mC demethylation [37]. 5hmC is mostly found in the vicinity of binding sites of transcription factor, and several reports have implicated the association of 5hmC in transcription regulation, epigenetic reprogramming, and pluripotency [38].

Methyl-CpG binding proteins (MBPs) specifically read the 5mC mark on DNA. MBPs are divided into three major families on based on their structural features:

1. Methyl-CpG binding Domain (MBD) protein family,
2. Kaiso protein family
3. SET (Su(var)3-9, enhancer of Zeste and Trithorax) and RING (Really Interesting New Gene) finger associated (SRA) domain protein family.

The MBD proteins serve as a very important component of epigenetic regulation because they are involved in various critical mechanisms that influence the regulation of transcription. These proteins play a main role in silencing transcription by blocking transcription or by affecting other proteins to bind DNA or by inducing chromatin remodeling through their binding partners [39]. MBD family members, such as methyl-CpG binding protein 2 (MeCP2) and methyl-CpG binding domain proteins 1–4 (Mbd1–4) are also capable of binding to (hydroxy) methylated CpG dinucleotides [35,40]. Unlike MBD proteins, Kaiso-like family protein family, including, Kaiso, ZBTB4(Zinc finger and BTB domain containing protein 4), and ZBTB38 consist of a conserved BTB/POZ (Bric-a-brac, tramtrack, broad complex/poxvirus and zinc finger) domain, associated with protein– protein interactions and three Kruppel-like C<sub>2</sub>H<sub>2</sub> zinc finger motifs. Kaiso preferentially binds to two consecutively methylated CpG sites. Like MBD proteins, zinc-finger

domain containing proteins involved in transcriptional repression in a DNA methylation-dependent manner [39,41].

In mammals, SRA domain present in RING domain associated proteins like Np95 (mouse homolog of human ICBP90), also known as UHRF1 (Ubiquitin-like with PHD and Ring Finger Domains 1), recognizes methylated DNA. While in plants, SRA domain present on SET domain associated SUVH (Su(Var 3-9) Homolog) family HKMTs, like SUVH4 and SUVH5 proteins [35]. In contrast to many methyl binding proteins, the main function of UHRF1 is not only to bind to DNA for transcriptional repression, but also to bind DNMT1 and direct it to the hemimethylated DNA to maintain DNA methylation in the genome, particularly during DNA replication process [41].

#### 1.1.7 Linking DNA methylation with histone mark

There is a interdependent relationship between 5mC and H3K9me, i.e., dependent on one another for their establishment and maintenance in the genome [42]. It has been demonstrated that DNA binding proteins such as MeCP2, MBD1, and Kaiso are associated with histone modifications. These proteins bind to the promoter region by recognizing the methylation mark on DNA, and recruit protein complexes such as HDACs and HKMTs for histone deacetylation and methylation. Thus, DNA methylation induces chromatin structural changes through histone modifications [43]. MeCP2 associated with SUV39H1/2, a H3K9 methyltransferase, plays a role in gene repression and heterochromatin formation [44]. Both DNMT1 and DNMT3a assist H3K9 methylation by recruiting SUV39H1 H3K9 methyltransferase to chromatin. Also, DNMT1 and DNMT3b can assist the removal of acetylation mark from histones upon binding to HDACs, which leads to tight packing of DNA and shields it from transcription machinery [41]. G9a, a H3K9 methyltransferase from mammals, forms a complex with HDAC and DNMT1, which facilitate the local deacetylation of histones, again resulting in transcriptional repression. Similarly,

recruitment of G9a by DNMT3a and 3b is required for final stages of gene silencing through 5mC and H3K9me [45].

In plants, the KYP (KRYPTONITE/SUVH4), a SUVH protein, contains an N-terminal SRA domain which binds methylated DNA and C-terminal SET domain which methylates H3K9. Recently, crystal structure of KYP complexed with 5mC containing DNA, SAH and unmodified H3 peptide substrate reveal the recognition of 5mC by SRA domain through a flip-out mechanism, where 5mC is flipped out from the DNA duplex and inserted into the pocket within the SRA domain. Structural and *in vivo* studies have shown that KYP is recruited to nucleosomes through binding to methylated DNA and subsequently methylates H3K9 [42]. Therefore, these studies have established that DNA methylation and histone modification pathways are interdependent [45].

#### 1.1.8 UHRF1

UHRF1 is a cell cycle regulator protein which is primarily expressed in proliferating cells and is crucial for S phase entry. UHRF1 mainly resides in a pericentromeric heterochromatin (PCH) and involved in the formation of heterochromatin [46]. UHRF1 is also present in euchromatin region and regulates gene expression; particularly, tumor suppressor genes silencing, probably affecting both DNA methylation and histone modifications. It is up-regulated in breast, prostate, and lung cancer cells. Therefore, it has become a central target for design and development of drug for cancer chemotherapy and a biomarker for diagnosis [46]. Recently, intriguing functions of UHRF1 have been discovered. The most noteworthy is that UHRF1 serves as a sensor for inter-strand cross-links illustrating its role in DNA repair processes. It is characterized as an E3 ligase with auto ubiquitination activity. Its deletion results in genome hypo methylation that leads to cell cycle arrest. Thus, UHRF1 exhibits three important roles during cell proliferation: (1) DNA methylation inheritance (2) DNA crosslinks sensing (3) DNA demethylation facilitation during development.



As an epigenetic regulator, UHRF1 connects DNA methylation with lysine methylation on histone. It binds to H3K9me mark and is also required for DNMT1-mediated DNA methylation through recruitment of DNMT1 to the hemimethylated on replication forks during DNA replication. Apart from DNMT1, UHRF1 also interacts with Proliferating Cell Nuclear Antigen (PCNA), Ubiquitin Specific Processing protease (USP7), HDAC1, HKMTs (SUV39H1, G9a), and HAT (Tip60). UHRF1 and these proteins work together to form a multicomponent complex, termed as Epigenetic Code Replication Machinery (ECREM), for the regulation of DNA replication [46].

### 1.1.9 UHRF1 domain architecture

UHRF1 is a multi-domain protein, composed of five functional domains (Fig. 1.10):

1. UBQ (Ubiquitin-like) domain
2. TTD (Tandem Tudor Domain)
3. PHD (Plant Homeodomain)
4. SRA (SET and RING Associated) domain
5. RING (Really Interesting New Gene) domain.

UHRF1 interacts with several proteins through its domains to form a large macromolecular protein complex called ECREM [47].

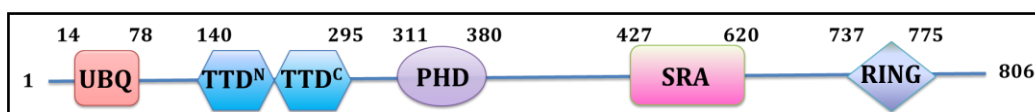


Figure 1.10: UHRF1 domain architecture.

#### 1.1.9.1 Ubiquitin like domain (UBQ)

UBQ like domain is positioned in the N-terminus of UHRF1. This domain contains  $\alpha/\beta$  ubiquitin fold and has conserved surface lysines 31 and 50, which are supposed to be targeted for mono- or poly-ubiquitination (PDB ID: 2FAZ) [48]. UHRF1 controls the ubiquitination state and stability of DNMT1 and the USP7 [48].

#### 1.1.9.2 Tandem Tudor Domain (TTD)

The TTD consists of two subdomains i.e., TTD<sup>N</sup> and TTD<sup>C</sup>. These two domains belong to tudor family 5-stranded  $\beta$ -barrel fold and are tightly packed together. TTD<sup>N</sup> has a cage, formed by aromatic residues F152, Y188 and Y191, which recognizes di and tri-methylated lysine residues [48]. Recognition of H3K9me by the TTD domain, connects DNA methylation to histone modifications [47].

#### 1.1.9.3 Plant Homeodomain (PHD)

Plant Homeodomain (PHD) is a Zn-finger domain consisting of three zinc atoms coordinated to cysteine residues, which stabilize the structure. Crystal structure of PHD domain with bound H3 tail revealed that alanine 1 and arginine 2 on H3 form hydrogen bond interactions to specific residues of the domain. The PHD finger domain of UHRF1 recognizes unmodified arginine 2 of histone H3 (H3R2) [49]. It has been found that PHD of UHRF1 is associated with extensive reorganization of PCH [48,50].

#### 1.1.9.4 SRA domain

The SRA (SET and RING associated) domain of UHRF1 is a DNA-binding module which binds to 5mC and 5hmC containing DNA duplex in context of hemimethylated CpG. As already mentioned, UHRF1 mediates the recruitment of DNMT1 for faithful inheritance of DNA-methylation, more specifically, it is the SRA domain which binds to hemimethylated DNA during replication and maintains the DNA-methylation. The crystal structure analysis of SRA domain (PDB ID: 2PB7) reveals that it has twisted  $\beta$  sheets packed together that appears like a crescent moon-like structure, and only the inner surface is involved in binding to DNA [51]. Two loops sticking out of this structure hold into the DNA minor and major groove. R491 in N-K-R finger loops, forms hydrogen bond interaction with CpG sequences. When SRA domain binds to hemi-methylated DNA, cytosine from the double helix is flipped out because of the contacts discussed above. Two

aromatic residues, Y478 and Y466, are involved in forming  $\pi$ - $\pi$  stacking interactions with the 5mC in pocket. It was established that SRA domain displays highest preference for hemi-methylated DNA (me1/2CpG). This interaction is facilitated by contacts between N489 in N-K-R finger loop and non-methylated cytosine on the second strand of DNA. Methylation of this cytosine disturbs positioning of the N-K-R finger and therefore impairs SRA binding. This cytosine methylation disturbs the N-K-R positioning and impairs SRA binding [48].

#### 1.1.9.5 RING domain

RING domain is situated at C-terminus of UHRF1. It has E3 ubiquitin ligase activity towards histone H3. It contains two zinc-fingers and three  $\alpha$ -helical bundles (PDB ID: 3FL2). A recent report demonstrates that RING domain establishes interactions with H3K23 and H3K18. It was suggested that these marks are required for recruitment of the maintenance DNMT1 to the UHRF1 targeted sites in the genome [48].

## 1.2 Aim of the thesis

Histone and DNA binding by UHRF1 is regulated by long-range inter-domain and inter-domain linker interactions within the full-length protein. UHRF1 TTD-PHD act together to recognize the N-terminus of histone H3 along with H3K9me3, which may play a role in establishing and maintaining histone H3K9 methylation patterns during the cell cycle. Crystal structures of the UHRF1 TTD-PHD bound to H3R2K9me3 peptide show that UHRF1 can simultaneously engage H3R2 and H3K9me3 on a single H3 tail through TTD and PHD linked recognition module. The degree of methylation (also known as methylation status) may confer distinct functions. Recently, it has been demonstrated that UHRF1 TTD-PHD functions as a single histone recognition module in cells, providing a combinatorial readout of a histone H3 tail in ‘cis’ that is required for the epigenetic inheritance of DNA methylation. Different histone modifications can influence each other either in a positive or negative manner through effector/reader-mediated readout.

The aim of this thesis is to gain mechanistic and dynamic insights into the interaction of UHRF1 protein with histone lysine methylation and methylation statuses, and binding to 5mC oxidation derivatives.

The focus of my thesis research is:

1. Characterization of histone lysine methyl mark binding by TTD domain of UHRF1.
2. Molecular Dynamic simulation and binding studies to explore the H3R2K9 methylation statuses recognition by the TTD-PHD of UHRF1.
3. Molecular Dynamic simulation and binding studies to explore the effect of H3K4 methylation on H3R2K9me recognition by the TTD-PHD of UHRF1.
4. Characterization of 5mC oxidation derivative recognition by SRA domain of UHRF1.

## Chapter 2

Characterization of histone lysine methyl mark binding by TTD domain of UHRF1

## 2.1 Introduction

Lysine and arginine methylation pattern on histones and non-histone proteins governs the chromatin structure, genomic stability, transcriptional regulation and RNA metabolism. Such methylation marks are recognized by the reader domains of different families [52]. Among royal family domain, tudor domain is considered as the largest and diverse set of reader domain involved in different epigenetic functions such as chromatin remodeling, binding of histones, pre-RNA-processing, RNA-silencing and transposon silencing [53].

### 2.1.1 Tudor domain

Tudor domain binds to methylated histone tails and facilitates protein-protein interactions [54]. Despite different functions, Tudor domain family proteins have a common feature, i.e., a conserved 50 amino acid domain, which serves to mediate intermolecular protein interactions [55]. The tudor domain is composed of four- or five-stranded  $\beta$ -barrel fold along with one or two helices packed against the  $\beta$ -sheet to form an aromatic cage for the recognition of methylated histone peptides (Fig. 2.1) [54].

Tudor domain containing proteins are categorized into two groups based on the number of tudor domain [54]. First group, those having single tudor domain, for instance, Esa1 and SMN proteins. And other, containing multiple tudor domain repeats, called Tandem Tudor Domains (TTDs), for example, TTDs of spindlin1 and TDRD1. Structures of TTDs of several proteins, like JMJD2A and 53BP1, complexed with H3K4me, H3K9me, and H4K20me peptides, show that only one of the tudor domain in TTDs interacts with methylated histone peptide [54]. Owing to the internal packaging and the linker between two domains, the orientation, of the tudor domains to each other, is distinct among TTDs [52,56].

Tudor domain containing proteins are divided into four groups based on the existing functional data in the literature. Group 1 proteins bind to the methylated lysines/ arginines on histone tails, for instance, Tdrd3, PHF1, PHF20, the JMJD family and TP53BP1. Group 2 tudor proteins recognize methylated arginine and forms complex with pre-mRNA splicing factors which is essential for efficient assembly of small nuclear ribonucleoproteins (snRNPs), for instance, include SMN and SMNDC1. Group 3 tudor proteins, e.g. SND1 is a main component of the RNA-Induced Silencing Complex (RISC) and binds to hyper-edited, double-stranded RNA and promotes its cleavage, and also involved in other cellular pathways. Group 4 tudor proteins, such as Tdrd1-9 and Tdrd11 are identified in methylation-dependent association with PIWI proteins [53].

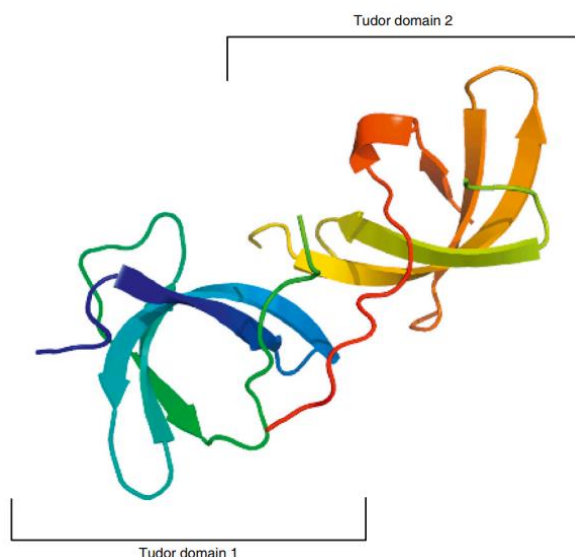


Figure 2.1: Ribbon representation of UHRF1 TTD (PDB ID: 4GY5).

### 2.1.2 Tudor Domain of 53BP1

53BP1 is a p53-binding protein 1, which plays a key role DNA double-strand breaks (DSBs) repair. Domains architecture of 53BP1 is given in Fig. 2.2 [56,57]. 53BP1 TTD binds to H4K20me2 and localizes 53BP1 to the DSB sites thereby linking lysine methylation to chromatin structures and DSB signaling. 53BP1 TTD also binds to dimethyllysine 370 of p53 for the transactivation modulation of p53 on the target sites thereby involved in transcription regulation [58].

53BP1 TTD is also found to recognize H3K79me2 and this interaction was seemingly necessary for targeting DSBs. Since oligomerization of 53BP1 is essential for efficient DSB recognition. Previous studies have reported that the binding pocket contains five aromatic residues, as a result, it can accommodate small sized dimethyl or monomethyl lysine marks [56].

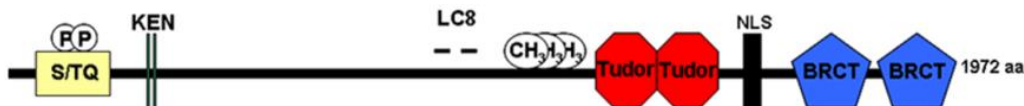


Figure 2.2: Domain architecture of 53BP1.

### 2.1.3 Sequence comparison of UHRF1 TTD and 53BP1 TTD binding pocket

UHRF1 TTD reads the methylation mark on H3K9. Recently, it was demonstrated that the UHRF1 TTD recognize lysine 4, threonine 6, and tri-methylated lysine 9 of the H3K9me3 peptide [46]. TTD simultaneously recognizes H3K9me3 and H3K4 by tudor1 and groove between the tandem tudor domains, respectively [59]. The conserved aromatic residues Phe152, Tyr188 and Tyr191 of N-terminal region of UHRF1 TTD interact with tri-methyl ammonium group of H3K9me3. However, 53BP1 TTD recognizes H4K20me2 by using five aromatic amino acids including Tyr1500, Trp1495, Tyr1502, Phe1519 and Tyr1523 [60]. There is a structural resemblance between the aromatic cages present on N-terminus of UHRF1 TTD (residues 133–217) and 53BP1 (residues 1484–1538) (Fig. 2.3) [59].

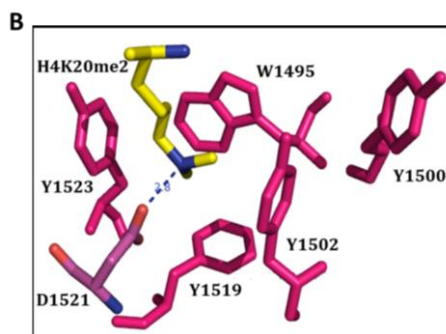




Figure 2.3: Sequence alignment of TTD of UHRF1 and 53BP1 and residues of 53BP1 involved in recognition of H4K20me2. (A) Sequence comparison between Tudor 1 of TTDs of UHRF1 and 53BP1. Secondary structural elements of the UHRF1 TTD are indicated above the sequences ( $\beta$ -strands are in green arrows and loops are in dotted lines). Residues highlighted in a background color-code correspond to their conservation level: fully-conserved, red; conservative substitutions, yellow. Residues that form the aromatic amino acid-lined cage for K9me2 recognition in UHRF1 and 53BP1 are indicated as inverted blue and red upright triangles, respectively. Filled blue and red circles designate the acidic residues, which interact with methyllysine in UHRF1 TTD and 53BP1 TTD, respectively. Green and red rectangles correspond to residues that form van der Waals contacts with the aliphatic and ammonium group of the K9me2 side chain in UHRF1, respectively. (B) Binding pocket residues of 53BP1 TTD involved in recognition of H4K20me2.

Based on the sequence and structural alignment of TTD domain of UHRF1 and 53BP1, we hypothesized that UHRF1 TTD may also recognize di-methyllysine 20 on histone H4. We tested this hypothesis by isothermal titration calorimetry (ITC) binding studies by comparing  $K_D$  of UHRF1 TTD towards H3K9me2 and H4K20me2 peptides.

#### 2.1.4 Structural relationship between tudor and Chromodomain

Chromodomain is a methyllysine reader domain associated with various processes such as gene expression/repression in cellular differentiation, stem cell maintenance and cancer progression. Chromodomain is made of 40 to 60 amino acid residues and has three-stranded anti-parallel  $\beta$ -sheets flanked by a C-terminal  $\alpha$ -helix [61,62]. It recognizes methyllysine histone marks through a shallow surface recognition pocket that consists of 3 to 4 aromatic residues (aromatic cage). Chromodomain containing proteins, HP1, Polycomb group (PcG) and Chromodomain Helicase DNA-binding (CHD) protein, bind to H3K9me3, H3K27me3 and H3K4me3 marks, respectively [61–64].

TTD and chromodomain are structurally superimposable and bind to methylated peptides by using equivalent amino acid position (Fig. 2.4).

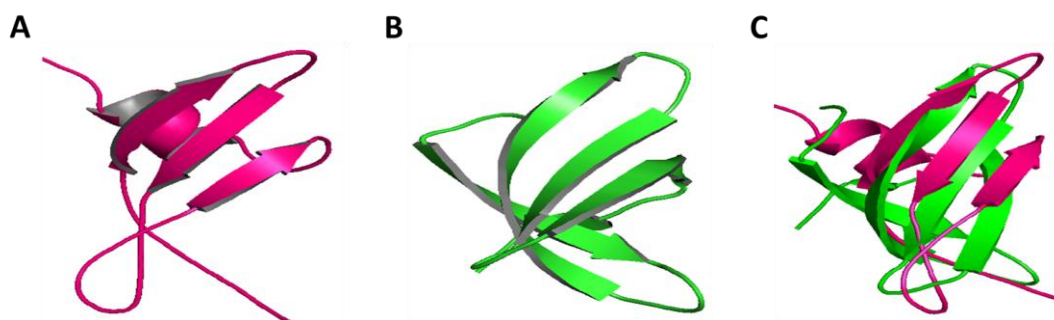


Figure 2.4: Cartoon representation of the (A) Chromo, (B) Tudor domain and their (C) Structural superposition

Analysing the sequence and structural similarity with tudor1 domain of 53BP1 and chromodomain, we hypothesized that UHRF1 TTD may have affinity towards dimethylated lysine present at distinct positions in histones H3 and H4. To investigate this, we performed ITC binding studies of UHTF1 TTD with H3K4me2, H3K9me2 and H3K27me2 peptides.

#### 2.1.5 Role of flanking sequence on methyllysine recognition

Most of the methyllysine binders accommodate methyllysine by using hydrophobic aromatic cage to discriminate different PTMs and methylation marks [24]. Once the methylated peptide binds to the protein, it induces change in  $\beta$ -sheet conformation and form antiparallel alignment to the reader surface groove. This antiparallel interaction enhances the binding affinity and also influences the orientation of methyllysine [24].

On the other hand, flanking-sequence contacts can influence methyllysine binder's selectivity for particular histone lysine methylation mark. On the histone H3 tail, methylation on lysine 9 (H3K9) and lysine 27 (H3K27) are involved in epigenetic repression. The amino acid sequences immediately surrounding H3K9 and H3K27 are very similar and share a consensus sequence motif of ARKS (Fig. 2.5) [24,65,66].

However, consensus motif flanking residues are unrelated. Residues from -2 to +1 positions can make the methylation binders highly selective for particular histone mark [22].

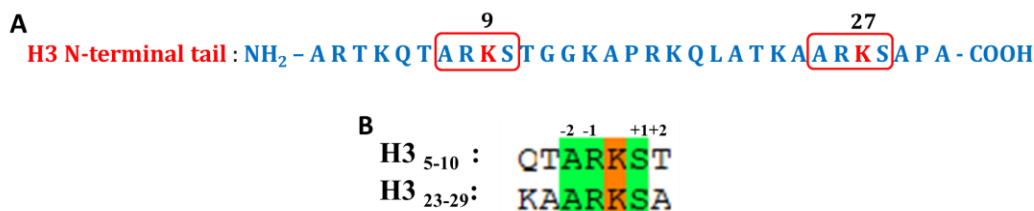


Figure 2.5: Flanking region similarity between H3K9 and H3K27. (A) Sequence of the N-terminus (residues 1–31) of histone H3. The “neighborhoods” of the H3K9 and H3K27 methylation sites are very similar; identical sequence stretches surrounding both sites are boxed and H3K9 and H3K27 are marked in red. (B) Sequence comparison of the H3K9 and H3K27 methylation sites surrounding residues (highlighted with green shade).

Since the -2 to +1 flanking residues of H3K9 and H3K27 are identical, we hypothesized that UHRF1, which binds H3K9me, may also have similar affinity towards H3K27me. To test this, we performed ITC binding studies on UHRF1 with H3K9me2 and H3K27me2 peptides.

### 2.1.6 Methylation status specific readout

Basic amino acid residues on histones such as arginine and lysine undergo methylation. The well-established lysine methylation sites include H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20. Most studied sites of arginine methylation include H3R2, H3R8, H3R17, H3R26 and H4R3. Methylation modifies the hydrophobicity and size of the modified residue. Unlike, lysine acetylation, lysine methylation doesn't have significant effect on histone-DNA interaction. Therefore, histone lysine methylation functions mainly through reader domain. These ‘reader’ modules contain methyllysine-binding motifs, including PHD, chromo, tudor, PWWP, WD40, BAH, ADD, Ankyrin repeat, MBT and Zn-CW domains, and also

have the ability to distinguish target methyllysines based on their status of methylation and surrounding amino-acid sequence.

The chromodomains of HP1 and PcG recognize H3K9me3 and H3K27me3, respectively. The single tudor domains of PHF1 and PHF19 are recognize H3K36me3, whereas the TTD of 53BP1 binds to H4K20me2 and the hybrid TTD domain of JMJD2A recognises H3K4me3 and H4K20me3. The other two members of the Royal family, PWWP and chromo-barrel preferentially bind H3K36me3/me2 and H4K20me1 respectively. PHD finger is different from the royal family and it is well-characterized reader of H3K4me3. BPTF PHD binds tightly to H3K4me3 compared to H3K4me2 and is associated with chromatin remodeling. The WD40 domain of EED interacts with H3K27me3, H3K9me3, H4K20me3 and H1K26me3. The G9a and GLP ankyrin repeats and the ORC1 BAH domain shown to recognize high specificity towards H3K9me2/me1 and H4K20me2, respectively.

Depending on the position and degree of methylation (mono, di or tri), histone lysine methylation can have role on repression or activation of transcription. Effect of position of methylation is as follows; H3K4me3 is normally associated with active transcription, whereas H3K27me3 is involved in chromatin repression. Effect of degree of methylation is as follows; H3K4me1 is mainly associated with enhancer activity whereas H3K4me3 is connected to promoter function; H3K79me2 is associated in cell cycle regulation whereas H3K79me3 is connected to the Wnt signaling pathway [67,68]. However, there exist some examples that can be associated with opposite activities; for instance, H3K4me2 and H3K4me3 are involved in transcriptional activation and/or repression [67].

The high density of H3K36me3 was found on 3' end of an actively expressed gene. On the other hand, H3K9me3 and H4K20me3 act as repressive marks and are involved in constitutive heterochromatin formation, and H3K27me3 is involved in facultative heterochromatin. Outcome of particular lysine methylation mark

readout is also dependent on the type of the effector module recognizing the mark (Table 2.1) [7]. For example, recognition of H3K4me2 by the Tudor domain of JMJD2A module promotes the H3K9me3 demethylation. In contrast, recognition of same mark by the tudor domain of Spindlin1 is required for rRNA transcription. Thus reader domains act as important determinant for the functional outcome of histone modifications [67].

Table 2.1: Reader domains that recognize different lysine methylation.

S.No	Protein	Recognition Module	Ligand	Biological functions
1.	JMJD2A	Tudor	H3K4me3 H4K20me3 H4K20me2	H3K9me3 and H3K36me3-specific demethylase; transcriptional regulation and regulator of DNA damage response
2.	53BP1	Tudor	H4K20me2	Substrate of ATM; Promote nonhomologous end joining DNA repair
3.	Spindlin1	Tudor	H3K4me3	Nucleolar protein; promote rRNA transcription
4.	PHF1	Tudor	H3K36me3	Accessory component of PRC2 complex; promote Transcriptional repression
5.	ORC1	BAH	H4K20me2	DNA replication licensing
6.	HP1	Chromo	H3K9me3	Heterochromatin assembly and maintenance
7.	Polycomb	Chromo	H3K27me3	Cellular differentiation
8.	CHD1	DCD	H3K4me3	Chromatin remodelling
9.	MRG15	Chromo	H3K36me3	RNA splicing
10.	BPTF	PHD	H3K4me3	Nucleosome Sliding
11.	L3MBTL1	MBT	H4K20me2	Chromatin compaction
12.	Pdp1	PWWP	H4K20me3	Chromatin localization of H4K20 methyltransferase

Based on these information, we investigated the binding affinities of UHRF1 TTD for mono, di and trimethylated H3K9 (i.e. H3K9me0/1/2/3) and H4K20 (i.e. H4K20me1/2/3), to understand the preferred methylation status readout by UHRF1 TTD.

### 2.1.7 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a powerful, label free and sensitive technique for studying the thermodynamics of macromolecular interactions such as proteins with peptide, nucleic acid, drugs and metal ions. In ITC experiments two reactants are titrated against one another and their binding is determined by direct measurement of heat exchange [69,70]. We can calculate thermodynamic parameters such as enthalpy ( $\Delta H$ ), entropy ( $\Delta S$ ), association constant ( $K_A$ ), binding stoichiometry ( $n$ ), free energy of binding ( $\Delta G$ ), and potential site-site interactions (cooperativity) in a single experiment [71].

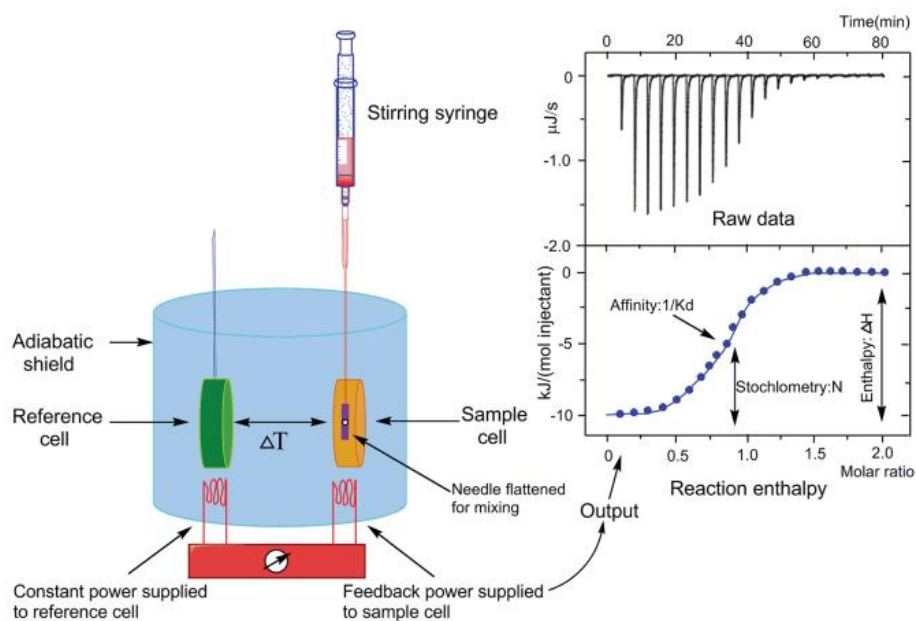


Figure 2.6: Schematic representation of the Basic principle of isothermal titration calorimetry (left), a characteristic curve of the titration experiment (upper right) with its convolution integral enthalpy plot (lower right).

The ITC instrument consists of a reference cell and a reaction cell. They are made up of a conductive inert (non-reactive) metal (Fig. 2.6). The cells are connected by a thermoelectric device (TED) that is sensitive to temperature and is connected to a feedback power supply. Constant power supply is applied to the sample cell heater. The temperature difference between the two cells is detected by the thermocouples and it is measured. The advantage of ITC is that it determines the direct enthalpy which is important for drug design.

## 2.2 Materials and Methods

### 2.2.1 Sub-Cloning of UHRF1 TTD to generate hexahistidine-SUMO tagged construct

The cDNA encoding full-length human UHRF1 was obtained from Open biosystems. Residues 140-295, that corresponds to the UHRF1 TTD, were sub-cloned. Amplification was carried out using forward and reverse primers (Appendix I) containing NdeI and BamHI restriction sites, respectively. The amplified PCR product was confirmed by agarose gel electrophoresis, followed by DNA gel extraction and subsequent digestion with NdeI and BamHI. The purified digested product was ligated with NdeI-BamHI digested pET28-N-His-SUMO vector using T4 DNA ligase. Schematic represents the workflow of cloning strategy (Fig. 2.7). Clones obtained were screened by using colony PCR and confirmed by sequencing.

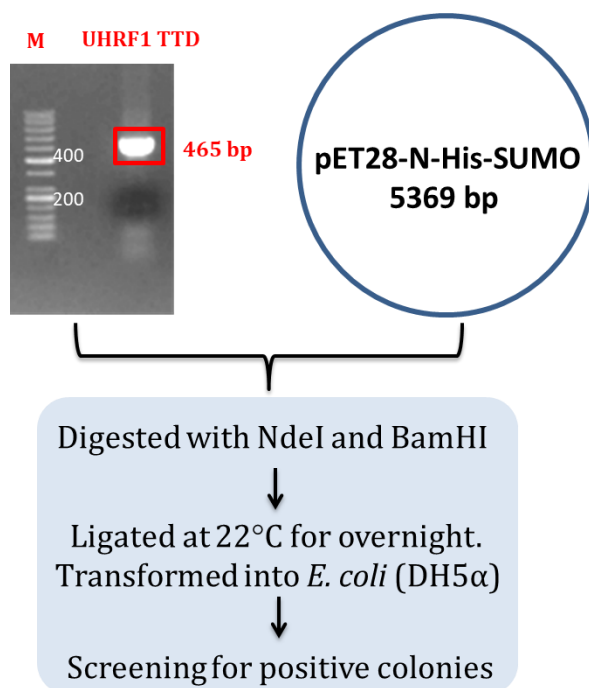


Figure 2.7: Flowchart of UHRF1 TTD sub-cloning.

### 2.2.2 Expression and Purification of UHRF1 TTD

UHRF1 TTD protein was expressed in *E. coli* Rosetta2(DE3) (Novagen). Cells were grown at 37°C, till the OD<sub>600</sub> reached 0.5-0.6, then the temperature was decreased to 20°C and the culture was induced with 0.4 mM of isopropyl-1-thio-D-galactopyranoside (IPTG). The cell culture was grown for 15 hrs, following which the cells were harvested and re-suspended in lysis buffer (25 mM Tris-HCl, pH 7.5, 500 mM NaCl, 10 mM imidazole and 3 mM β-mercaptoethanol). Cells were lysed by ultrasonic homogenizer and then the lysate was clarified by centrifugation at 40,000g for 1 hr. The hexahistidine-SUMO fusion protein was purified on a nickel-charged column (HisTrap HP, GE healthcare). After elution with a 750 mM imidazole containing buffer, the fusion protein was cleaved with Ulp1 protease at 25 Uml-1 to remove hexahistidine-SUMO tag, during a 16 hr dialysis using dialysis tube (5kDa Cutoff), at 4°C. The protein was further purified by gel filtration chromatography (HiLoad Superdex 75 16/600) and equilibrated with equilibration buffer (15 mM Tris-HCl, pH 7.5, 50 mM NaCl, 3 mM DTT). Purified UHRF1 TTD



was concentrated to  $20 \text{ mg mL}^{-1}$  at  $4^\circ\text{C}$  in Vivaspin 20 mL (Vivascience AG) 5,000 cut-off concentrator. 12% SDS-PAGE gel of gel filtration purified UHRF1 TTD along with the chromatogram showing the elution profile, is given in Fig. 2.8.

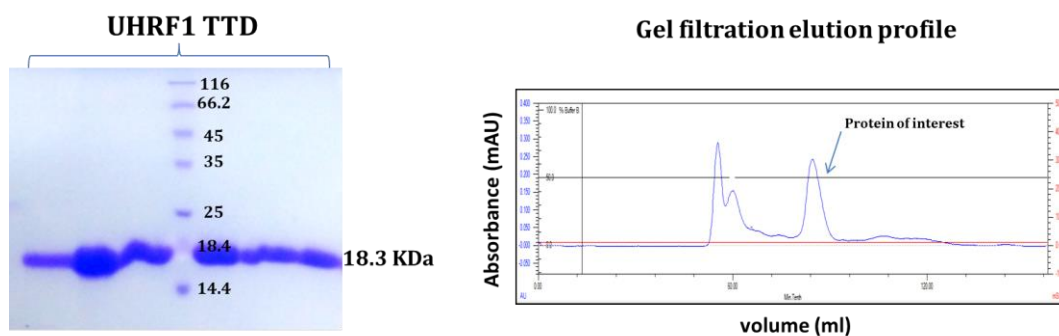


Figure 2.8: SDS-PAGE of gel filtrations purified UHRF1 TTD and the chromatogram showing its elution profile.

### 2.2.3 ITC measurements

The equilibrium dissociation constant ( $K_D$ ), molar ratio ( $N$ ) and thermodynamic parameters of the UHRF1 TTD bound to methylated lysine or unmethylated H3 or H4 histone peptides (H3[1-12]K9me0/1/2/3, H3[1-12]K4me2, H3[20-34]K27me2, H4[1-23]K20me1/2/3) were determined using a VP-ITC calorimeter (MicroCal, LLC) at  $25^\circ\text{C}$ . The protein was dialyzed in dialysis tube overnight using dialysis buffer (40 mM Tris-HCl, pH 7.5, 50 mM NaCl, and 2 mM  $\beta$ -mercaptoethanol), at  $4^\circ\text{C}$ . Lyophilized peptides were dissolved in same buffer used for protein dialysis. The protein and peptides concentrations used were  $100 \mu\text{M}$  to  $250 \mu\text{M}$  and 1 mM to 2.5 mM, respectively. The final volume of the reaction cell was  $210 \mu\text{L}$ , and the reference cell was filled with deionized water. The peptide was sequentially added in  $2.3 \mu\text{L}$  (for a total of 15-16 injections) aliquots at 3-min intervals. The data was processed using MicroCal Origin software. The ITC data was deconvoluted based on a binding model containing “one set of sites” using a nonlinear least-squares algorithm. The binding enthalpy change ( $\Delta H$ ), association constant ( $K_A$ ), and binding stoichiometry ( $N$ ) were permitted to vary during the least-squares

minimization process and taken as the best-fit values for UHRF1 TTD bound to the histone peptides. In the cases where the C-value (which is the product of the receptor concentration and the binding constant,  $K_A$ ) is low, 'N' was fixed to 1.0, and ' $K_A$ ' and ' $\Delta H$ ' were permitted to float. The reported values are the best values from three titrations. The thermodynamic parameters for histone peptides binding to UHRF1 are provided in Table 2.2.

## 2.3 Results

### 2.3.1 UHRF1 TTD selectively recognizes K9 methylation mark on histone H3

Sequence comparison of UHRF1 TTD with that of 53BP1 reveals that out of five as in 53BP1, UHRF1 TTD has three conserved aromatic residues that form the methyllysine binding hydrophobic cage. Thus, the UHRF1 TTD should be able to recognize methylated lysine histone peptides, as that of 53BP1 TTD. We used ITC-based binding studies to monitor the potential sequence specific recognition. Our binding measurements reveal that the UHRF1 TTD binds to H3(1-12)K9me2 peptide with a  $K_D$  of 1.1  $\mu\text{M}$  with 1:1 stoichiometry (Fig. 2.9A). The binding affinities for H3(1-12)K4me2 and H3(20-34)K27me2 with a  $K_D$  of 85.0  $\mu\text{M}$  and 62.0  $\mu\text{M}$ , respectively, are weaker than H3K9me2 (Fig. 2.9B and C). Thus, the UHRF1 TTD exhibit highest specificity for methylated H3K9 mark on histone H3, with a 104 and 66-fold preference over methylated H3K4 and methylated H3K27, respectively. The thermodynamic parameters for histone peptides binding to UHRF1 are given in Table 2.2.

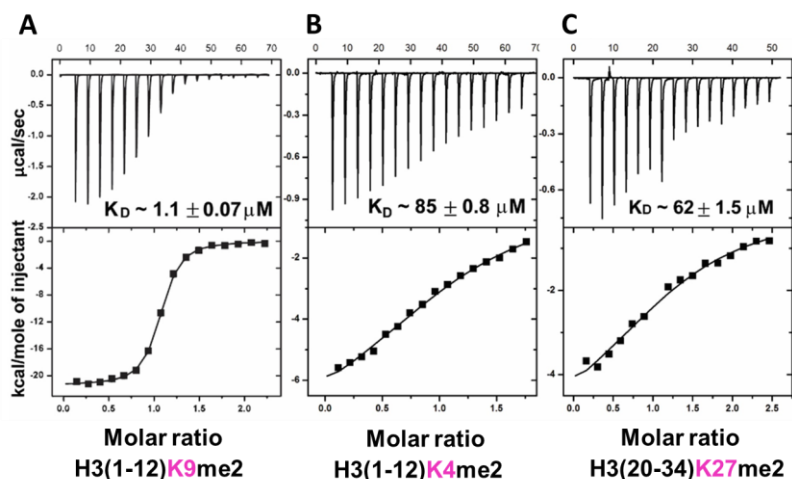


Figure 2.9: Raw ITC data (upper panel) and normalized integration data (lower panel) of enthalpy plots the binding of the UHRF1 TTD to (A) H3K9me2 (B) H3K4me2 (C) H3K27me2 peptides. Binding stoichiometry ( $N$ ) is 1 for ITC measurements.

### 2.3.2 Recognition of H4K20me mark by the UHRF1 TTD

Structure based sequence comparison of UHRF1 TTD with that of 53BP1 indicated that residues of 53BP1 involved in H4K20me2 recognition are significantly conserved in UHRF1 (Fig. 2.3A). Therefore, we presume that UHRF1 TTD may also recognize the H4K20me mark. Indeed, ITC-based binding studies establish that the UHRF1 TTD binds H4(1-23)K20me2 peptide with a  $K_D$  of  $9.8 \mu\text{M}$  (Fig. 2.10A). UHRF1 TTD has very weak binding affinities for K20me3 ( $K_D$  of  $55 \mu\text{M}$ ) and K20me1 ( $K_D$  of  $101 \mu\text{M}$ ) peptides (Fig. 2.10B and C). However, the binding affinity of UHRF1 TTD for H4(1-23)K20me2 is 9 times lower than H3(1-12)K9me2 peptide.

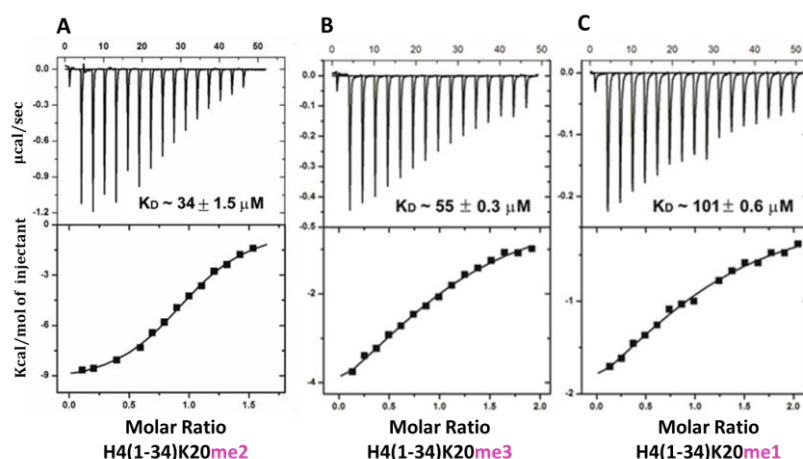


Figure 2.10: Raw ITC data (upper panel) and normalized integration data (lower panel) of enthalpy plots for the binding of the UHRF1 TTD to (A) H4K20me2, (B) H4K20me3 and (C) H4K20me1 peptides.

### 2.3.3 Methylation status specific readout of H3K9me by UHRF1 TTD

As previously reported [72], ITC-based binding studies establish that the UHRF1 TTD binds H3(1-12)K9me3 peptide with a  $K_D$  of  $5.6 \mu\text{M}$ , which is 5-fold weaker than binding to H3(1-12)K9me2 peptide (Fig. 2.9A and 2.11A). Similarly, the binding is also weaker for methylation statuses lower than dimethylation ( $K_D$  of  $12.0 \mu\text{M}$  for H3(1-12)K9me1 (Fig. 2.11B) and  $160 \mu\text{M}$  for H3(1-12)K9me0 peptides) (Fig. 2.11C)}. The thermodynamic parameters for histone peptides binding to UHRF1 are provided in Table 2.2.

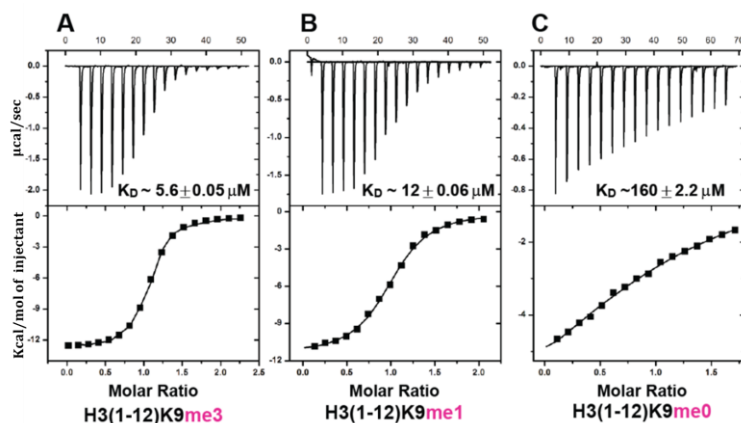


Figure 2.11: Raw ITC data (upper panel) and normalized integration data (lower panel) of enthalpy plots for binding of UHRF1 TTD to (A) H3K9me3, (B) H3K9me1 and (C) H3K9me0 peptides.

Table 2.2: Dissociation constant and thermodynamic data for binding of methylated or unmethylated lysines of histone H3 and H4 peptides to the UHRF1 TTD. Errors reported for  $K_D$  and sum of the squared deviations between the data and the model curve.

Peptides	$K_D$ ( $\mu\text{M}$ )	$\Delta H$ (kcal/mol)	$T\Delta S$ (kcal/mol)
UHRF1 TTD			
H3(1-12)K9me2	$1.1 \pm 0.07$	$-21 \pm 0.01$	-1.08
H3(1-12)K4me2	$85 \pm 0.8$	$-8.06 \pm 0.1$	-0.19
H3(20-34)K27me2	$62 \pm 1.5$	$-5.8 \pm 0.2$	-0.0
H3(1-12)K9me3	$5.6 \pm 0.05$	$-13.1 \pm 0.09$	-0.4
H3(1-12)K9me1	$12.0 \pm 0.06$	$-11.5 \pm 0.1$	-0.36
H3(1-12)K9me0	$160 \pm 2.2$	$-8.9 \pm 0.7$	-0.31
H4(1-34)K20me2	$34 \pm 1.5$	$-10.5 \pm 0.15$	0.1

## 2.4 Discussion

Our ITC binding studies show that UHRF1 TTD, like 53BP1 TTD, binds to H4K20me2 with high affinity ( $K_D = 34 \mu\text{M}$ ). Since, it is also structurally similar to chromodomain, which recognizes methyllysine present at different positions in H3 and H4, our study shows that among H3K4me2, H3K9me2, H3K27me2 and H4K20me2, UHRF1 has higher affinity for H3K9me2 ( $K_D = 1.1 \mu\text{M}$ ). For rest other dimethyllysine, it shows moderate binding affinity (Table 2.2 and Fig. 2.9A). These results indicate that UHRF1 TTD selectively recognize the dimethyl mark on H3K9.

Then we investigated the methyl status specific readout by UHRF1 TTD. Our ITC studies on H3K9 with different methyl status shows that it has a comparable binding affinity for H3K9me3 ( $K_D = 5.6 \mu\text{M}$ ) and H3K9me2 ( $K_D = 1.1 \mu\text{M}$ ). These results indicate that UHRF1 TTD binds to both H3K9me3 and H3K9me2

methylation marks but has higher preference for H3K9me2. Higher preference of UHRF1 TTD for lower methylation status (dimethyllysine) may be due to presence of negatively charged residues in the binding pocket, as in 53BP1. Our ITC binding results are supported by the lower lysine methylation status-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger where presence of an acidic residue in the aromatic cage in these reader modules restrict the binding pocket to recognize the lower methylation status, di and monomethyllysine, of H1.4K26, H3K4, H3K9, H3K27, H3K36 and H4K20.

We also hypothesized that, since, the flanking residues may also contribute towards the recognition of epigenetics mark, whereas -2 to +1 flanking residues of H3K9 and H3K27 are identical, UHRF1 TTD, which binds to H3K9me2 with highest affinity, may also bind to H3K27me2 with similar affinity. On the contrary, our ITC studies on UHRF1 TTD with H3(1-12)K9me2 and H3(20-34)K27me2 peptides show that it has 60 times less affinity for H3K27me2 ( $K_D = 62 \mu\text{M}$ ). This result contradicts our hypothesis. One possible explanation may be the variation in stretched flanking residues.

Thus, UHRF1 TTD as a stand-alone domain binds to H3K9me2 with a significantly higher binding affinity compared to H3K4me2, H3K27me2 and H4K20me2. It is possible that UHRF1 TTD can nonspecifically bind other methyllysine, H3K4me2 and H3K27me2 marks in different amino acids sequence contexts.

In this study, we have found that UHRF1 TTD preferentially recognizes H3K9me2 mark. In UHRF1, TTD<sup>C</sup> is linked to PHD domain which is known to assist the UHRF1 TTD to recognize H3K9me. In the next chapter, we have investigated the mechanistic insights of the preferential recognition of H3R2K9me2 by UHRF1 TTD-PHD dual domain, using computational (molecular modelling and molecular dynamic simulation) and biochemical approaches.

## Chapter 3

Molecular Dynamic simulation and binding studies to explore the H3R2K9 methylation statuses recognition by the TTD-PHD of UHRF1

### 3.1 Introduction

Many effector modules function together through recognition of two or more histone PTMs on the histone tails. The paired effector modules appear in the different combinations such as reader-reader, reader-writer and reader-eraser pairs [73]. Combinatorial actions of these effector modules is involved in the recruitment of particular chromatin-associated proteins to the specific genomic region for further modulation [74].

A number of chromatin-associated proteins contain more than one reader domain, in the form of the same reader domain and/or a combination of different reader domains which are specific for particular histone PTMs [22]. More than 10 families of different histone modification-specific reader domains, that bind to the same histone tail or multiple histone tails in *cis* or *trans* positions, have been structurally characterized (Figure 3.1) [73].

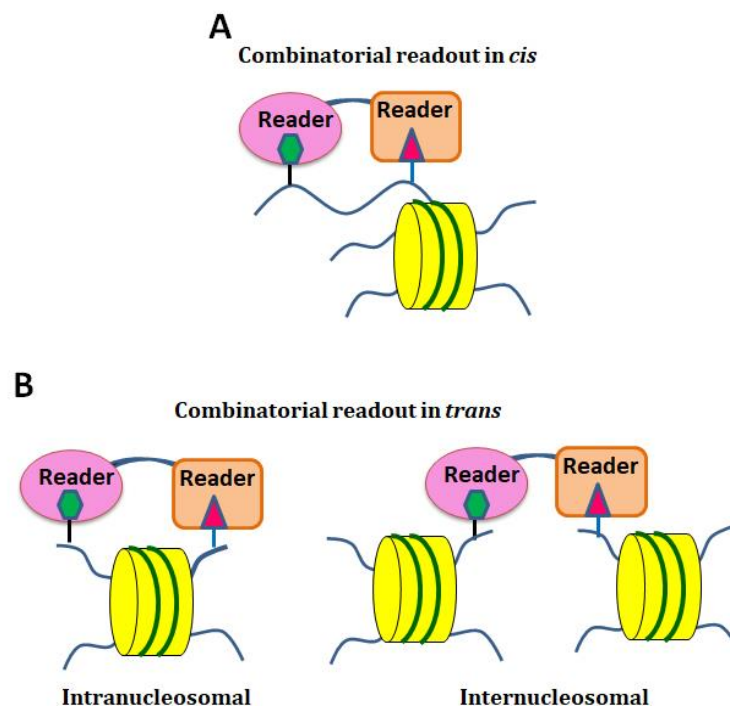


Figure 3.1: Schematic representation of combinatorial recognition of histone PTMs by multiple domains of same protein. (A) In *cis* mode where PTMs on same histone



is read and (B) In *Trans* where PTMs on different histones is read. The first and second panel in section B represents the combinatorial binding of intra-nucleosomal and inter-nucleosomal histone modification by two reader domains of same protein.

### 3.1.1 Combinatorial recognition of histone marks in single histone tail (*cis*)

Epigenetic marks on the same histone tails can be recognized by the multiple tandem linked reader domains. Combinatorial recognition of multiple histone PTMs is supported by the coexistence of different effector domains in different combination within a given polypeptides or complexes involved in epigenetic regulation (Fig. 3.2A and B) [75].

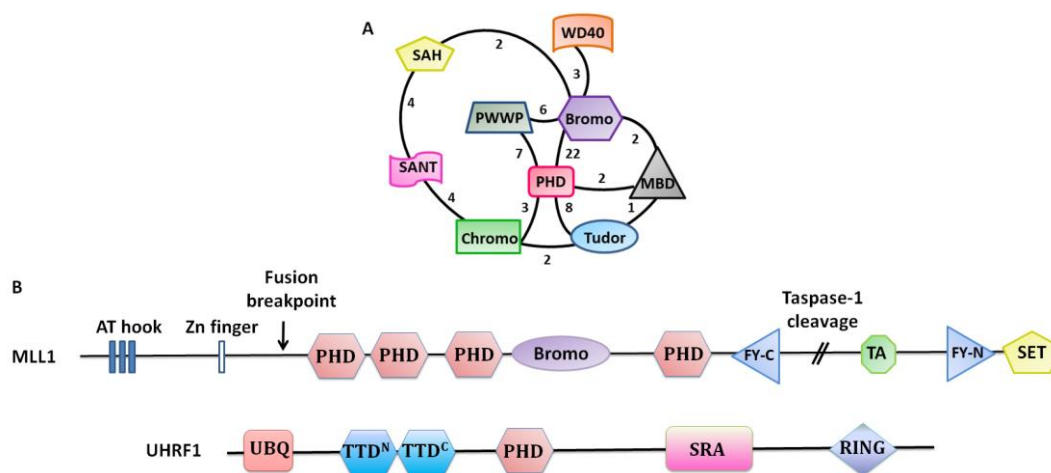


Figure 3.2: (A) Schematic representation of the coexistence of possible reader domains within single polypeptides. Numbers in the line connecting the domains indicate the number of instances any of the two domains are linked. (B) Depicts the examples of proteins containing multiple reader domains.

[Adapted from Ruthenberg et al., 2015]

Simultaneous recognition of multiple histone PTMs by the reader domains of the multi-domain protein or complex leads to high affinity binding compared to single readout. Thus, two or more reader domains recognizing multiple histone codes, cooperatively, would enhance the binding affinity. For instance, TAF1 (TBP-associated factor, 250 kDa), the largest subunit of TFIID (RNA polymerase II transcription factor D) is the first reported paired chromatin associated reader

domain. It contains two bromodomains that simultaneously engage two or more acetyllysine histone marks (Fig. 3.3A) [73]. The two bromodomains exhibit a binding affinity of 39  $\mu$ M towards the H4K16ac peptide, which was enhanced by 7 to 28 folds for di/tetra-acetyllysine-modified H4 peptides (H4K8ac/K16ac and H4K5ac/K8ac/K12ac/K16ac), respectively [76]. The linked bromodomains of TAF1 simultaneously recognize two acetyllysine mark on a single histone tail. Unlike TAF1, TRIM24 and TRIM33 binds to two different histone PTMs by distinct set of reader domains called PHD-bromodomain cassettes. The TRIM33 PHD binds to K9me3, and the bromodomain recognizes K18ac of the H3 tail [73,75,76]. TRIM24 PHD and bromodomain associates with H3K4me0 and H3K23ac, respectively (Fig. 3.3B and C) [74].

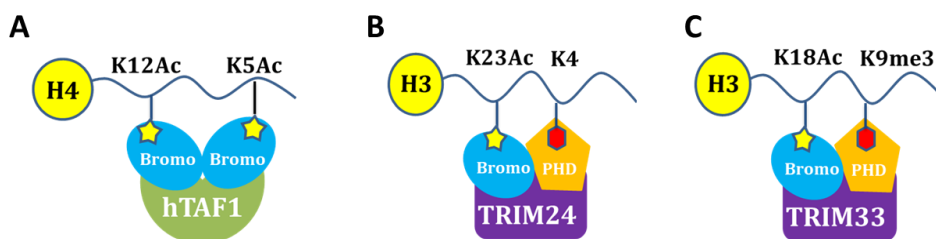


Figure 3.3: Schematic representation of combinatorial readout of two histone marks by paired chromatin-associated reader domains in *cis* mode. (A) TAF1 double bromodomains reads H4K5ac and H4K12ac marks. (B) TRIM24 PHD-bromodomain cassette reads H3K4me0 and H3K23ac modification pair and (C) The TRIM33 PHD-bromodomain cassette reads H3K9me3 and H3K18ac modification pair.

### 3.1.2 Combinatorial recognition of PTMs in different histone tails (*trans*)

Some chromatin associated proteins recognize multiple PTMs on separate histone tails that may be on single nucleosome or on adjacent nucleosomes. These histone marks are present in close proximity for the recognition by reader domains. For instance, BPTF (Bromodomain PHD finger Transcription Factor), a subunit of NURF chromatin remodeling complex, contains PHD linked bromodomain simultaneously recognizes H3K4me3 and H4K12ac, H4K16ac or H4K20ac, respectively (Fig. 3.4A) [74,75]. Some reader domains are linked by a short linker

region in such a way that they can bind histone marks from different nucleosome in the *trans* mode. For instance, CHD4 contains two PHD finger domains connected by short linker. The PHD1 preferentially binds to unmodified H3K4 and PHD2 binds H3K9me3 (Fig. 3.4B) [74,75,77].

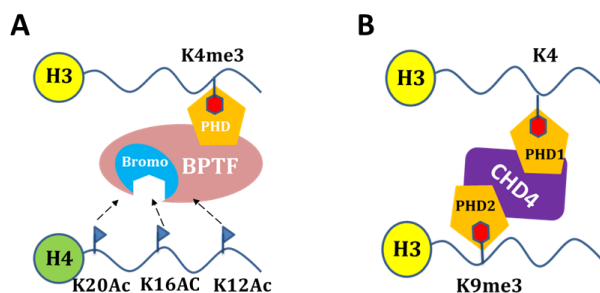


Figure 3.4: Schematic representation of combinatorial readout of two histone marks by paired chromatin-associated domains in *trans*. (A) The BPTF PHD-bromodomain cassette reads H3K4me3 and H4K12ac modification pair. (B) Two PHD domains of CHD4 read H3K4 and K9me3 marks on two different H3.

Thus, the linked reader domain shows the multivalent recognition of modified histone peptide interaction in *cis* and *trans* position. UHRF1 also has multiple reader domains that can exhibit multivalent histone PTMs recognition.

### 3.1.3 Multivalent histone engagement of UHRF1

UHRF1 contains two reader domains, namely, TTD and PHD. The UHRF1 TTD and PHD are separated by a 17 amino acids linker. The association of domains is maintained by extensive contacts between TTD and the linker residues [78]. UHRF1 TTD-PHD is known to be involved in combinatorial recognition of unmodified R2 and K9me3 (H3R2K9me3) marks on histone H3 tail (Fig. 3.5) [72].

UHRF1 TTD is reported to bind the N-terminal H3(1-10)K9me3 with the binding affinity of 0.97 $\mu$ M, and PHD binds to unmodified H3R2 with a  $K_D$  of 2.1  $\mu$ M. Recently, it was found that the linked TTD-PHD cassette binds H3K9me3 with a binding affinity of 0.15  $\mu$ M which is 6 fold higher than isolated TTD alone [22,79]. It indicates that the interaction between the PHD and the N-terminus of H3 is

important for H3K9me3 recognition by TTD-PHD. Importantly, disruption of the function of individual reader domain, or disturbing the dual domain function to engage the H3 tail in *cis*, was shown to inhibit the interaction between UHRF1 and the chromatin [80]. It suggests that the combinatorial interaction of H3R2K9me3 marks containing H3 histone tail by UHRF1 TTD-PHD involved in the role in establishment and maintenance of H3K9me mark in the cell cycle regulation[78].

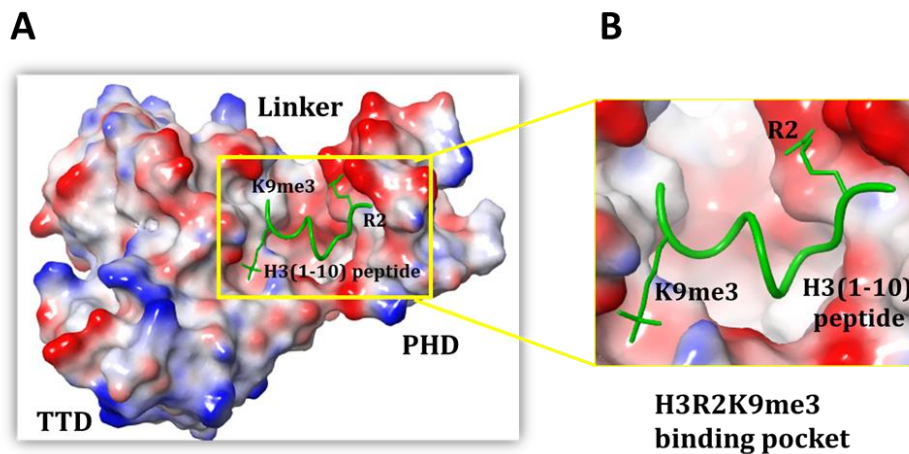


Figure 3.5: Structure of the UHRF1 TTD-PHD bound to H3 (1-10) K9me3 peptide.

(A) Protein and peptide are in electrostatic surface model and cartoon representation, respectively. H3K9me3 and H3R2 are in stick representation. (B) Magnified view of the peptide in UHRF1 TTD-PHD-H3R2K9me3 binding pocket.

It is reported that recognition of first two N-terminal amino acids, i.e. alanine 1 and arginine 2 of H3 (H3A1R2), by PHD domain induces the conformational change in the linker region and the TTD domain, so that TTD can specifically bind to methylated H3K9 [79,81]. It suggests that PHD is necessary and sufficient for H3R2 recognition, whereas both the PHD and TTD are required for H3R2K9me3 recognition. Thus, TTD and PHD work together to interact H3R2K9me and localize the UHRF1 to PCH, which is important for DNA methylation maintenance by DNMT1 during DNA replication. The independent recognition of H3R2 mark by PHD domain is important for regulation of the euchromatic gene expression by UHRF1 [46].

From the binding studies discussed in previous chapter, we know that UHRF1 TTD has the highest binding affinity for H3K9me2 mark. Also, the role of UHRF1 PHD to aid UHRF1 TTD in recognition of H3K9me3 through H3R2 binding is well established. Based on these information, we hypothesized that UHRF1 PHD may contribute significantly towards the combinatorial recognition of H3K9me2 and H3R2 by UHRF1 TTD-PHD cassette.

To validate the hypothesis, MD simulation studies were carried out to: (A) capture the dynamic of TTD-PHD binding to H3R2K9me peptides in different methylation statuses; (B) identify the dynamics of binding pocket residues contributing to peptide recognition; (C) compute free energy of binding of TTD-PHD to H3 peptides, along with ITC and FP. Through MD simulation studies, it was found that Asp145 may contribute for preferential recognition of H3K9me2. To confirm the role of Asp145, present in the binding pocket of UHRF1 TTD, in preferential recognition of H3K9me2, binding and MD simulation studies using Asp145 mutants are performed. Also, contribution of PHD finger of UHRF1 TTD-PHD to recognize H3K9me2 is studied.

We used MD simulation for computational characterization of peptide-protein interactions. ITC and Fluorescence Polarization techniques were used for binding studies. Principle of ITC technique is discussed in section 2.1.7, whereas MD simulations and Fluorescence Polarization principles are discussed below.

### 3.1.4 Fluorescence polarization

Fluorescence polarization (FP) is a homogenous and powerful technique in which the polarization of the fluorescent light emitted from the sample is measured. It is based on the rotational behavior of the labelled molecule. A monochrome light of certain wavelength (absorption wavelength) is incident on the sample, which in turns emits light of higher wavelength (emission wavelength). Polarization is an inverse function of rotational dynamics of the molecule, which is, again, an inverse

function of the molecular weight (MW). This polarization value gives the kinetic details of the interaction. Increase in the MW reduces the rotation of the molecule and hence increase the polarization (Fig. 3.6). In the biological processes, including protein-peptide, protein-protein, protein-nucleic acid and protein-small molecules, the MW of the complex increases upon interaction [82].

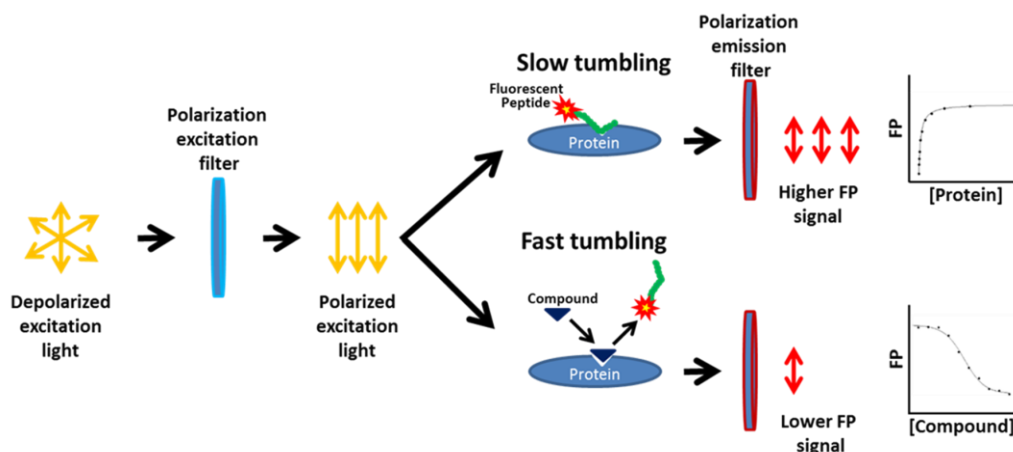


Figure 3.6: Principle of Fluorescence Polarization

Fluorescence polarization is used to study the biological processes that involve changes in molecular weight, such as a binding event or the enzymatic cleavage of a substrate. FP is used to measure emission intensities such as parallel ( $I_{||}$ ) and perpendicular ( $I_{\perp}$ ) of the fluorescent sample which is excited by the polarized light and calculate fluorescence anisotropy (FA) or fluorescence polarization (FP) (Fig. 3.6) [82,83]. The equations of FP and FA are given below.

$$FP = \frac{I_{||} - I_{\perp}}{I_{||} + I_{\perp}} \quad FA = \frac{I_{||} - I_{\perp}}{I_{||} + 2I_{\perp}}$$

### 3.1.5 Molecular Dynamic Simulation

Proteins undergo different conformational changes and they interact with one another, with drugs and hormones dynamically. An MD simulation is a standard method for simulating dynamic motion of the molecules. It helps in visualizing and

understanding the dynamic behavior of the system at an atomic level, under a given temperature, pressure and energy constants. When combined with a visualization software, it can display the structural parameters in a time dependent way. MD simulation routinely calculates the following quantities: 1) thermodynamics of system; 2) interface related terms like order parameter, density of groups, electrostatic potential; 3) Time average structure; 4) RMSD difference between two structures; and 5) Radius of Gyration [84–86]. Schematic representation of the steps involved in MD simulation of protein is shown in the fig. 3.7

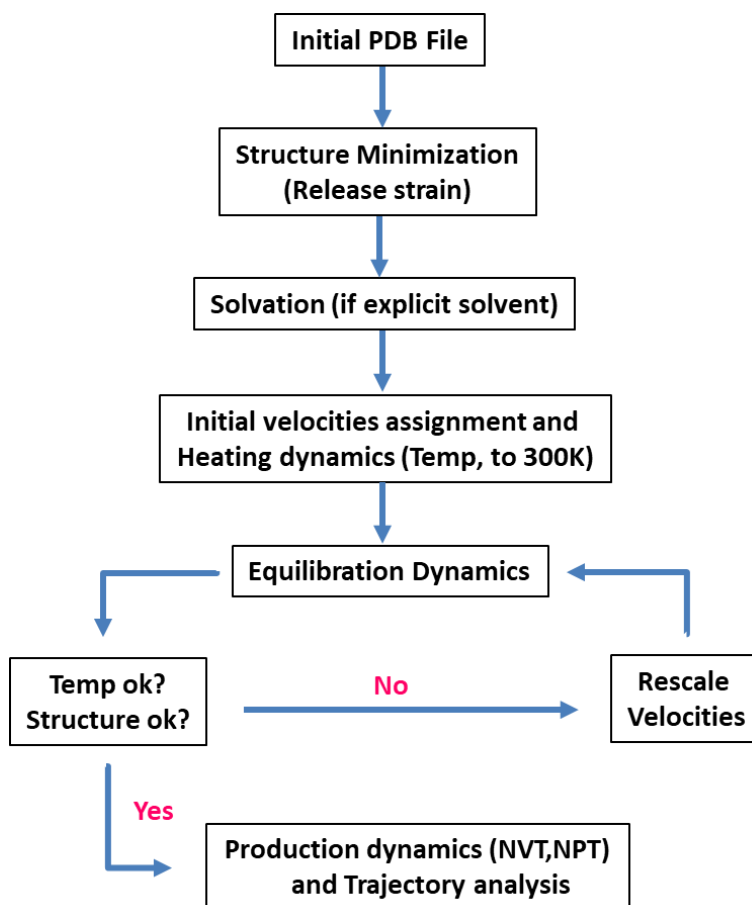


Figure 3.7: Schematic depicts steps involved in MD simulation of Protein

## 3.2 Material and Methods

### 3.2.1 Preparation of UHRF1 TTD-PHD-H3K9me2 complex for simulation

The crystal structure of UHRF1 TTD-PHD with H3(1-10)K9me3 (PDB ID: 4GY5) [46] was prepared for MD simulation using Schrodinger's Maestro Molecular modeling suite. The H3K9me3 was modified to H3K9me2 by virtually editing the methylation marks. Two amino acids threonine 11 and glycine 12 were also added virtually to the existing H3(1-10) peptide to get H3(1-12) peptide and the added amino acids were energy minimized. Bond orders were assigned and hydrogen atoms were added followed by H-bond optimization and restrained minimization using OPLS3 force field. Protonation statuses were determined at physiological pH 7.0 using PROPKA [87,88]. All the water molecules were removed.

### 3.2.2 Preparation of UHRF1 TTD-PHD Asp145Glu and Asp145Ala mutants for simulations

The crystal structure of UHRF1 TTD-PHD complexed with H3(1-10)K9me3 (PDB ID: 4GY5) was used to prepare Asp145Glu and Asp145Ala mutants. The mutation was carried out using Schrodinger's Maestro Molecular modeling suit. Then the methylation marks on H3K9 were virtually edited to generate different combinations of protein-peptide complex, i.e. UHRF1 TTD-PHD Asp145Glu and UHRF1 TTD-PHD Asp145Ala mutants complexed with H3(1-12)K9me2 and H3(1-12)K9me3 peptides. All the parameters were set as that of UHRF1 TTD-PHD-H3(1-12)K9me2 complex.

### 3.2.3 Molecular dynamics (MD) simulations

To investigate the time dependent protein-peptide interactions and conformational dynamics of the studied complex systems, in the current study, we carried out classical MD simulations for H3(1-12)K9me3 and H3(1-12)K9me2 peptides complexed with UHRF1 TTD-PHD and its Asp145Glu and Asp145Ala mutants, using Desmond MD simulation program [Desmond Molecular Dynamics System,



Version 2.2, D.E. Shaw Research, New York, NY, 2009]. All systems were solvated in an orthorhombic box ( $a=b=c=10\text{\AA}$  and  $\alpha=\beta=\gamma=90^\circ$ ) with explicit SPC (Single Point Charge) water model. The interactions were calculated with the OPLS3 force field. The complex was neutralized in buffer system with 0.15M NaCl. The particle-mesh Ewald method [89] was used to calculate the long-range electrostatic interactions. A cut-off radius of  $9.0\text{\AA}$  was applied for short-range van der Waals and Coulomb interactions. The systems were simulated under an isothermal-isobaric ensemble (NPT) with the temperature of 300K and the pressure of 1 bar. Nose-Hoover thermostat [90] and Martyna-Tobias-Klein [91] methods were implemented to maintain the temperature and the pressure of the systems, respectively. An integral time step of 2 fs was used for the overall simulations. The systems were minimized and equilibrated with the default protocols of the Desmond and the systems were relaxed. Finally, a 10 ns non-constrained MD simulation was performed for each system and the coordinates were saved for every time step.

#### 3.2.4 Protein-peptide interaction analysis

Protein-peptide interactions were screened throughout the MD simulations. To identify the hydrogen bond, maximum distance of  $3.5\text{\AA}$  and minimum donor angle of  $120.0^\circ$  were considered. In the case of a face to face  $\pi$ - $\pi$  stacking interaction, angle and distance between the rings less than  $30^\circ$  and less than  $4.4\text{\AA}$  were set

#### 3.2.5 Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) calculations

Protein-peptide free energy of binding ( $\Delta G_{\text{bind}}$ ) after 10ns of the simulation were estimated using MM-GBSA method (Prime module of the Schrodinger's molecular modeling package) [Schrodinger Release 2017-1: Prime, version 3.8, Schrodinger, LLC, New York, NY, 2014] [91] for all the complexes.

### 3.2.6 Sub-Cloning of UHRF1 TTD-PHD to generate hexahistidine-SUMO tagged construct

The cDNA encoding full-length human UHRF1 was obtained from Open biosystems. Residues 140-380, that corresponds to the UHRF1 TTD-PHD, were sub-cloned. Amplification was carried out using forward and reverse primers (Appendix I) containing NdeI and BamHI restriction sites, respectively. The amplified PCR product was confirmed by agarose gel electrophoresis, followed by DNA gel extraction and subsequent digestion with NdeI and BamHI. The purified digested product was ligated with NdeI-BamHI digested pET28-N-His-SUMO vector using T4 DNA ligase. Schematic represents the workflow of cloning strategy (Fig. 3.8). Clones obtained were screened by using colony PCR and confirmed by sequencing.

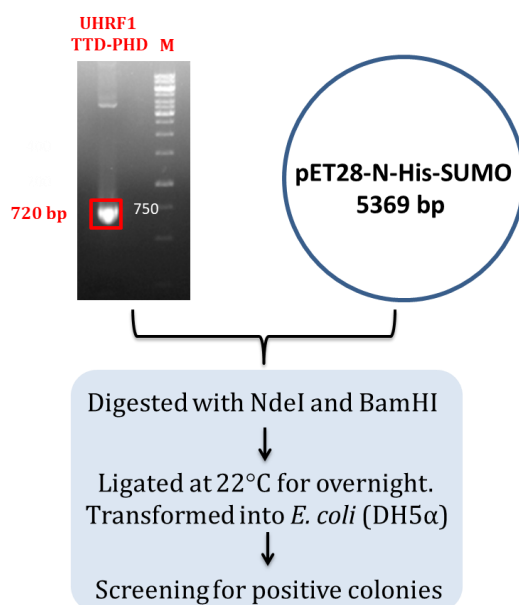


Figure 3.8: Flowchart of UHRF1 TTD-PHD sub-cloning.

### 3.2.7 Expression and Purification of UHRF1 TTD-PHD

Expression and purification of UHRF1 TTD-PHD carried out as discussed in section 2.2.2. The purification gel picture and the chromatogram of UHRF1 TTD-PHD depicted in Fig. 3.9

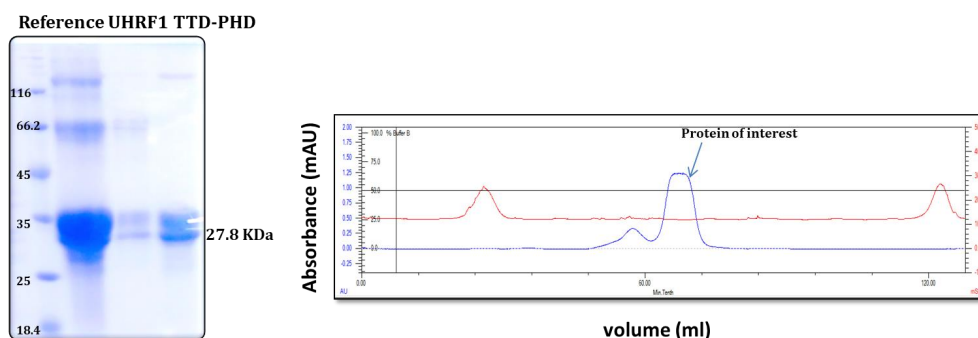


Figure 3.9: SDS-PAGE image after gel filtrations purification and the chromatogram showing elution profile of UHRF1 TTD-PHD.

### 3.2.8 Generation of UHRF1 TTD-PHD Asp145Glu mutant

Mutant was generated by using QuikChange II XL Site Directed Mutagenesis Kit (Stratagene) on a plasmid carrying the cDNA of the UHRF1 TTD-PHD and mutation was confirmed by sequencing. Mutated protein was expressed and purified as described for UHRF1 TTD-PHD.

### 3.2.9 ITC measurements

The equilibrium dissociation constant ( $K_D$ ), molar ratio ( $N$ ) and thermodynamic parameters of the UHRF1 TTD-PHD and UHRF1 TTD-PHD Asp145Glu bound to H3(1-12)K9me3 and H3(1-12)K9me2 peptides were determined using a VP-ITC calorimeter (MicroCal, LLC) at 25°C. The proteins were dialyzed overnight against a buffer containing 40 mM Tris-HCl, pH 7.5, 50 mM NaCl, and 2 mM  $\beta$ -mercaptoethanol at 4°C. Lyophilized peptides were dissolved in buffer used for protein dialysis. The concentrations of protein and peptides used in this study were 100-250  $\mu$ M and 1-2.5 mM, respectively. The reaction cell was filled with 210  $\mu$ L of protein and the reference cell was filled with deionized water. The peptide was sequentially added in 2.3  $\mu$ L (for a total of 15-16 injections) aliquots at 3 min intervals. The data were processed using MicroCal Origin software. The titration data were deconvoluted based on a binding model containing “one set of sites” using a nonlinear least-squares algorithm. The binding enthalpy change ( $\Delta H$ ), association

constant ( $K_A$ ), and binding stoichiometry ( $N$ ) were permitted to vary during the least-squares minimization process and taken as the best-fit values for UHRF1 TTD-PHD bound to the histone peptides. In the cases where the  $C$ -value (which is the product of the receptor concentration and the binding constant,  $K_A$ ) is low, ' $N$ ' was fixed to 1.0, and ' $K_A$ ' and ' $\Delta H$ ' were permitted to float. The reported values are the mean values from three titrations. The equilibrium molar dissociation constant ( $K_D$ ) of the mutant bound to H3K9me2 and H3K9me3 peptides was determined using a VP-ITC calorimeter (MicroCal, LLC).

### 3.2.10 Fluorescence Polarization (FP) measurements

To confirm the ITC binding studies on UHRF1 TTD-PHD with peptides, fluorescence polarization measurements were performed at 20°C using C-terminal fluorescein labeled peptides. The labeled peptide (1  $\mu$ M) was added to an increasing concentration of the protein. The polarization (in millipolarization [mP] units) of 20  $\mu$ L of reaction mixture was measured after 45 minutes of incubation at room temperature. Polarization values were referenced against a blank sample buffer (20 mM Tris-HCl pH 7.5, 50 mM NaCl and 1 mM DTT) and a reference containing 1  $\mu$ M of peptide. The polarization data were analyzed using GraphPad Prism 5.0 software (GraphPad, San Diego, CA, USA) to calculate a binding dissociation constant ( $K_D$ ) by fitting the experimental data in a non-linear regression equation using a one-site specific binding model accounting for ligand depletion. Experiments were performed in triplicate.

## 3.3 Results

### 3.3.1 MD simulation studies on binding of H3R2K9me2 and H3R2K9me3 peptides to TTD-PHD UHRF1 cassette

Insignificant fluctuation of RMSDs of protein and ligand, and RMSFs of ligand has indicated that not only TTD-PHD but also methyllysine peptides are stable during simulation time (Fig. 3.10 and 3.11). Overall, H3R2K9me2 and H3R2K9me3

peptides have similar network of interactions throughout the simulation (Fig. 3.12), as well as peptides exhibit minimum internal conformation (Fig. 3.12C, D).

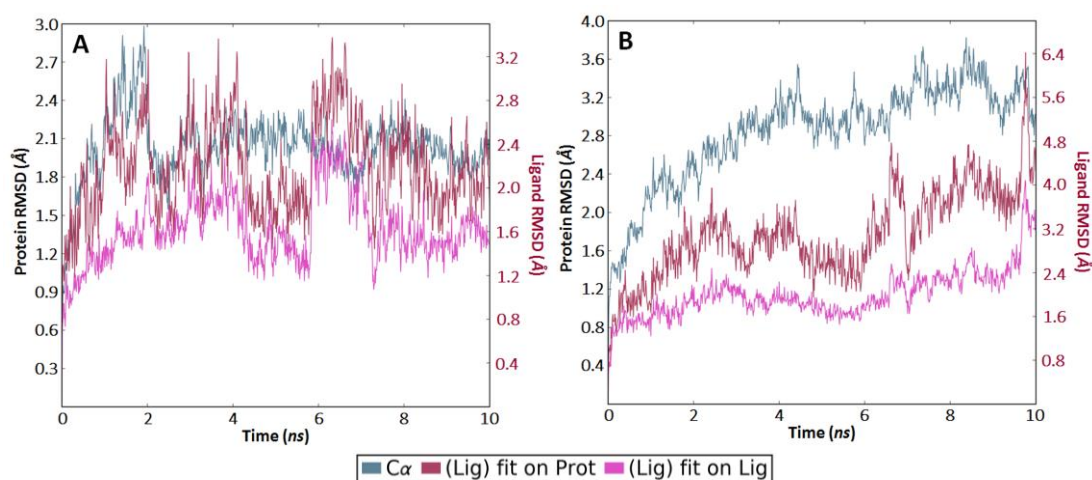


Figure 3.10: Protein-Ligand root-mean square deviation (RMSD) plotted against the stimulation time. (A) UHRF1 TTD-PHD-H3(1-12)K9me2 complex (B) UHRF1 TTD-PHD-H3(1-12)K9me3 complex RMSD of C $\alpha$  of the UHRF1 TTD-PHD. ‘Lig fit on Prot’ shows the RMSD of the respective peptide when the protein-peptide complex is aligned on the protein backbone (at t=0ns). RMSD of the ligand is measured; ‘Lig fit on Lig’ shows the RMSD of the peptide with respect to its conformation at t=0ns.

H3R2 mark is recognized by the negatively charged surface groove of PHD finger. Guanidinium group of Arg2 is recognized by the Asp337, Asp334 and Cys333 through the network of salt bridge and hydrogen bond interactions in TTD-PHD-H3R2K9me3 complex structure [46,49,60,72]. Above network of interactions are similar throughout the simulation in H3R2K9me2 and H3R2K9me3 peptides binding (Fig. 3.12C, D). Phe152, Tyr188 and Tyr191 residues from TTD domain of TTD-PHD form aromatic cage binding pocket that binds to K9me3 mark through hydrophobic and CH- $\pi$  interactions (Fig. 3.12) [46,72]. All these aromatic residues interact with K9me3 and K9me2 mark of the peptides mainly through hydrophobic interactions throughout the simulation time (Fig. 3.12A, B). Asp145 has hydrogen bond interaction with dimethyl ammonium group of H3R2K9me2 peptide (Fig.

3.12A, C), surprisingly, which is, absent in H3R2K9me3 peptide (Fig. 3.12B, D). Another notable difference is TTD-PHD has significantly lower free energy of binding for H3R2K9me2 compared to H3R2K9m3 (Table 3.1).

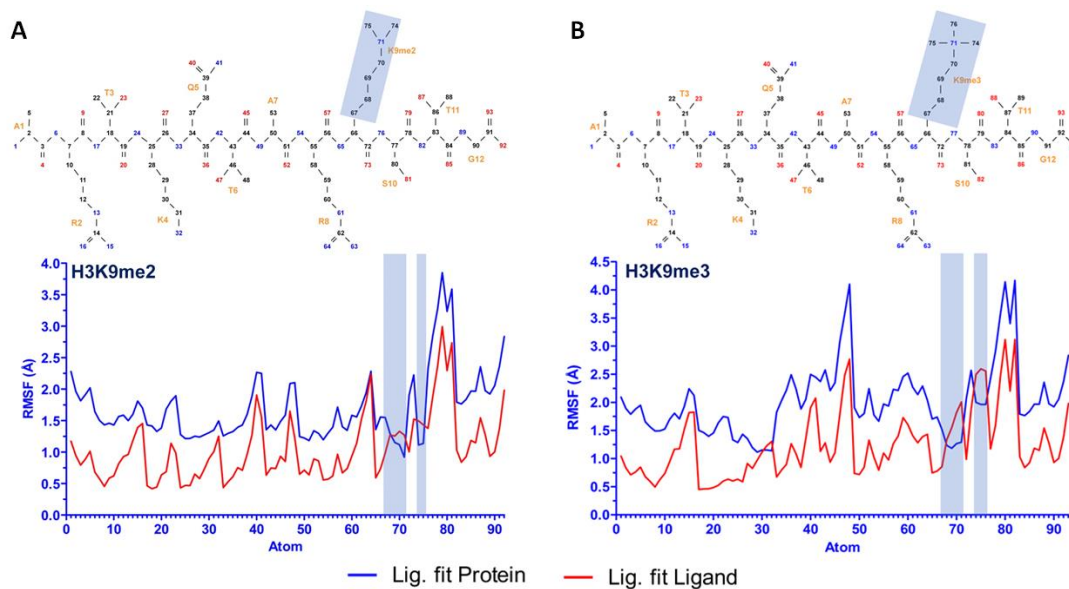


Figure 3.11: Ligand root-mean square fluctuation (RMSF) - peptide's atoms fluctuation throughout the simulation. RMSF of atoms of (A) H3K9me2 (B) H3K9me3 during 10ns MD simulation. The figure shows the ligand's fluctuations broken down by atom, corresponding to the 2D structure in the top panel of every curve. K9 methylations are highlighted in light blue in all figures.

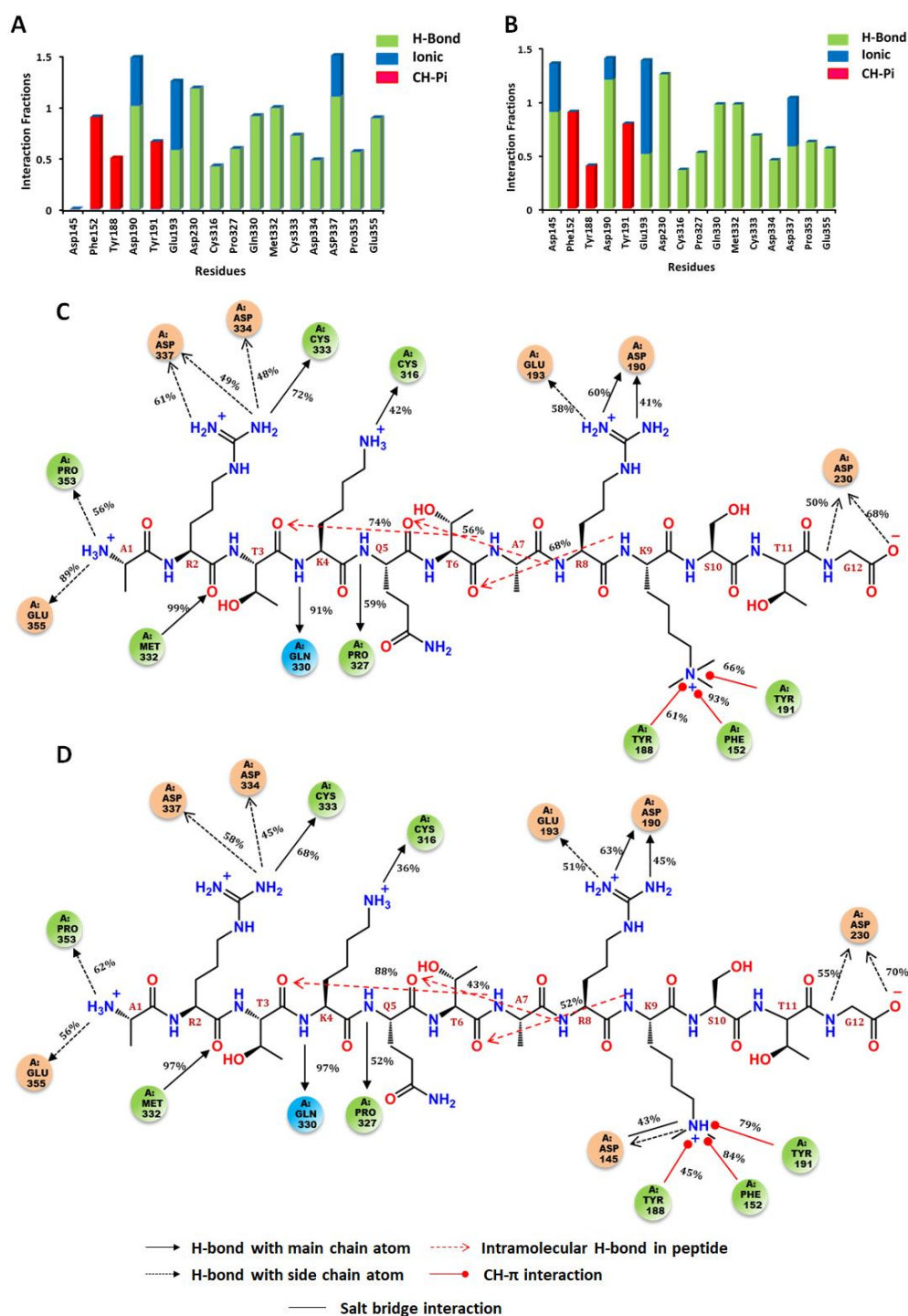


Figure 3.12: Residual interaction profile of UHRF1 TTD-PHD with di- and trimethylated lysine 9 of H3(1-12) peptide. Histograms display interaction fraction of protein residues during simulation time with (A) H3R2K9me3 and (B) H3R2K9me2. 2D-interaction diagrams of the protein residues contact with the peptides (C) H3R2K9me3 and (D) H3R2K9me2.

### 3.3.2 UHRF1 TTD-PHD cassette preferentially binds H3R2K9me2

Our ITC binding study on UHRF1 TTD-PHD with H3(1-12)K9me2 and H3(1-12)K9me3 peptides reveal that UHRF1 TTD-PHD cassette exhibits 6-folds higher affinity for dimethyllysine mark ( $K_D = 0.15 \mu\text{M}$ ) (Fig. 3.13A) than trimethyllysine mark (Fig. 3.13B). To confirm this, we also conducted FP studies on the UHRF1 TTD-PHD cassette with fluorescent labelled H3(1-12)K9me2 and H3(1-12)K9me3 peptides. FP studies also revealed that the UHRF1 TTD-PHD has higher affinity for H3(1-12)K9me2 ( $K_D$  of  $5.2 \mu\text{M}$ ) compared to H3(1-12)K9me3 ( $K_D$  of  $23.1 \mu\text{M}$ ) (Fig. 3.14). The variation between the  $K_D$  values of ITC and FP studies may be due to difference in the temperature constants used in the experiment.

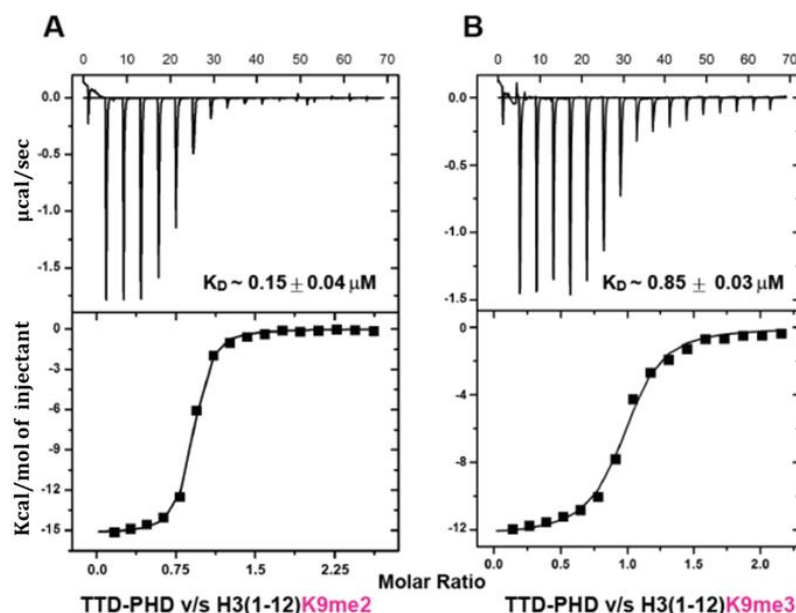


Figure 3.13: ITC binding studies of H3R2K9me2 and H3R2K9me3 peptides to UHRF1 TTD-PHD cassette. Enthalpy plots for the binding of the UHRF1 TTD-PHD to (A) H3R2K9me2 and (B) H3R2K9me3 peptides. The inset lists the measured molar dissociation constant.



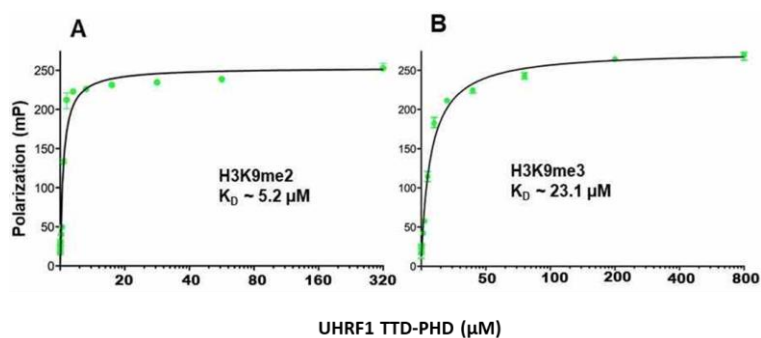


Figure 3.14: Fluorescence Polarization (FP) Binding of the UHRF1 TTD-PHD to fluorescein labelled peptides (A) H3(1-12)K9me2 (B) H3(1-12)K9me3. The inset lists the measured molar dissociation constant.

### 3.3.3 UHRF1 TTD-PHD Asp145Glu mutant preferentially binds di-methyllysine 9 over tri-methyllysine 9

Our MD simulation studies show that Asp145Glu and Asp145Ala mutants of UHRF1 TTD-PHD shows that the presence of negatively charged residue in the binding pocket may be required for preferential recognition of H3K9me2 mark over H3K9me3. ITC binding studies on UHRF1 TTD-PHD Asp145Glu mutant with H3(1-12)K9me2 and H3(1-12)K9me3 peptides reveal that the Asp145Glu mutant of UHRF1 TTD-PHD also exhibits higher affinity for H3(1-12)K9me2 ( $K_D = 0.51 \mu\text{M}$ ) compared to H3(1-12)K9me3 ( $K_D = 1.21 \mu\text{M}$ ) (Fig. 3.15).

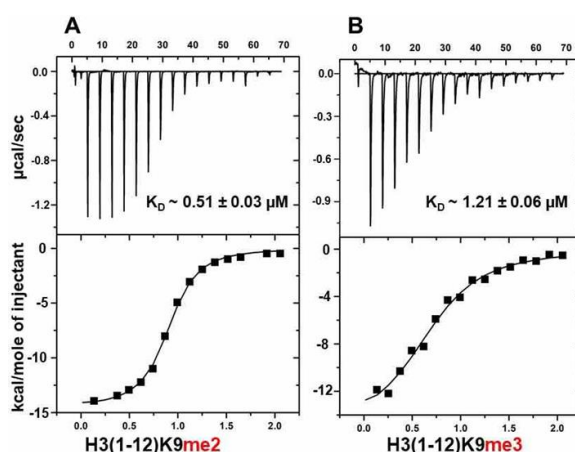


Figure 3.15: ITC binding studies of H3R2K9me2 and H3R2K9me3 peptides to UHRF1 TTD-PHD Asp145Glu mutant. Enthalpy plots for the binding of the

UHRF1 TTD-PHD to (A) H3R2K9me2 and (B) H3R2K9me3 peptides. The inset lists the measured molar dissociation constant.

Table 3.1: MM/GBSA calculated  $\Delta G_{\text{bind}}$  for UHRF1 TTD-PHD bound to H3(1-12) peptide with different lysine methylation statuses.

H3 Peptide vs TTD-PHD	$\Delta G$ of binding (Kcal/mol)
H3R2K9me3	-111.98
H3R2K9me2	-130.42
H3 Peptide vs TTD-PHD Asp145Glu	
H3R2K9me3	-120.57
H3R2K9me2	-137.68
H3 Peptide vs TTD-PHD Asp145Ala	
H3R2K9me3	-111.59
H3R2K9me2	-109.222

### 3.3.4 Effect of UHRF1 PHD domain on histone H3 lysine methyl mark recognition by the UHRF1 TTD

To unravel the effect of PHD domain on recognition of H3K9me2 by TTD of UHRF1, we have carried out ITC binding studies on UHRF1 TTD and UHRF1 TTD-PHD with H3(1-12)K9me2 and H3(5-12)K9me2 peptides, respectively. UHRF1 TTD-PHD binds H3(5-12)K9me2 peptide with a  $K_D$  of 2.5  $\mu\text{M}$  (Fig. 3.16B), which is similar to H3(1-12)K9me2 peptide recognition by the UHRF1 TTD (Fig. 3.16A) indicating that PHD may not have any role in recognition of H3K9me2 by UHRF1 TTD.

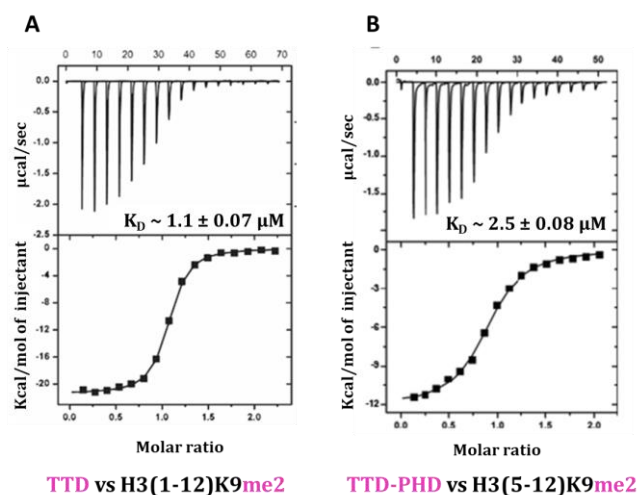


Figure 3.16: ITC binding studies of TTD and TTD-PHD of UHRF1 to methylated lysine H3. Enthalpy plots for the binding of the (A) UHRF1 TTD to H3(1-12)K9me2 and (B) TTD-PHD to H3(5-12)K9me2.

### 3.4 Discussion

#### 3.4.1 Effect of negatively charged residue in the binding pocket on lower methyllysine status recognition

Our MD simulation studies show that there is an additional hydrogen bond and salt bridge interaction between K9me2 of H3 and Asp145 of TTD domain throughout the simulation time, which is lost in trimethylated H3K9. This shows that Asp145 in the binding pocket of UHRF1 TTD domain may play a key role in preferential recognition of H3K9me2 over H3K9me3 (Fig. 3.12, Table 3.1). To confirm the role of this negatively charged residue, Asp145, we mutated it to a negatively charged residue (Glu) and a hydrophobic residue (Ala). In the case, where Asp145 was mutated to Glu145, MD simulation and  $\Delta G_{\text{bind}}$  (Table 3.1) show the persistence of preference for H3R2K9me2 mark. Our ITC binding studies have also confirmed that the Asp145Glu mutant of UHRF1 TTD-PHD has similar binding affinity for H3K9me2 with respect to wild type UHRF1 TTD-PHD (Fig 3.13 and 3.14). The preferential recognition of H3R2K9me2 over H3R2K9me3 peptide by the UHRF1 TTD-PHD Asp145Glu mutant, similar to wild type UHRF1 TTD-PHD highlights

the role of negatively charged residue for preferential recognition of lower methylation status (Fig. 3.17A). On the other hand, UHRF1 TTD-PHD Asp145Ala mutant completely lost this interaction in MD simulation study (Fig. 3.17B). These observations are supported by the previous studies on lower methyllysine status-specific readout by MBT repeats of L3MBTL1 and 53BP1 TTD. Those studies reported that the presence of negatively charged residue in the aromatic cage binding pocket is linked to lower methyllysine status recognition of H1.4K26, H3K4, H3K9, H3K27, H3K36 or H4K20 peptides [60,92,93]. Substitution, of one of the aromatic cage residues in the binding pocket to the negatively charged residue (Asp/Glu), was the strategy used to engineer the PHD finger of BPTF and the Chromodomain of HP1 $\alpha$  to preferentially bind dimethyllysine of histone H3 [92,94]. In contrast to the current study, the previous study has reported that UHRF1 TTD preferentially binds H3R2K9me3 over H3R2K9me2 [60]. Use of N-terminal fluorophore labelled H3 peptide for binding study that may affect the recognition of H3 peptide by the TTD, might be the reason for an anomaly in the previous study [60]. FP binding studies using the C-terminal fluorescein labelled, H3(1-12)K9me2 and H3(1-12)K9me3, peptides validate the preferential recognition of H3K9me2 methylation status by the UHRF1 TTD (Fig 3.14).

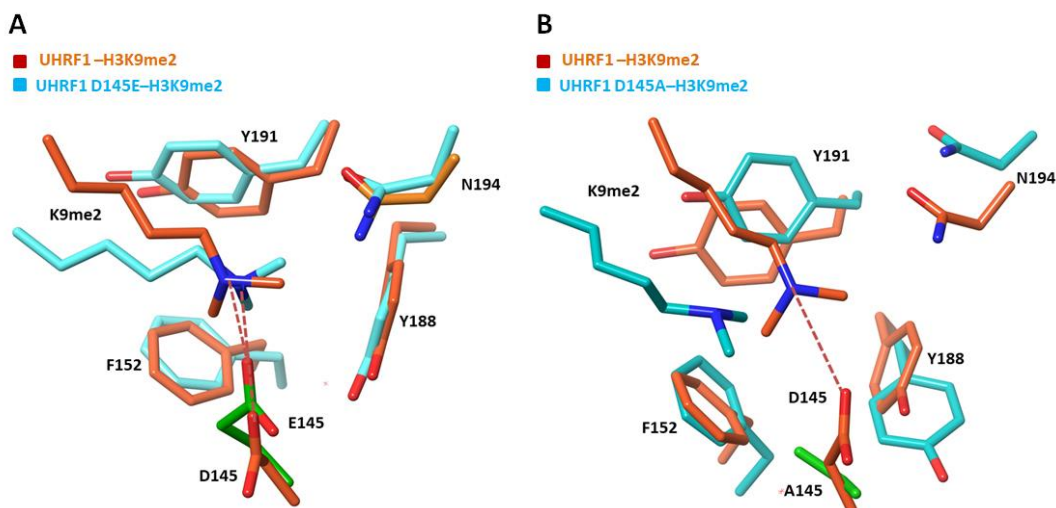


Figure 3.17: Comparison of K9me2 recognition in the H3K9me binding pocket of UHRF1 TTD at the end of 10ns simulation of UHRF1 TTD-PHD, UHRF1 TTD-PHD Asp145Glu and UHRF1 TTD-PHD Asp145Ala complexed with H3(1-12)K9me2 peptide. Binding pocket of UHRF1 TTD (reddish-brown stick) bound to K9me2 (reddish-brown stick) is superimposed on UHRF1 TTD Asp145Glu (A) and UHRF1 TTD Asp145Ala bound K9me2 (cyan stick) (B). Asp145Glu and Asp145Ala mutation is represented in green stick.

#### 3.4.2 Comparison of dynamics of the dimethyllysine 9 and the trimethyllysine 9 recognition by the UHRF1 TTD

To understand the dynamics of methyllysine statuses recognition by the UHRF1 TTD domain, we compared the MD simulation results of UHRF1 TTD-PHD-H3R2K9me2 and UHRF1 TTD-PHD-H3R2K9me3 complexes (Fig. 3.18). Unlike 53BP1 TTD, UHRF1 TTD can accommodate both dimethyllysine and trimethyllysine statuses but prefers former over later (Fig. 3.18). Conformations of dimethyl- and tri-methyllysines, in the binding pocket, are such that only dimethyllysine has hydrogen bond interaction with the Asp145. Among binding pocket residues, the carboxyl plane of Asp145 is oriented towards the methyllysine in UHRF1 TTD-PHD-H3R2K9me2 complex, in contrast, it is oriented away from the methyllysine in the UHRF1 TTD-PHD-H3R2K9me3 complex. Above conformational changes are necessary for preferential recognition of H3R2K9me2 over H3R2K9me3, and to avoid the exposure of one of the methyl groups to the negatively charged environment that might weaken the recognition of trimethyllysine status by the UHRF1 TTD (Fig. 3.18). Preferential recognition of H3R2K9me2 is supported by global loss of H3R2K9me2 in the genome of embryonic stem cells (ESCs) that leads to loss of UHRF1 recruitment to chromatin.

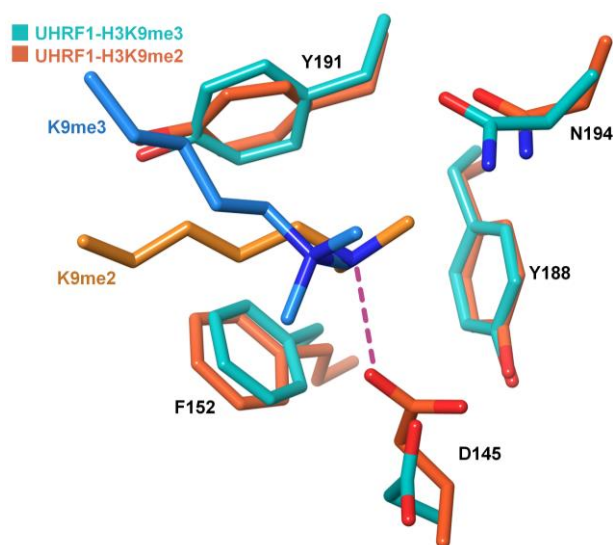


Figure 3.18: Comparison of K9me3 and K9me2 recognition in the H3K9me binding pocket of UHRF1 TTD at the end of 10ns simulation of UHRF1 TTD-PHD-H3K9me2 and UHRF1 TTD-PHD-H3K9me3 complexes. Binding pocket of UHRF1 TTD (reddish-brown stick) bound to K9me2 (brown stick) is superimposed on UHRF1 TTD (cyan stick) bound K9me3 (blue stick).

### 3.4.3 UHRF1 PHD doesn't contribute in recognition of H3K9me2 by UHRF1 TTD domain

There has been a recent interest and appreciation of multivalent or combinatorial readout of two or more histone marks by a corresponding number of 'reader' domains, highlighting the advantages of combinatorial recognition at the histone and/or nucleosome level [76]. The linker region between TTD and PHD in UHRF1 acts as a ruler for combinatorial recognition of H3K9me2 and H3R2 marks by dual domains [72]. Therefore, recognition of N-terminal amino group, H3A1 and H3R2 by PHD can prevent the non-specific histone H3 methyllysine recognition by the TTD, but PHD finger has negligible contribution for binding affinity to recognize H3K9me2 mark by the TTD (Fig. 3.16). Our hypothesis is supported by previous study showing that UHRF1 TTD-PHD exhibit negligible binding affinity to trimethylated histone H3 lysine 27 or 36 peptides [46].

However, combinatorial recognition of UHRF1 TTD-PHD to H3R2K9me2 peptide is marginally better than recognition of individual, H3K9me2 and H3R2 marks by TTD and PHD domains, respectively. It appears that the relative linker-dependent separation and orientation of TTD and PHD finger within the cassette may not conducive to, simultaneously, target H3R2 and H3K9me3 with the expected higher binding affinity compared to individual mark recognition. In addition, ordered histone H3 in the binding pockets of UHRF1 TTD-PHD could also restrict the recognition of dual marks by the reader domain.

## Chapter 4

Molecular Dynamics (MD) simulation and binding studies to explore the effect of H3K4 methylation on H3R2K9me recognition by the TTD-PHD of UHRF1



## 4.1 Introduction

It is established that some residues on histones undergo several PTMs, like methylation (lysine and arginine), acetylation (lysine) and phosphorylation (serine and threonine) (Fig. 4.1) [95,96]. Such modification, in most of the cases, affect the charge distribution on histones where it is involved in binding to DNA, that lead to modulation of chromatin structure/conformation, which in turn regulates the DNA associated processes. A synergistic combination of two or more histone marks works together to ensure that no gene is inadvertently turned on or off. There are some protein motifs or reader domains evolved to bind combinations of histones and initiate crosstalk [96,97].

A single PTM on histone is able to modulate or regulate protein function by creating binding sites for protein or disturbing the protein-protein interaction. However, many histone proteins undergo modifications at multiple sites and act combinatorically to enrich the information. For instance, N-terminal tail of histone H3 can be simultaneously methylated at K4 and acetylated at different positions include K9, K14, K18 and K27. These combinatorial modifications affect the transcription. Among all modifications, lysine side chain is the main target of different PTMs. These lysine PTMs can be involved in the cross regulation [98].

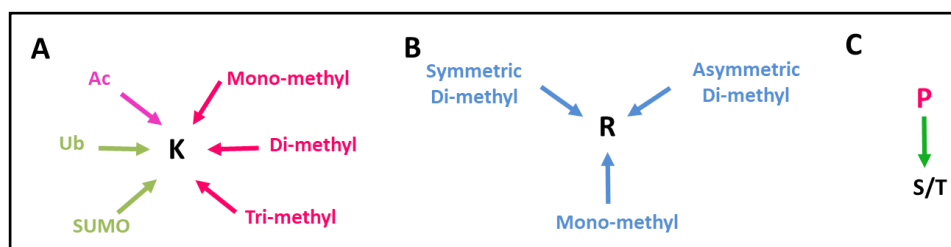


Figure 4.1: Potential ‘choices’ of the modification states of different histone residues.

Acetylation (Ac) of lysine (K), mono-, di- and tri-methylation of lysine, mono-ubiquitination (ub) of lysine and SUMOylation (SUMO) of lysine; Mono- or di-methylation (symmetric or asymmetric) of arginine (R); phosphorylation (P) of serine or threonine (S/T).

## 4.2 Histone modifications crosstalk

The crosstalk between different modifications provides an extra level of complexity to fine tune the overall control. There are three mechanisms that will explain crosstalk of modifications. One or more modification occurs in the same residue that is targeting different modification pathway. For example, lysine can be acetylated, methylated or ubiquitylated. Crosstalk can occur on *cis* (same histone) and *trans* (between histone or nucleosome) positions. One modification may disturb the protein or reader module binding of other modification [99,100].

These crosstalks can also be regulated positively or negatively. In positive crosstalk, the addition or removal of one PTM can act as a signal for the second modification/s recognition/readout or removal. In negative crosstalk, the competition between histone PTM on a single residue, the recognition site of one PTM masked by another PTM (Fig. 4.2) [98].

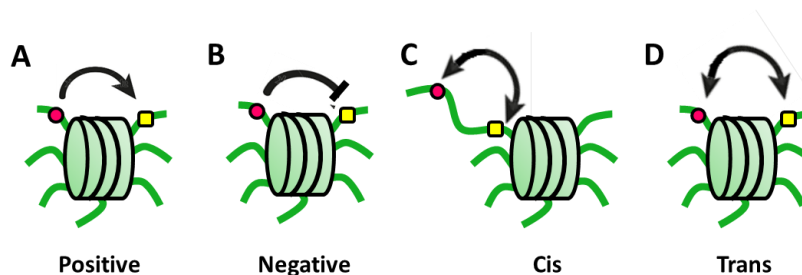


Figure 4.2: Modes of histone modification crosstalk. (A) One PTM positively regulates another PTM. (B) One PTM negatively regulates another PTM. (C) Regulation of PTM is on the same histone tail. (D) Regulation of PTM is on different histone tail of same or another histone.

## 4.3 Reader domain mediated histone crosstalk

A major function of PTMs is to create binding sites for specific modular binding domains providing a mechanism for inducing protein-protein interactions and propagating signals. Proteins commonly contain more than one protein interaction domain, and in some cases more than one reader domains can recognize a PTM

[98]. In fact, many readers associate with a substantial stretch of the histone tail, allowing for the sensing of multiple marks. In many cases, the biological outcome depends on a combinatorial readout of multiple PTMs by spatially linked readers that lead to crosstalk between PTMs. The readout and crosstalk can be cooperative, antagonistic, or independent in nature [101].

#### 4.4 Phospho/methyl switch

Methylation and phosphorylation are one of the best examples of cross regulation in *cis*. These two sites are situated nearby or adjacent to each other. For example, most of the methylation sites are situated nearby the phosphorylated serine or tyrosine such as H3T3phK4me, H3K9meS10ph, H3K27meS28ph and H3S31phK36me [102]. The existence of lysine nearby the phosphorylation site is called phospho/methyl switch and these switches are involved in the effector protein regulation [101].

For instance, the binding of HP1 chromodomain to H3K9me3 is inhibited by the phosphorylation of S10 and this event is involved in the negative regulation of HP1 localization at chromatin [74]. The CHD1 double chromodomain (DCD) binds tightly towards H3K4me3 mark when the T3 is unphosphorylated. It exhibits 25-fold lower affinity to H3T3phK4me3 peptide, because of weakening inter-molecular interactions upon phosphorylation. But the phosphorylation on H3S10 residue doesn't show any effect on the interaction of CHD1 DCD with H3K4me3 [101]. The TTD domain of KDM4A recognizes H3K4me3. This interaction is abolished by phosphorylation at H3T3, and significant reduction in binding affinity was observed when the H3T6 was phosphorylated [103]. The PHD2 finger domain of CHD4 protein preferentially binds H3K9me3 compared to unmodified histone H3 tail. T3, T6 or S10 phosphorylation affects the interaction between PHD2 and H3K9me3 [101]. MLL1 methylates H3K4. But methylation of H3R2 prevents methylation of H3K4 by MLL1 [104].

#### 4.5 Roles of histone modification crosstalk

Many chromatin marks, such as H3S10ph and H3T3ph, do not act independently but influence each other's function. Histone PTMs crosstalk also has effect on catalytic activity and substrate specificity of eraser and writer enzymes. In some cases, the existing PTM blocking the modification site of another modification. For an example, the H3K9ac blocks the methylation of H3K9. In contrast, activation of some enzymes depends on the interaction of chromatin mark or existence of PTM. For instance, ubiquitinated histone H2B activates the enzyme Dot1 KMT for intra-nucleosomal lysine methylation. Thus, histone modification crosstalk will provide knowledge to understand synergistic and antagonistic effects of multiple modifications [105].

#### 4.6 Conformation of N-terminal H3 peptide in the TTD-PHD of UHRF1

The isolated PHD recognition of H3 is well conserved in UHRF1 TTD-PHD-H3R2K9me3 complex. Residues from R2 to K4 of histone H3 involved in hydrogen bond interaction with PHD finger. The H3R2 side chain fits into the negatively charged groove in the PHD finger, in which the guanidinium nitrogen atoms are recognized by residues C333, D334, and D337. H3K9me3 and H3S10 form contacts only with the tudor1 domain region. The aliphatic side chain of K9me3 fits into a surface groove, and the trimethyl ammonium group is trapped in an "aromatic cage" formed by residues F152, Y188, and Y191 of the first tudor domain, as commonly observed for methylated lysine recognition in similar reader domain (Fig. 4.3B) [72].

Interestingly, crystal structures of the UHRF1 TTD-PHD bound to H3(1-11)K9me3 (PDB: 4GY5) revealed that lysine 4 through alanine 7 of H3 adopt a single  $\alpha$ -helix turn (Fig. 4.3A), which is stabilized by intramolecular H-bonds [46,80]. The extended conformation of histone H3 in complex with the PHD led to the assumption that TTD-PHD multivalent engagement induces a conformational change in the histone H3. It is also seen that lysine and arginine residues are rich in

N-terminus of H3 tail and they lack secondary structure in peptides and in free histone. These findings have led that N-terminus of histone H3 is disordered [80]. UHRF1 PHD recognizes unmodified H3R2, and it exhibit modest reduction in binding affinity when the H3K4 is trimethylated. But trimethylation on H3K9 does not show any impact on binding. On the other hand, UHRF1 PHD shows very weak affinity towards histone H3 ( $K_D > 500 \mu\text{M}$ ) when the T3 is phosphorylated. This indicates that instability of the complex due to steric or electrostatic repulsion of the T3 phosphorylation [49,72].

Thus N-terminus of H3 peptide undergoes single helical conformation upon binding to TTD-PHD. We hypothesize that lysine 4 methylation may affect the conformation of peptide in the binding pocket, that inturn its recognition by the TTD-PHD. Also, the side chain of K4 has hydrogen bond interaction with main chain residue of protein (Fig. 4.3C) which may be lost upon methylation. Therefore, we want to study the effect of lysine 4 dimethylation on combinatorial recognition of H3R2 and H3K9me3 marks by TTD-PHD dual domains of UHRF1 through MD simulation and ITC binding studies.

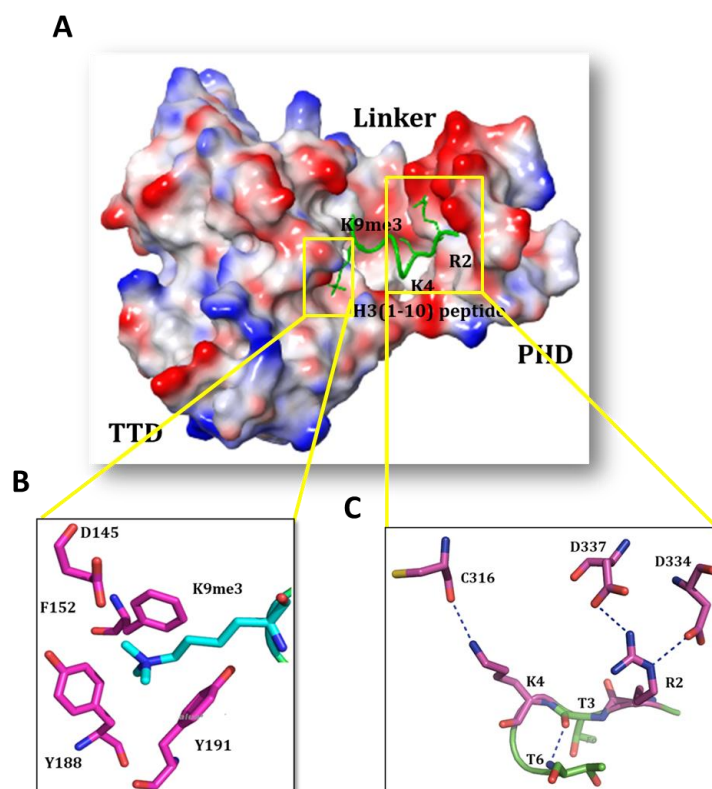


Figure 4.3: Structure of the UHRF1 TTD-PHD bound to H3(1-10)K9me3 peptide (PDB ID: 4GY5). (A) Protein and peptide are represented in electrostatic surface and sticks respectively. (B) Magnified view of K9me3 positioned between the aromatic rings of UHRF1 TTD binding pocket. (C) Zoomed view of H3R2K4 recognition by the UHRF1 PHD.

## 4.7 Material and Methods

### 4.7.1 MD Simulation

#### 4.7.1.1 Preparation of UHRF1 TTD-PHD for simulation

The crystal structure of UHRF1 TTD-PHD with H3(1-10)K9me3 (PDB ID: 4GY5) [46] was prepared for simulation. Using Schrodinger's Maestro Molecular modeling suit, the H3K9me3 was modified to H3K4me2K9me3 and H3K4me2K9me2 by virtually editing the methylation marks. Two amino acids threonine 11 and glycine 12 were also added virtually to the existing H3(1-10) peptide to get H3(1-12)

peptide and the added amino acids were energy minimized. Bond orders were assigned and Hydrogen atoms were added followed by H-bond optimization and restrained minimization using OPLS3 force field. Protonation statuses were determined at physiological pH 7.0 using PROPKA [87,88]. All the water molecules were removed.

#### 4.7.1.2 Molecular dynamics (MD) simulations

MD simulation was done for time dependent investigations of protein-ligand interactions and conformational dynamics of studied complex systems. In the current study, we carried out classical MD simulations for different methylated lysine of H3 using Desmond MD simulations program [Desmond Molecular Dynamics System, Version 2.2, D.E. Shaw Research, New York, NY, 2009]. All systems were solvated in an orthorhombic box ( $a=b=c=10\text{\AA}$  and  $\alpha=\beta=\gamma=90^\circ$ ) with explicit SPC (Single Point Charge) water model. The interactions were calculated with the OPLS3 force field. The complex was neutralized in buffer system with 0.15M NaCl. The particle-mesh Ewald method [89] was used to calculate the long-range electrostatic interactions. A cut-off radius of 9.0  $\text{\AA}$  was applied for short-range van der Waals and Coulomb interactions. The systems were simulated under an isothermal-isobaric ensemble (NPT) with the temperature of 300K and the pressure of 1 bar. A Nose-Hoover thermostat [90] and Martyna-Tobias-Klein [91] methods were implemented to maintain the temperature and the pressure of the systems, respectively. An integral time step of 2 fs was used for the overall simulations. The systems were minimized and equilibrated with the default protocols of the Desmond. Finally, 10 ns non-constrained MD simulation was performed for each system, and the coordinates were saved for every time step.

#### 4.7.1.3 Protein-ligand interaction analysis

Protein-peptide interactions are screened throughout the MD simulations. To identify the hydrogen bond, maximum distance of 3.5  $\text{\AA}$  and minimum donor angle

of  $120.0^\circ$  were considered. In the case of a face to face pi-pi stacking interaction, angle and distance between the rings less than  $30^\circ$  and less than  $4.4 \text{ \AA}$  were set.

#### 4.7.1.4 Molecular Mechanics-Generalized Born Surface Area (MM-GBSA)

##### Calculations

Protein-ligand free energy of binding ( $\Delta G_{\text{bind}}$ ) of the interactions at 0ns and 10ns of the simulation were estimated using MM-GBSA method (Prime module of the Schrodinger's molecular modeling package) [Schrödinger Release 2017-1: Prime, version 3.8, Schrödinger, LLC, New York, NY, 2014;] [91] for UHRF1 TTD-PHD with H3K4me2K9me3 and H3K4me2K9me2 peptides.

#### 4.7.2 Sub-Cloning of UHRF1 TTD-PHD to generate hexahistidine SUMO tagged construct

The cDNA encoding full-length human UHRF1 was obtained from Open biosystems. Residues 140-380, that corresponds to the UHRF1 TTD-PHD, was sub-cloned. Amplification was carried out using forward and reverse primers (Appendix 1) containing NdeI and BamHI restriction sites, respectively. The amplified PCR product was confirmed by agarose gel electrophoresis, followed by DNA gel extraction and subsequent digestion with NdeI and BamHI. The restriction digested product, was purified then, ligated with NdeI-BamHI digested pET28-N-His-SUMO vector, using T4 DNA ligase. Schematic represents the workflow of cloning strategy (Fig. 4.4). Clones obtained were screened by using colony PCR and confirmed by sequencing.



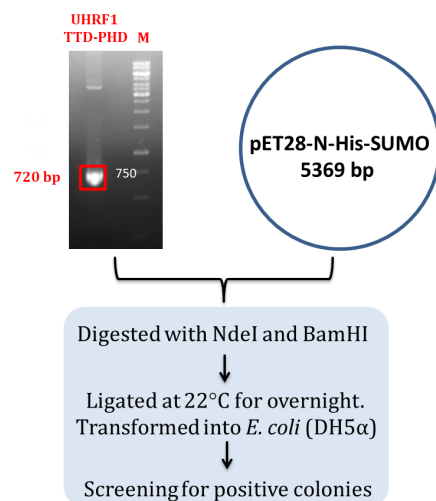


Figure 4.4: Flowchart of UHRF1 TTD-PHD sub-cloning.

#### 4.7.3 Expression and Purification of UHRF1 TTD-PHD

UHRF1 TTD-PHD protein was expressed in *E. coli* Rosetta2(DE3) (Novagen). Cells were grown at 37°C till OD600 reached 0.5-0.6, then the temperature was decreased to 20°C and the culture was induced with 0.4 mM of IPTG. The medium is supplemented with 100  $\mu\text{M}$  of  $\text{ZnCl}_2$ . The hexa-histidine-SUMO fusion protein was purified on a nickel-charged column (HisTrap HP, GE healthcare). After elution with a 750 mM imidazole containing buffer, the fusion protein was cleaved with Ulp1 protease 25  $\text{U ml}^{-1}$  during a 16 hr dialysis step at 4°C in the buffer containing 25 mM Tris-HCl, pH 8.0, 50 mM NaCl, 3 mM DTT. Protein was further purified on an anion exchange column (HiTrap Q; GE healthcare). Protein, eluted from the ion exchange column, was concentrated and then loaded on gel filtration column (HiLoad Superdex 75 26/600) column, equilibrated with a buffer containing 15 mM Tris-HCl, pH 7.5, 50 mM NaCl, 3 mM DTT. Purified UHRF1 TTD-PHD was concentrated to 10  $\text{mg mL}^{-1}$  at 4°C in Vivaspin 20 mL (Vivascience AG) 10,000 cut-off concentrator.

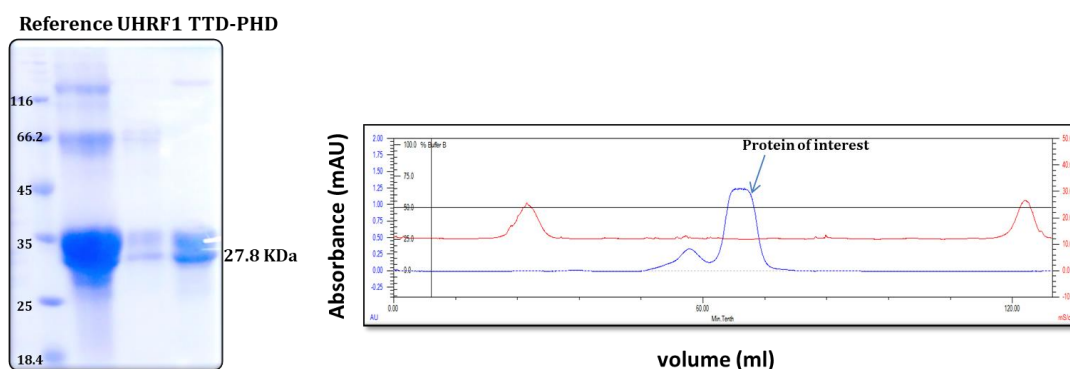


Figure 4.5: SDS PAGE analysis of purified UHRF1 TTD-PHD and gel filtrations chromatogram of UHRF1 TTD-PHD.

#### 4.7.4 ITC measurements

The  $K_D$ , molar ratio ( $N$ ) and thermodynamic parameters of the UHRF1 TTD-PHD bound to H3K4me2K9me2/3 peptide were determined using a VP-ITC calorimeter (MicroCal, LLC) at 25°C. The proteins were dialyzed overnight against a buffer containing 40 mM Tris-HCl pH 7.5, 50 mM NaCl, and 2 mM  $\beta$ -mercaptoethanol at 4°C. Lyophilized peptides were dissolved in buffer used for protein dialysis. The protein and peptides concentrations used for ITC binding study were 100 to 250  $\mu$ M and 1 to 2.5 mM, respectively. The volume of UHRF1 TTD-PHD in the reaction cell was 210  $\mu$ L, and the reference cell was filled with deionized water. The peptide was sequentially added in 2.3  $\mu$ L (for a total of 15-16 injections) aliquots at 3-min intervals. The data were processed using MicroCal Origin software. The titration data were deconvoluted based on a binding model containing “one set of sites” using a nonlinear least-squares algorithm. The binding  $\Delta H$ ,  $K_A$ , and binding stoichiometry ( $N$ ) were permitted to vary during the least-squares minimization process and taken as the best-fit values for UHRF1 TTD-PHD bound to the histone peptides. In the cases where the C-value (which is the product of the receptor concentration and the binding constant,  $K_A$ ) is low, ‘ $N$ ’ was fixed to 1.0, and ‘ $K_A$ ’ and ‘ $\Delta H$ ’ were permitted to float. The reported values are the best values from three titrations.

## 4.8 Results

### 4.8.1 MD simulation studies on binding of H3R2K4me2K9me2 and H3R2K4me2K9me3 peptides to UHRF1 TTD-PHD cassette

To gain insights into the effect of lysine 4 methylation on combinatorial binding of H3K9me2/3 and H3R2 to the UHRF1 TTD-PHD, we have undertaken MD simulation studies on, energy minimized, TTD-PHD-H3R2(1-12)K4me2K9me2 and TTD-PHD-H3R2(1-12)K4me2K9me3 complexes. Insignificant fluctuation of RMSDs of protein and ligand, and RMSFs of ligand has indicated that not only TTD-PHD but also methyllysine peptides are stable throughout the simulation (Fig. 4.6 and 4.7). Overall, H3R2K4me2K9me2 and H3R2K4me2K9me3 peptides have similar network of interactions throughout the simulation (Fig. 4.8), as well as peptides exhibit minimum internal conformation (Fig. 4.8C, D).

Dynamics of TTD-PHD and H3R2K4me2K9me2 peptide showed that the intermolecular contacts are similar to those observed between TTD-PHD and H3R2K9me2 (Fig. 3.10D and 4.8C). Similarly, H3R2K4me2K9me3 and H3R2K9me3 peptides have highly comparable intermolecular contacts with TTD-PHD (Fig. 3.10C and 4.8D). In addition, UHRF1 TTD-PHD has similar  $\Delta G_{\text{bind}}$  for H3R2K4me2K9me2 and H3R2K9me2 peptides (Table 3.1 and 4.1). Similarly, H3R2K4me2K9me3 and H3R2K9me3 peptides exhibit similar  $\Delta G_{\text{bind}}$  for UHRF1 TTD-PHD (Table 3.1 and 4.1). Taken together, these studies suggest that methyllysine 4 has negligible effect on recognition of H3R2K9me2/3 peptides by the UHRF1 TTD-PHD.

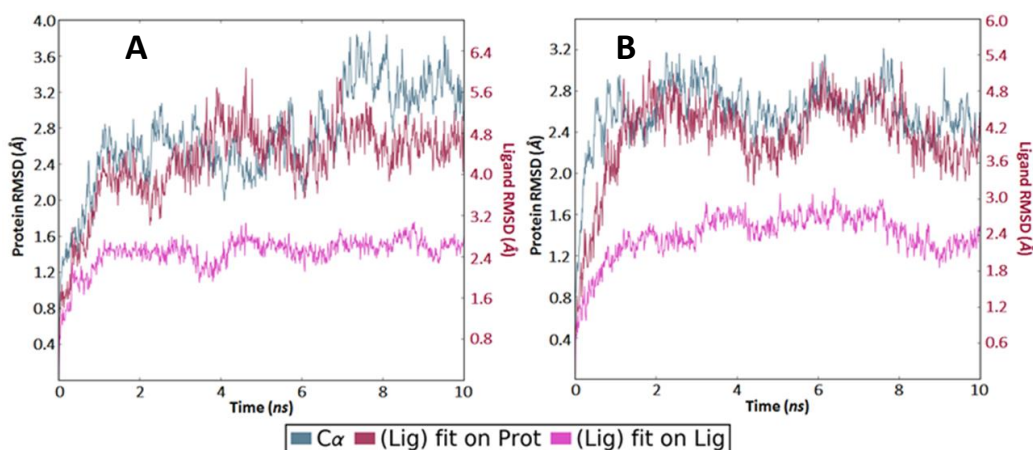


Figure 4.6: Protein-peptide root-mean square deviation (RMSD) plotted against the stimulation time. (A). UHRF1-H3R2(1-12)K4me2K9me2 complex (B). UHRF1-H3R2(1-12)K4me2K9me3 complex. RMSD of  $C_{\alpha}$  of the UHRF1 TTD-PHD backbone. ‘Lig fit on Prot’ shows the RMSD of the respective ligand (peptide) when the protein-ligand complex is first aligned on the protein backbone (at  $t=0$ ns) and then RMSD of the ligand is measured; ‘Lig fit on Lig’ shows the RMSD of the ligand that is aligned on its conformation at  $t=0$ ns.

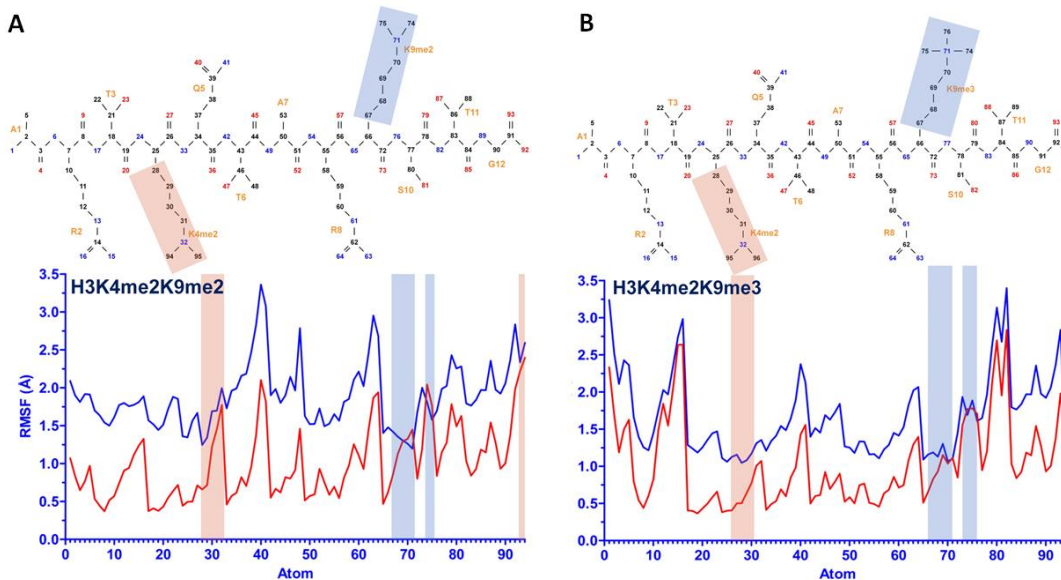


Figure 4.7: Peptide root-mean square fluctuation (RMSF) - peptide's atoms fluctuation throughout the simulation. RMSF of atoms of (A) H3R2K4me2K9me2 and (B) H3R2K4me2K9me3 during 10ns MD simulation. The figure shows the ligand's fluctuations broken down by atom, corresponding to the 2D structure in

the top panel of every curve. K9 methylations are highlighted in light blue in all figures.

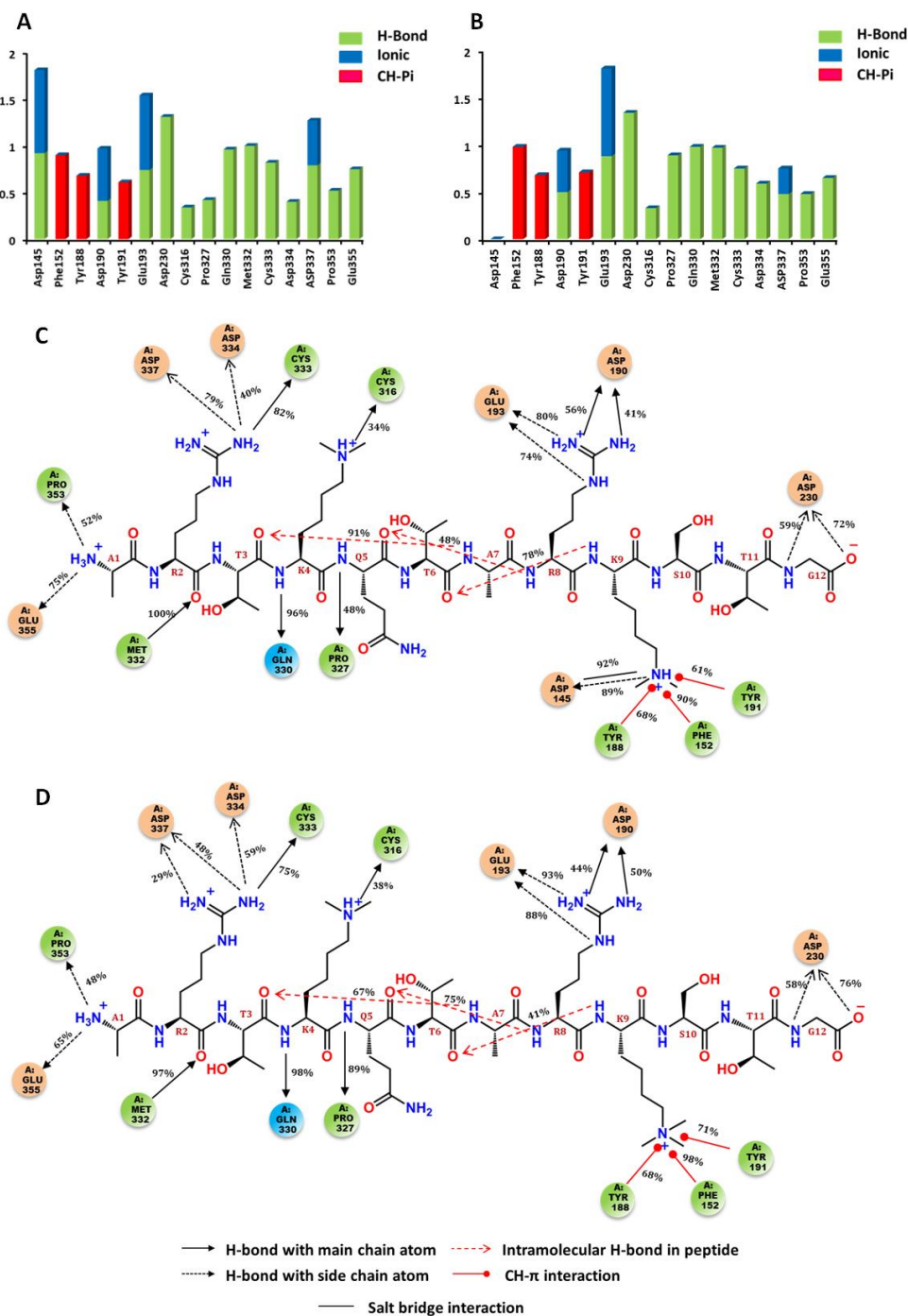


Figure 4.8: Residual interaction profile of UHRF1 TTD-PHD with dual methylated H3(1-12) peptides. Histograms display interaction fraction of protein residues during simulation time with (A) H3R2K4me2K9me2 and (B) H3R2K4me2K9me3. 2D-

interaction diagrams of the protein residues contact with the peptides (C)

H3R2K4me2K9me2 and (D) H3R2K4me2K9me3.

Table 4.1: MM/GBSA calculated  $\Delta G_{\text{bind}}$  for UHRF1 TTD-PHD bound to H3(1-12) peptide with different lysine methylation statuses.

H3 Peptide vs TTD-PHD	$\Delta G$ of binding (Kcal/mol)
H3R2K4me2K9me3	-111.680
H3R2K4me2K9me2	-123.139

4.8.2 Dimethyllysine 4 on H3 has insignificant effect on H3R2K9me2 recognition by the UHRF1 TTD-PHD Cassette

To test MD simulation results, we carried out ITC binding study using the H3R2(1-12)K4me2K9me2 peptide and the UHRF1 TTD-PHD. UHRF1 TTD-PHD cassette binds dual lysine-methylated peptide, H3R2K4me2K9me2, with a  $K_D$  of 0.21  $\mu\text{M}$ , which is similar to that of TTD-PHD binding to H3(1-12)R2K9me2 (Fig. 4.9), thereby supporting the MD simulation results.

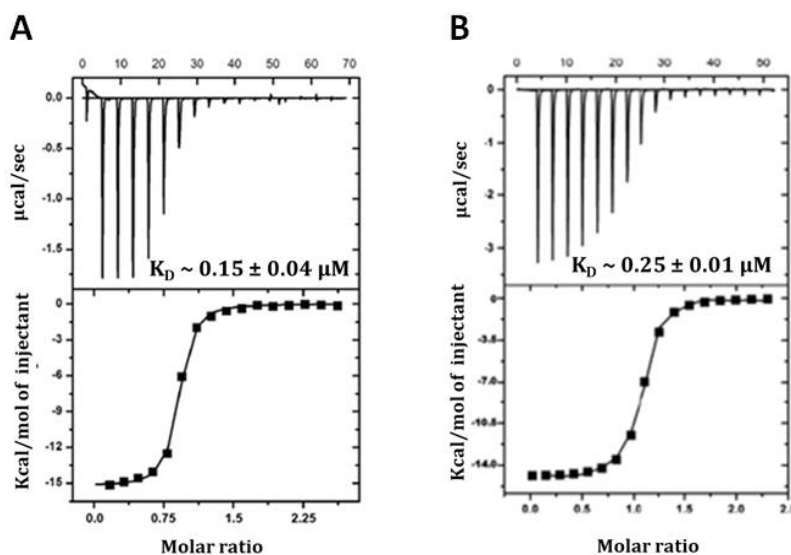


Figure 4.9: Raw ITC data (upper panel) and normalized integration data (lower panel) of enthalpy plots for the binding of the UHRF1 TTD-PHD to (A) H3R2K9me2 and (B) H3R2K4me2K9me3 peptides.

## 4.9 Discussion

Our MD simulation and binding studies show that the interaction between the dimethylated and the unmethylated H3K4 remains conserved. It may be because of the absence of aromatic residues near H3K4, i.e. H3K4me2 may not be recognized by the UHRF1 TTD-PHD. These results may indicate that H3K4me2 neither affects the H3 tail conformation nor induce any conformational change in the PHD finger or the linker region of the protein. Due to this, the orientation or the position of the aromatic cage responsible for recognition of H3K9me3 remains able to recognize H3K9me2 mark. Results also indicate that dimethylation on lysine 4 has insignificant effect on the combinatorial recognition of K9me2 and R2 marks by the TTD and PHD domains, respectively.

## Chapter 5

Characterization of 5mC oxidation derivative recognition by SRA domain of UHRF1.



## 5.1 Introduction

DNA methylation in the form of 5mC provides critical regulatory mechanism beyond the simple genomic sequence as it impacts a variety of biological processes from gene regulation to disease pathogenesis [106,107]. The balance between methylation and demethylation processes is involved in maintenance of dynamic of DNA methylation on the genome. Disruption of this balance could lead to aberrant methylation patterns seen in human diseases such as cancer [107].

The oxidized forms of 5mC represent important dynamic epigenetic states in the modulation of transcriptional programs and it may serve as signals for several specific “binders,” “readers” or “erasers.” It has been demonstrated that 5mC and its derivatives recruit distinct transcriptional regulators, polymerases and a large number of DNA repair proteins. Maintenance of the appropriate levels of 5mC and its derivatives is essential for cell homeostasis [108].

### 5.1.1 Oxidation of 5mC

The demethylation of CpG sequences can occur in passive or in an active manner [109,110]. Passive demethylation takes place due to the absence of DNMT1 or its cofactor SAM, during several rounds of DNA replication. After DNA replication the created hemimethylated site remains as such, and one daughter cell carries a completely unmethylated CpG site after the second round of replication. Passive demethylation plays important roles in certain biological contexts such as DNA demethylation in primordial germ cells, conversion of embryonic stem cells to a naïve pluripotent state and epigenome reprogramming after fertilization [109]. On the other hand, active DNA demethylation is mediated by TET proteins in mammals, which progressively oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), and by the thymine-DNA glycosidase (TDG) protein that binds and excises 5fC and 5caC residues to allow the restoration of unmodified C by the base excision repair machinery (Fig. 5.1) [111]. Thus, in mammals and plants, these modified bases may represent new

epigenetic state in the genomic DNA or intermediates in the process of 5mC demethylation [112,113].

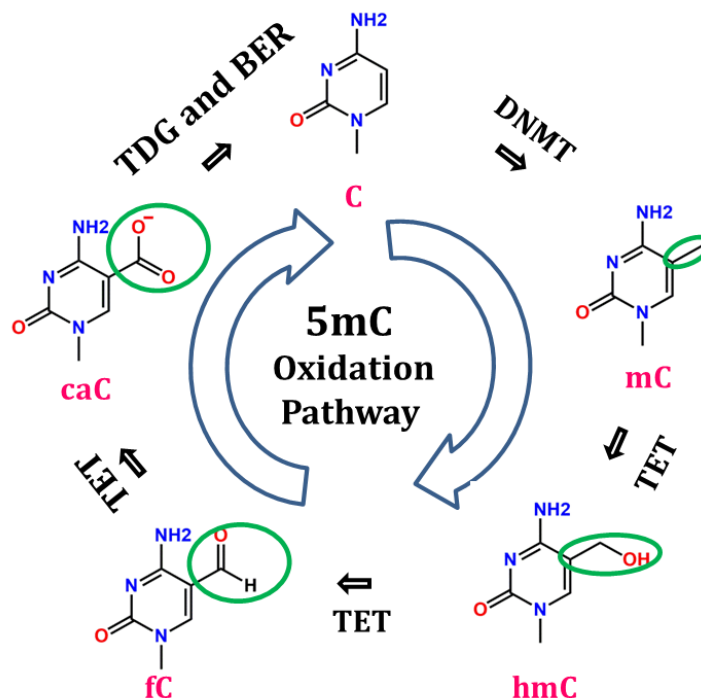


Figure 5.1: TET mediated 5mC demethylation.

### 5.1.2 Recognition of 5mC oxidation derivatives by the SRA domain

The DNA free structure of the 5hmC and 5 glucosylhydroxymethylcytosine (5ghmC)-specific endonuclease PvuRts1I indicated that its C-terminal SRA domain might accommodate a flipped out 5 hmC or 5ghmC base. Another 5hmC and 5 ghmC-specific restriction enzyme, AbaSI, recognizes intra-helical 5hmC through an SRA-like domain [114,115]. The SRA domain of UHRF1 (UHRF1 SRA) selectively recognizes 5mC over 5hmC, and preferentially binds hemi-5mCG over fully-5mCG. But SRA UHRF1 binds to other oxidation derivatives remains unknown.

Until recently, 5mC was considered the only epigenetic mark in genomic DNA. However, newly identified marks either antagonize the read-out and interpretation of 5mC or they can act as new set of epigenetic marks that are recognized by different epigenetic reader modules [116]. UHRF2, a close relative of UHRF1,

specifically binds 5hmC in neuronal progenitor cells. It is intriguing that the SRA domains of UHRF1 and UHRF2 selectively recognize hemi-5mCG/hemi-5hmCG and fully-5hmCG, respectively [38], even though they have 88% sequence similarity and high structural identity (RMSD: 0.73Å). UHRF1 binds to both hemi-5mCG and DNA methyltransferase 1 (DNMT1) to maintain the DNA methylation patterns in mammals. Our studies reveal that UHRF1 SRA recognizes 5hmC similar to 5mC. There is high possibility that UHRF1 can bind both hemi-5hmCG and DNMT1 on the genomic DNA, therefore, it may involve in maintenance of CG methylation even in the presence of 5hmC. Recently, it has been shown that the oxidation of 5mC to 5hmC clearly interferes with the DNA binding to the MBD domain of MeCP2, MBD1 and MBD2 [117,118].

To investigate whether UHRF1 SRA binds to 5mC oxidation derivatives in the context of a hemi-CG sequence, and if so, to know the preferred oxidation derivative recognition, we carried out binding studies using EMSA and FP techniques.

## 5.2 Material and methods

### 5.2.1 Sub-Cloning of UHRF1 SRA to generate hexahistidine-SUMO tagged construct

The cDNA encoding full-length human UHRF1 was obtained from Open biosystems. Residues 427-630, that corresponds to the UHRF1 SRA, was sub-cloned. Amplification was carried out using forward and reverse primers containing NdeI and BamHI restriction sites, respectively. The amplified PCR product was confirmed by agarose gel electrophoresis, followed by DNA gel extraction and subsequent digestion with NdeI and BamHI. The purified then restriction digested product was ligated with NdeI-BamHI digested pET28-N-His-SUMO vector using T4 DNA ligase. Schematic represents the workflow of cloning strategy (Fig. 5.3). Clones obtained were screened by using colony PCR and confirmed by sequencing.

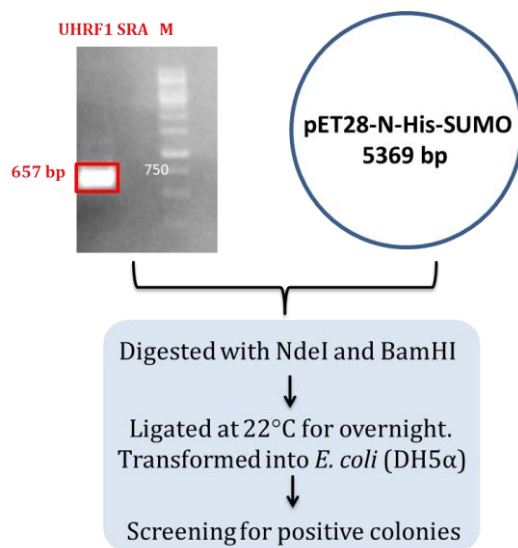


Figure 5.3: Flowchart of UHRF1 SRA sub-cloning.

### 5.2.2 Expression and Purification of UHRF1 SRA

The above construct containing hexahistidine-SUMO tagged UHRF1 SRA (residues 427-630) was expressed in *E. coli* Rosetta2(DE3). The expressed protein was purified on a nickel charged column (HisTrap HP, GE healthcare). The fusion protein was cleaved with Ulp1 protease at 15  $\mu\text{M}$  concentrations. The protein was further purified by cation-exchange (HiTrap Heparin HP) chromatography. Finally, the protein was purified using gel filtration chromatography using a 'HiLoad Superdex 75 16/600' column, which was equilibrated with a buffer containing 15 mM Tris-HCl, pH 7.5, 100 mM NaCl, 3 mM DTT and 2.5% Glycerol (Fig. 5.4). Purified protein was concentrated to 15  $\text{mg mL}^{-1}$  at 4°C in Vivaspin 20 mL (Vivascience AG) 10,000 cut-off concentrator.

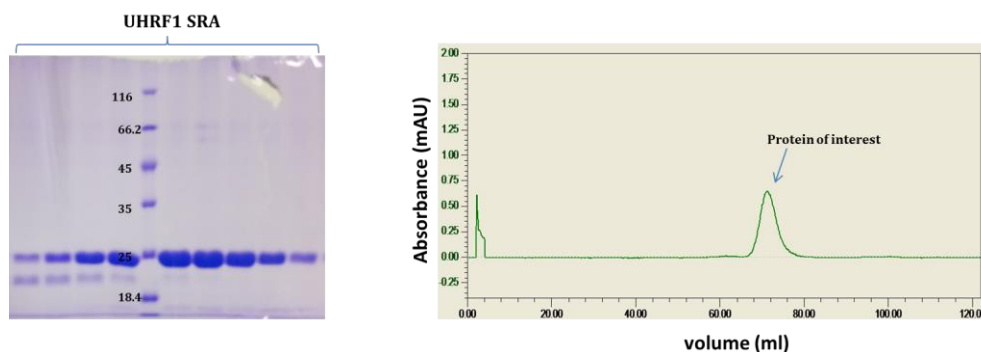


Figure 5.4: SDS-PAGE analysis of purified UHRF1 SRA and its gel-filtration chromatogram.

### 5.2.3 Qualitative analysis of SRA-DNA complex by SEC

Binding of UHRF1 SRA domain and 5'-6FAM-labeled duplex DNA was carried out by using SEC (Size Exclusion Chromatography) method. DNA oligos were added in the ratio of 1:1.1 FAM labeled to modified (5mC/5hmC/5caC/5fC) unlabeled complementary DNA and annealing is carried out in the buffer containing 10mM Tris pH:7.5, 50mM NaCl and 3mM MgCl<sub>2</sub>. The oligos were heated at 95°C for 5min and cooled at 4°C for overnight. UHRF1 SRA incubated with FAM labeled DNA duplex in the molar ratio of 2:1. Finally, the complexes are resolved using Superdex 75 10/300 GL (GE Healthcare). Schematic flowchart represents the workflow of DNA annealing and SEC analysis of UHRF1 SRA-DNA complex (Fig. 5.5).

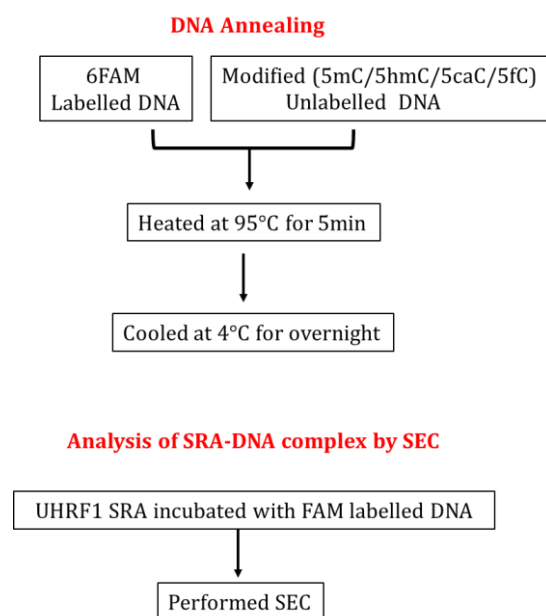


Figure 5.5: Schematic of DNA annealing and SEC analysis of UHRF1 SRA-DNA complex.

### 5.2.4 Electrophoretic Mobility Shift Assay (EMSA)

Binding between the UHRF1 SRA domain and 5'-6FAM-labeled duplex DNA was carried out in a buffer containing 25 mM Tris-HCl, pH 7.5, 5% Glycerol, 60 mM

NaCl, 10 mM MgCl<sub>2</sub>, 0.4 mg/mL BSA and 2 mM dithiothreitol with increasing amounts of UHRF1 SRA (200 to 1000 pmol for hemi-5mCG and hemi-5hmCG binding study and, 600 pmol to 3 nmol for hemi-5fCG and hemi-5caCG DNA), and fixed amount of modified DNA duplex (3 µg of hemi-5mCG and hemi-5hmCG DNA, and 9 µg of hemi-5fCG and hemi-5caCG DNA). The 20 µL of reaction mixtures were incubated at 4°C for 45 minutes then electrophoresed on a native 6% polyacrylamide gel at 100 V for 1.0 hour in a buffer containing 100 mM Tris-Borate-EDTA and 2.5% glycerol in the cold room. The gels were fluorescence scanned using Typhoon TRIO Variable Mode Imager

### 5.2.5 Fluorescence Polarization (FP) measurements

Fluorescence Polarization measurements of UHRF1 SRA domain binding to 5'-6FAM labeled duplex DNA was performed at 20 °C. The 5'-6FAM-labeled duplex DNA (50 nM) was added to a UHRF1 SRA domain with increasing concentration (50 nM to 50 µM). The polarization (in millipolarization [mP] units) of 20 µL of the reaction mixture was measured after 45 minutes of incubation at room temperature. Polarization values were measured against a blank sample buffer (20 mM Tris-HCl pH 7.5, 50 mM NaCl and 1 mM DTT) and a reference contained only 50 nM DNA. The polarization data was analyzed using GraphPad Prism 5.0 software (GraphPad, San Diego, CA, USA) to calculate a dissociation constant ( $K_D$ ) by fitting the experimental data in a non-linear regression equation using a one-site specific binding model accounting for ligand depletion. Experiments were performed in triplicate.

## 5.3 Results

We have determined the qualitative binding of UHRF1 SRA with hemi-5mCG and hemi-5hmCG. This gel filtration binding studies showed that the UHRF1 SRA binds to both hemi-5mCG and hemi-5hmCG containing DNAs (Fig. 5.6). Further, we have performed EMSA and FP studies of the UHRF1 SRA with hemi-5mCG,

hemi-5hmCG, hemi-5fCG and hemi-5caCG duplex DNAs. We attempted to determine the  $K_D$  for binding of 5mC oxidation derivative bases in hemi-CG sequence context towards UHRF1 SRA through ITC measurements, however, there was negligible heat change during ITC titration. Therefore, we decided to employ Electrophoretic Mobility Shift Assay (EMSA) and Fluorescence Polarization (FP) techniques.

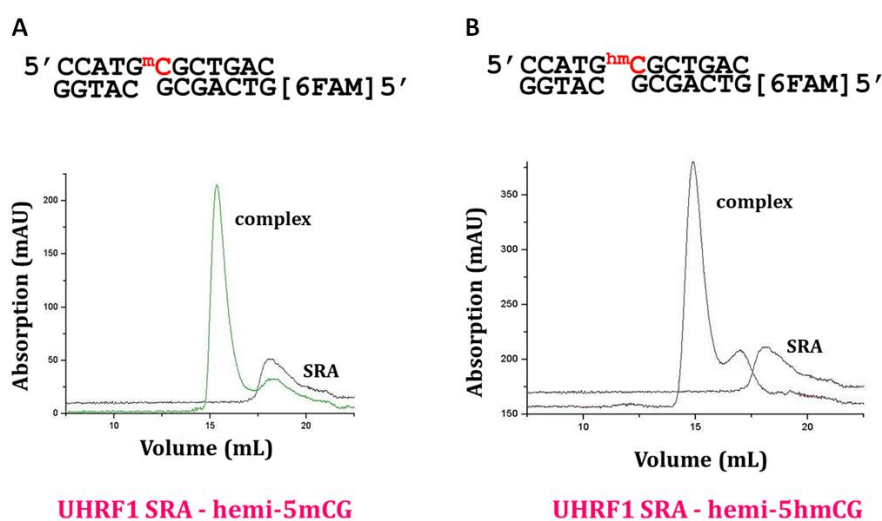


Figure 5.6: Gel filtration elution profile of UHRF1 SRA-DNA complex.

Although UHRF1 SRA binding to hemi-5mCG and hemi-5hmCG was reported, no data is available on its ability to recognize hemi-5caCG and hemi-5fCG modifications. Here, we performed EMSA, using 5'-6FAM labeled DNA duplexes in a CG sequence context, to investigate the UHRF1 SRA specificity for hemi-5caCG and hemi-5fCG DNA duplexes. Our control EMSA experiments revealed that SRA UHRF1 binds to hemi-5mCG and hemi-5hmCG DNA similar to published data (Fig. 5.7A and B). Though UHRF1 SRA could also recognize hemi-5caCG and hemi-fCG, the concentration of protein and modified DNA used for binding study is more than that used for 5hmC owing to weaker binding affinity towards 5caC and 5fC containing DNAs (Fig. 5.7C and D). Furthermore, our finding showed that the SRA domain of UHRF1 binds to all 5mC oxidation derivatives.

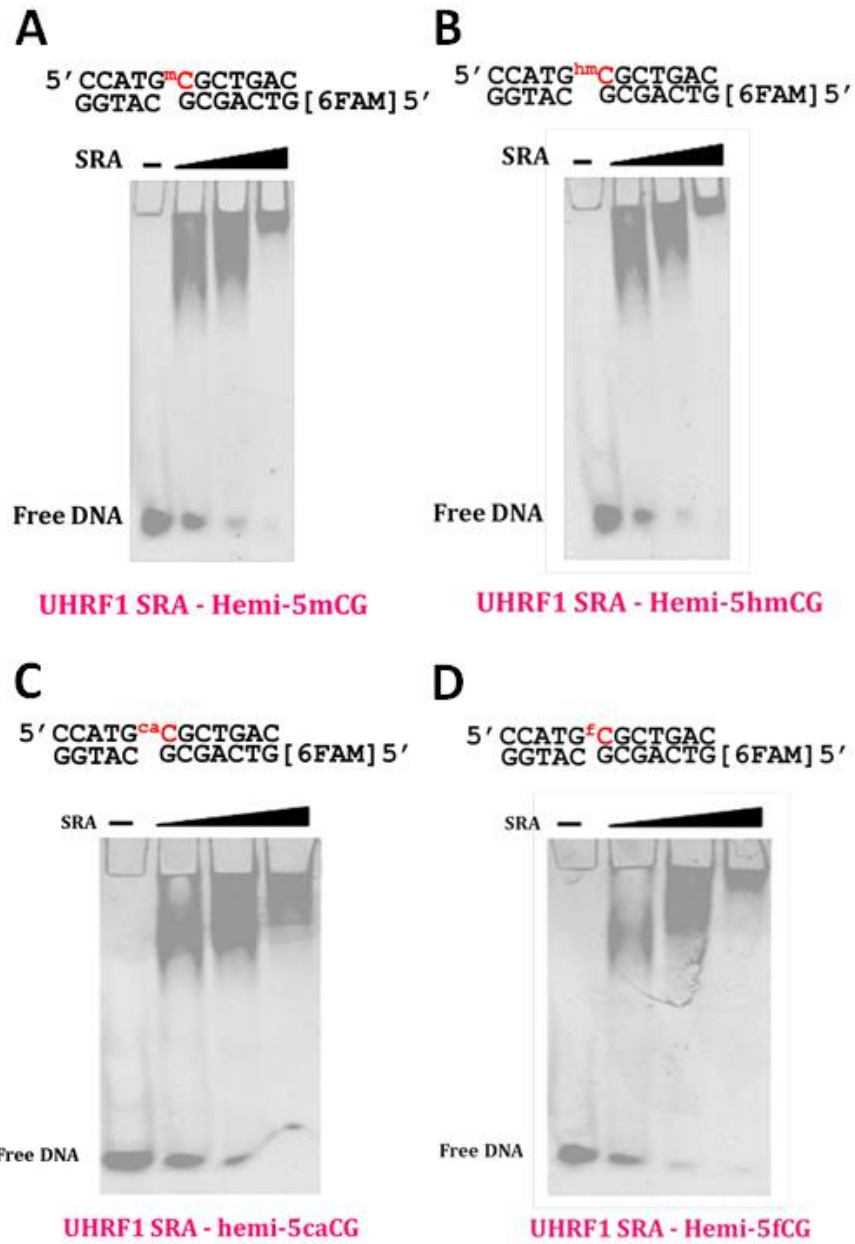


Figure 5.7: Binding of the UHRF1 SRA domain to hemi-5mCG, hemi-5hmCG, hemi-5caCG and hemi-5fCG containing duplex DNAs. (A) EMSA binding study of SRA UHRF1 with hemi-5mCG DNA. In the panel above the gel images, (—) indicates no protein but only DNA. Un-bound (free) DNA is indicated. (B) EMSA binding study of SRA UHRF1 with hemi-5hmCG DNA. (C) EMSA binding study of SRA UHRF1 with hemi-5caCG DNA. (D) EMSA binding study of SRA UHRF1 with hemi-5fCG DNA.



To quantify the UHRF1 SRA binding affinity for 5mC oxidation derivatives in hemi-CG sequence context, we performed FP measurements using 5'-6FAM labeled DNA duplexes. UHRF1 SRA domain binds to hemi-5mCG DNA with a  $K_D$  of 0.35  $\mu\text{M}$  which is similar (0.20  $\mu\text{M}$ ) to reported in literature [119]. However, the binding constant decreases by a factor of 4.3 ( $K_D = 1.5 \mu\text{M}$ ) for hemi-5hmCG (Fig. 5.8).

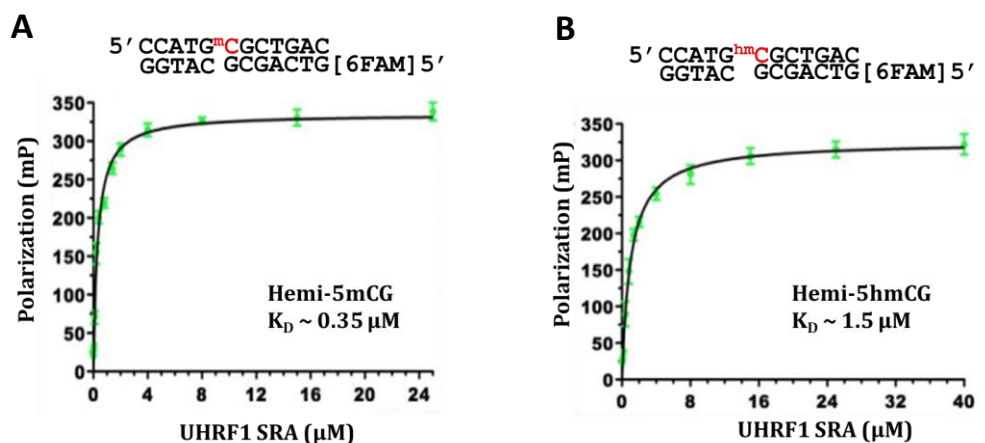


Figure 5.8: Fluorescence Polarization (FP) measurements of the binding of the UHRF1 SRA domain to (A) hemi-5mCG (B) hemi-5hmCG DNA.

#### 5.4 Discussion

The current study indicates that the SRA domains of UHRF1 exhibit lower binding affinity for 5fC- and 5caC-containing DNAs. UHRF1 involved in transcriptional repression through 5mC recognition, the oxidation of 5mC to 5caC could promote a switch from a repressive to an active transcriptional state of the chromatin, thereby changing the cellular interpretation of the 5mC epigenetic marks. Many hmC, fC and caC binders such as UHRF2, CARF and p53 are implicated in cancer. Finally, regulating the levels of mC and its oxidized derivatives is essential for normal cell homeostasis and that deregulation of the readers, writers, and erasers of these marks results in a disturbance of the balance between cell proliferation and differentiation during development.

## 6.1 Conclusion and future perspectives

In conclusion, in this study, we have characterized the histone methyllysine binding specificity of UHRF1 TTD. MD simulation and *in vitro* binding studies reveal that TTD, as standalone domain and also as PHD linked domain, preferentially binds H3K9me2 mark, and UHRF1 PHD has an insignificant effect on K9me2 mark recognition by the UHRF1 TTD domain in the context of UHRF1 TTD-PHD cassette. The methylation of H3K4 has not perturbed the combinatorial recognition of H3R2 and H3K9me2 marks by the linked PHD and TUDOR domains of UHRF1. Our study also demonstrates that the UHRF1 SRA binds all 5mC oxidation derivatives and recognizes 5hmC like 5mC.

*In vitro* binding analyses and/or structural studies to determine the effect of post-translational modifications on receptors, residue substitution or functional group changes on peptide or other ligands for receptor-ligand interaction will be laborious and costly. Therefore, this study suggests that MD simulation on modeled protein-modified ligand complexes may give insights on the effect of such modifications on ligand binding. Results of such MD simulation studies can be the basis for further biochemical and structural characterization, and for the lead optimization in drug discovery process. In future, we plan to use MD simulation based structural analysis to identify the residues to engineer UHRF1 TTD-PHD to switch its methyllysine status binding specificity. The engineered UHRF1 domain can be used to address the effect of a change in methyllysine status binding specificity on UHRF1's functions, in *in vivo*, such as heterochromatin formation, DNA repair, cell proliferation, and stem cell self-renewal and differentiation.

## References

- 1 Dupont, C., Armant, D. and Brenner, C. (2009) Epigenetics: Definition, Mechanisms and Clinical Perspective. *Seminars in Reproductive Medicine* 27, 351–357.
- 2 Peixoto, L. and Abel, T. (2013) The role of histone acetylation in memory formation and cognitive impairments. *Neuropsychopharmacology* 38, 62–76.
- 3 Trerotola, M., Relli, V., Simeone, P. and Alberti, S. (2015) Epigenetic inheritance and the missing heritability. *Hum Genomics* 9.
- 4 Mahmoud, S. A. and Poizat, C. (2013) Epigenetics and chromatin remodeling in adult cardiomyopathy. *J. Pathol.* 231, 147–157.
- 5 Turner, B. M. (2012) The adjustable nucleosome: an epigenetic signaling module. *Trends Genet.* 28, 436–444.
- 6 Quina, A. S., Buschbeck, M. and Di Croce, L. (2006) Chromatin structure and epigenetics. *Biochem. Pharmacol.* 72, 1563–1569.
- 7 Biterge, B. (2016) A Mini Review on Post-Translational Histone Modifications. *MOJ Cell Science & Report* 3.
- 8 Cutter, A. R. and Hayes, J. J. (2015) A brief review of nucleosome structure. *FEBS Letters* 589, 2914–2922.
- 9 Zhou, B.-R., Jiang, J., Feng, H., Ghirlando, R., Xiao, T. S. and Bai, Y. (2015) Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Molecular Cell* 59, 628–638.
- 10 Luger, K. (2001) Nucleosomes: Structure and Function. In *Encyclopedia of Life Sciences* (John Wiley & Sons, Ltd, ed.), John Wiley & Sons, Ltd, Chichester.
- 11 McGinty, R. K. and Tan, S. (2015) Nucleosome Structure and Function. *Chemical Reviews* 115, 2255–2273.
- 12 The Nucleosome From Genomic Review.pdf.
- 13 Tremethick, D. J. (2007) Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell* 128, 651–654.
- 14 Li, G. and Zhu, P. (2015) Structure and organization of chromatin fiber in the nucleus. *FEBS Letters* 589, 2893–2904.
- 15 Lawrence, M., Daujat, S. and Schneider, R. (2016) Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends in Genetics* 32, 42–56.
- 16 Kouzarides, T. (2007) Chromatin Modifications and Their Function. *Cell* 128, 693–705.
- 17 Zhang, T., Cooper, S. and Brockdorff, N. (2015) The interplay of histone modifications - writers that read. *EMBO reports* 16, 1467–1481.
- 18 Andrews, F. H., Strahl, B. D. and Kutateladze, T. G. (2016) Insights into newly discovered marks and readers of epigenetic information. *Nature Chemical Biology* 12, 662–668.
- 19 Gillette, T. G. and Hill, J. A. (2015) Readers, Writers, and Erasers: Chromatin as the Whiteboard of Heart Disease. *Circulation Research* 116, 1245–1253.

- 20 Pande, V. (2016) Understanding the Complexity of Epigenetic Target Space: Miniperspective. *Journal of Medicinal Chemistry* 59, 1299–1307.
- 21 Marmorstein, R. and Zhou, M.-M. (2014) Writers and Readers of Histone Acetylation: Structure, Mechanism, and Inhibition. *Cold Spring Harbor Perspectives in Biology* 6, a018762–a018762.
- 22 Su, Z. and Denu, J. M. (2016) Reading the Combinatorial Histone Language. *ACS Chemical Biology* 11, 564–574.
- 23 Wang, Y., Han, Y., Fan, E. and Zhang, K. (2015) Analytical strategies used to identify the readers of histone modifications: A review. *Analytica Chimica Acta* 891, 32–42.
- 24 Yun, M. (2011) Readers of histone modifications. *Cell Research* 21, 15.
- 25 Lomberk, G., Wallrath, L. L. and Urrutia, R. (2006) The Heterochromatin Protein 1 family. *Genome Biology* 8.
- 26 Canzio, D., Larson, A. and Narlikar, G. J. (2014) Mechanisms of functional promiscuity by HP1 proteins. *Trends in Cell Biology* 24, 377–386.
- 27 Clapier, C. R., Iwasa, J., Cairns, B. R. and Peterson, C. L. (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology* 18, 407–422.
- 28 Längst, G. and Manelyte, L. (2015) Chromatin Remodelers: From Function to Dysfunction. *Genes* 6, 299–324.
- 29 Chen, T. and Dent, S. Y. R. (2014) Chromatin modifiers and remodelers: regulators of cellular differentiation. *Nature Reviews Genetics* 15, 93–106.
- 30 Black, J. C., Van Rechem, C. and Whetstine, J. R. (2012) Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Molecular Cell* 48, 491–507.
- 31 Sims, R. J., Nishioka, K. and Reinberg, D. (2003) Histone lysine methylation: a signature for chromatin function. *Trends in Genetics* 19, 629–639.
- 32 Watanabe, H., Soejima, K., Yasuda, H., Kawada, I., Nakachi, I., Yoda, S., Naoki, K. and Ishizaka, A. (2008) Deregulation of histone lysine methyltransferases contributes to oncogenic transformation of human bronchoepithelial cells. *Cancer Cell Int* 8, 15.
- 33 Hyun, K., Jeon, J., Park, K. and Kim, J. (2017) Writing, erasing and reading histone lysine methylations. *Experimental & Molecular Medicine* 49, e324–e324.
- 34 Jin, B., Li, Y. and Robertson, K. D. (2011) DNA methylation: Superior or subordinate in the epigenetic hierarchy? *Genes and Cancer* 2, 607–617.
- 35 Ludwig, A. K., Zhang, P. and Cardoso, M. C. (2016) Modifiers and Readers of DNA Modifications and Their Impact on Genome Structure, Expression, and Stability in Disease. *Front Genet* 7.
- 36 Rajakumara, E., Nakarakanti, N. K., Nivya, M. A. and Satish, M. (2016) Mechanistic insights into the recognition of 5-methylcytosine oxidation derivatives by the SUVH5 SRA domain. *Sci Rep* 6, 20161.

- 37 You, C., Ji, D., Dai, X. and Wang, Y. (2014) Effects of Tet-mediated Oxidation Products of 5-Methylcytosine on DNA Transcription *in vitro* and in Mammalian Cells. *Scientific Reports* 4, 7052.
- 38 Zhou, T., Xiong, J., Wang, M., Yang, N., Wong, J., Zhu, B. and Xu, R.-M. (2014) Structural Basis for Hydroxymethylcytosine Recognition by the SRA Domain of UHRF2. *Molecular Cell* 54, 879–886.
- 39 Wood, K. H. and Zhou, Z. (2016) Emerging Molecular and Biological Functions of MBD2, a Reader of DNA Methylation. *Front Genet* 7.
- 40 Li, E. and Zhang, Y. (2014) DNA methylation in mammals. *Cold Spring Harb Perspect Biol* 6, a019133.
- 41 Moore, L. D., Le, T. and Fan, G. (2013) DNA Methylation and Its Basic Function. *Neuropsychopharmacology* 38, 23–38.
- 42 Du, J., Johnson, L. M., Jacobsen, S. E. and Patel, D. J. (2015) DNA methylation pathways and their crosstalk with histone methylation. *Nature Reviews Molecular Cell Biology* 16, 519–532.
- 43 Kondo, Y. (2009) Epigenetic Cross-Talk between DNA Methylation and Histone Modifications in Human Cancers. *Yonsei Medical Journal* 50, 455.
- 44 Rose, N. R. and Klose, R. J. (2014) Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1839, 1362–1372.
- 45 Cedar, H. and Bergman, Y. (2009) Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics* 10, 295–304.
- 46 Cheng, J., Yang, Y., Fang, J., Xiao, J., Zhu, T., Chen, F., Wang, P., Li, Z., Yang, H. and Xu, Y. (2013) Structural Insight into Coordinated Recognition of Trimethylated Histone H3 Lysine 9 (H3K9me3) by the Plant Homeodomain (PHD) and Tandem Tudor Domain (TTD) of UHRF1 (Ubiquitin-like, Containing PHD and RING Finger Domains, 1) Protein. *Journal of Biological Chemistry* 288, 1329–1339.
- 47 Alhosin, M., Omran, Z., Zamzami, M. A., Al-Malki, A. L., Choudhry, H., Mousli, M. and Bronner, C. (2016) Signalling pathways in UHRF1-dependent regulation of tumor suppressor genes in cancer. *Journal of Experimental & Clinical Cancer Research* 35.
- 48 Tauber, M. and Fischle, W. (2015) Conserved linker regions and their regulation determine multiple chromatin-binding modes of UHRF1. *Nucleus* 6, 123–132.
- 49 Rajakumara, E., Wang, Z., Ma, H., Hu, L., Chen, H., Lin, Y., Guo, R., Wu, F., Li, H., Lan, F., et al. (2011) PHD Finger Recognition of Unmodified Histone H3R2 Links UHRF1 to Regulation of Euchromatic Gene Expression. *Molecular Cell* 43, 275–284.
- 50 Bortoluzzi, A., Amato, A., Lucas, X., Blank, M. and Ciulli, A. (2017) Structural basis of molecular recognition of helical histone H3 tail by PHD finger domains. *Biochemical Journal* 474, 1633–1651.

- 51 Hashimoto, H., Horton, J. R., Zhang, X. and Cheng, X. (2009) UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics* 4, 8–14.
- 52 Jiang, Y., Liu, L., Shan, W. and Yang, Z.-Q. (2016) An integrated genomic analysis of Tudor domain-containing proteins identifies PHD finger protein 20-like 1 (PHF20L1) as a candidate oncogene in breast cancer. *Molecular Oncology* 10, 292–302.
- 53 Ying, M. and Chen, D. (2012) Tudor domain-containing proteins of *Drosophila melanogaster*: Functional groups of Tudor proteins. *Development, Growth & Differentiation* 54, 32–43.
- 54 Mathioudakis, N., Palencia, A., Kadlec, J., Round, A., Tripsianes, K., Sattler, M., Pillai, R. S. and Cusack, S. (2012) The multiple Tudor domain-containing protein TDRD1 is a molecular scaffold for mouse Piwi proteins and piRNA biogenesis factors. *RNA* 18, 2056–2072.
- 55 Adams, M. A. and Min, J. Structure And Function Of Histone Code Reading Proteins 46.
- 56 Yap, K. L. and Zhou, M.-M. (2010) Keeping it in the family: diverse histone recognition by conserved structural folds. *Critical Reviews in Biochemistry and Molecular Biology* 45, 488–505.
- 57 Carr, S. M., Munro, S., Zalmas, L.-P., Fedorov, O., Johansson, C., Krojer, T., Sagum, C. A., Bedford, M. T., Oppermann, U. and La Thangue, N. B. (2014) Lysine methylation-dependent binding of 53BP1 to the pRb tumor suppressor. *Proceedings of the National Academy of Sciences* 111, 11341–11346.
- 58 Kachirskaia, I., Shi, X., Yamaguchi, H., Tanoue, K., Wen, H., Wang, E. W., Appella, E. and Gozani, O. (2008) Role for 53BP1 Tudor Domain Recognition of p53 Dimethylated at Lysine 382 in DNA Damage Signaling. *Journal of Biological Chemistry* 283, 34660–34666.
- 59 Nady, N., Lemak, A., Walker, J. R., Avvakumov, G. V., Kareta, M. S., Achour, M., Xue, S., Duan, S., Allali-Hassani, A., Zuo, X., et al. (2011) Recognition of Multivalent Histone States Associated with Heterochromatin by UHRF1 Protein. *Journal of Biological Chemistry* 286, 24300–24311.
- 60 Botuyan, M. V., Lee, J., Ward, I. M., Kim, J.-E., Thompson, J. R., Chen, J. and Mer, G. (2006) Structural Basis for the Methylation State-Specific Recognition of Histone H4-K20 by 53BP1 and Crb2 in DNA Repair. *Cell* 127, 1361–1373.
- 61 Teske, K. A. and Hadden, M. K. (2017) Methyllysine binding domains: Structural insight and small molecule probe development. *European Journal of Medicinal Chemistry* 136, 14–35.
- 62 Milosevich, N. and Hof, F. (2016) Chemical Inhibitors of Epigenetic Methyllysine Reader Proteins. *Biochemistry* 55, 1570–1583.
- 63 Herold, J. M., Ingerman, L. A., Gao, C. and Frye, S. V. (2011) Drug Discovery Toward Antagonists of Methyl-Lysine Binding Proteins. *Current Chemical Genomics* 5, 51–61.

- 64 Eissenberg, J. C. (2012) Structural biology of the chromodomain: Form and function. *Gene* 496, 69–78.
- 65 Fischle, W. (2003) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & Development* 17, 1870–1881.
- 66 Yang, Y., Hu, L., Wang, P., Hou, H., Lin, Y., Liu, Y., Li, Z., Gong, R., Feng, X., Zhou, L., et al. (2010) Structural insights into a dual-specificity histone demethylase ceKDM7A from *Caenorhabditis elegans*. *Cell Research* 20, 886–898.
- 67 Greer, E. L. and Shi, Y. (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* 13, 343–357.
- 68 Mohan, M., Herz, H.-M., Takahashi, Y.-H., Lin, C., Lai, K. C., Zhang, Y., Washburn, M. P., Florens, L. and Shilatifard, A. (2010) Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). *Genes & Development* 24, 574–589.
- 69 Swamy, M. J. and Sankhala, R. S. (2013) Probing the thermodynamics of protein-lipid interactions by isothermal titration calorimetry. *Methods Mol. Biol.* 974, 37–53.
- 70 Pierce, M. M., Raman, C. S. and Nall, B. T. (1999) Isothermal Titration Calorimetry of Protein–Protein Interactions. *Methods* 19, 213–221.
- 71 Brown, A. (2009) Analysis of Cooperativity by Isothermal Titration Calorimetry. *International Journal of Molecular Sciences* 10, 3457–3477.
- 72 Arita, K., Isogai, S., Oda, T., Unoki, M., Sugita, K., Sekiyama, N., Kuwata, K., Hamamoto, R., Tochio, H., Sato, M., et al. (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 Sciences* 109, 12950–12955.
- 73 Wang, Z. and Patel, D. J. (2011) Combinatorial Readout of Dual Histone Modifications by Paired Chromatin-associated Modules. *Journal of Biological Chemistry* 286, 18363–18368.
- 74 Musselman, C. A., Lalonde, M.-E., Côté, J. and Kutateladze, T. G. (2012) Perceiving the epigenetic landscape through histone readers. *Nature Structural & Molecular Biology* 19, 1218–1227.
- 75 Du, J. and Patel, D. J. (2014) Structural biology-based insights into combinatorial readout and crosstalk among epigenetic marks. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1839, 719–727.
- 76 Ruthenburg, A. J., Li, H., Patel, D. J. and David Allis, C. (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nature Reviews Molecular Cell Biology* 8, 983–994.
- 77 Musselman, C. A., Ramirez, J., Sims, J. K., Mansfield, R. E., Oliver, S. S., Denu, J. M., Mackay, J. P., Wade, P. A., Hagman, J. and Kutateladze, T. G. (2012) Bivalent recognition of nucleosomes by the tandem PHD fingers of the

- CHD4 ATPase is required for CHD4-mediated repression. *Proceedings of the National Academy of Sciences* 109, 787–792.
- 78 Xie, S., Jakoncic, J. and Qian, C. (2012) UHRF1 Double Tudor Domain and the Adjacent PHD Finger Act Together to Recognize K9me3-Containing Histone H3 Tail. *Journal of Molecular Biology* 415, 318–328.
- 79 Gelato, K. A., Tauber, M., Ong, M. S., Winter, S., Hiragami-Hamada, K., Sindlinger, J., Lemak, A., Bultsma, Y., Houlston, S., Schwarzer, D., et al. (2014) Accessibility of Different Histone H3-Binding Domains of UHRF1 Is Allosterically Regulated by Phosphatidylinositol 5-Phosphate. *Molecular Cell* 54, 905–919.
- 80 Rothbart, S. B. and Strahl, B. D. (2014) Interpreting the language of histone and DNA modifications. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1839, 627–643.
- 81 Houlston, R. S., Lemak, A., Iqbal, A., Ivanochko, D., Duan, S., Kaustov, L., Ong, M. S., Fan, L., Senisterra, G., Brown, P. J., et al. (2017) Conformational dynamics of the TTD–PHD histone reader module of the UHRF1 epigenetic regulator reveals multiple histone-binding states, allosteric regulation, and druggability. *Journal of Biological Chemistry* 292, 20947–20959.
- 82 Hall, M. D., Yasgar, A., Peryea, T., Braisted, J. C., Jadhav, A., Simeonov, A. and Coussens, N. P. (2016) Fluorescence polarization assays in high-throughput screening and drug discovery: a review. *Methods and Applications in Fluorescence* 4, 022001.
- 83 Jameson, D. M. and Ross, J. A. (2010) Fluorescence Polarization/Anisotropy in Diagnostics and Imaging. *Chemical Reviews* 110, 2685–2708.
- 84 Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H. and Shaw, D. E. (2012) Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics* 41, 429–452.
- 85 Patodia, S. (2014) Molecular Dynamics Simulation of Proteins: A Brief Overview. *Journal of Physical Chemistry & Biophysics* 4.
- 86 Adcock, S. A. (2008) Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins 53.
- 87 Li, H., Robertson, A. D. and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Structure, Function, and Bioinformatics* 61, 704–721.
- 88 Bas, D. C., Rogers, D. M. and Jensen, J. H. (2008) Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins: Structure, Function, and Bioinformatics* 73, 765–783.
- 89 Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *The Journal of Chemical Physics* 98, 10089–10092.
- 90 Hoover, W. G. (1985) Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* 31, 1695–1697.



- 91 Martyna, G. J., Tobias, D. J. and Klein, M. L. (1994) Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* 101, 4177–4189.
- 92 Li, H., Fischle, W., Wang, W., Duncan, E. M., Liang, L., Murakami-Ishibe, S., Allis, C. D. and Patel, D. J. (2007) Structural Basis for Lower Lysine Methylation State-Specific Readout by MBT Repeats of L3MBTL1 and an Engineered PHD Finger. *Molecular Cell* 28, 677–691.
- 93 Min, J., Allali-Hassani, A., Nady, N., Qi, C., Ouyang, H., Liu, Y., MacKenzie, F., Vedadi, M. and Arrowsmith, C. H. (2007) L3MBTL1 recognition of mono- and dimethylated histones. *Nature Structural & Molecular Biology* 14, 1229–1230.
- 94 Tuning HP1 $\alpha$  chromodomain selectivity for di- and.pdf.
- 95 Banerjee, T. and Chakravarti, D. (2011) A Peek into the Complex Realm of Histone Phosphorylation. *Molecular and Cellular Biology* 31, 4858–4873.
- 96 Prakash, K. and Fournier, D. (2017) Deciphering the histone code to build the genome structure.
- 97 Fischle, W., Wang, Y. and Allis, C. D. (2003) Histone and chromatin cross-talk. *Current Opinion in Cell Biology* 15, 172–183.
- 98 Hunter, T. (2007) The Age of Crosstalk: Phosphorylation, Ubiquitination, and Beyond. *Molecular Cell* 28, 730–738.
- 99 Izzo, A. and Schneider, R. (2010) Chatting histone modifications in mammals. *Briefings in Functional Genomics* 9, 429–443.
- 100 Bannister, A. J. and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Research* 21, 381–395.
- 101 Andrews, F. H., Gatchalian, J., Krajewski, K., Strahl, B. D. and Kutateladze, T. G. (2016) Regulation of Methyllysine Readers through Phosphorylation. *ACS Chemical Biology* 11, 547–553.
- 102 Kunowska, N., Rotival, M., Yu, L., Choudhary, J. and Dillon, N. (2015) Identification of protein complexes that bind to histone H3 combinatorial modifications using super-SILAC and weighted correlation network analysis. *Nucleic Acids Research* 43, 1418–1432.
- 103 Su, Z., Wang, F., Lee, J.-H., Stephens, K. E., Papazyan, R., Voronina, E., Krautkramer, K. A., Raman, A., Thorpe, J. J., Boersma, M. D., et al. (2016) Reader domain specificity and lysine demethylase-4 family function. *Nature Communications* 7, 13387.
- 104 Di Lorenzo, A. and Bedford, M. T. (2011) Histone arginine methylation. *FEBS Letters* 585, 2024–2031.
- 105 Fischle, W. (2008) Talk is cheap--cross-talk in establishment, maintenance, and readout of chromatin modifications. *Genes & Development* 22, 3375–3382.
- 106 Klungland, A. and Robertson, A. B. (2017) Oxidized C5-methyl cytosine bases in DNA: 5-Hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine. *Free Radical Biology and Medicine* 107, 62–68.

- 107 Song, C.-X. and He, C. (2013) Potential functional roles of DNA demethylation intermediates. *Trends in Biochemical Sciences* 38, 480–484.
- 108 Liu, J., Cui, X., Jiang, J., Cao, D., He, Y. and Wang, H. (2017) Uncoordinated expression of DNA methylation-related enzymes in human cancer. *Epigenetics & Chromatin* 10.
- 109 Song, J. and Pfeifer, G. P. (2016) Are there specific readers of oxidized 5-methylcytosine bases? *BioEssays* 38, 1038–1047.
- 110 Ficiz, G. (2015) New insights into mechanisms that regulate DNA methylation patterning. *Journal of Experimental Biology* 218, 14–20.
- 111 Wu, H. and Zhang, Y. (2011) Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes & Development* 25, 2436–2452.
- 112 Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Anselmi, F., Parlato, C., Medana, C., Dal Bello, F. and Oliviero, S. (2015) Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Reports* 10, 674–683.
- 113 Pfeifer, G. P., Kadam, S. and Jin, S.-G. (2013) 5-hydroxymethylcytosine and its potential roles in development and cancer. *Epigenetics & Chromatin* 6, 10.
- 114 Kazrani, A. A., Kowalska, M., Czapinska, H. and Bochtler, M. (2014) Crystal structure of the 5hmC specific endonuclease PvuRts1I. *Nucleic Acids Research* 42, 5929–5936.
- 115 Horton, J. R., Borgaro, J. G., Griggs, R. M., Quimby, A., Guan, S., Zhang, X., Wilson, G. G., Zheng, Y., Zhu, Z. and Cheng, X. (2014) Structure of 5-hydroxymethylcytosine-specific restriction enzyme, AbaSI, in complex with DNA. *Nucleic Acids Research* 42, 7947–7959.
- 116 Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W. T. C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (2013) Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* 152, 1146–1159.
- 117 Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M. C., Antes, I. and Leonhardt, H. (2011) Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain. *PLoS ONE* (Xu, S., ed.) 6, e21306.
- 118 Hashimoto, H., Liu, Y., Upadhyay, A. K., Chang, Y., Howerton, S. B., Vertino, P. M., Zhang, X. and Cheng, X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research* 40, 4841–4849.
- 119 Qian, C., Li, S., Jakoncic, J., Zeng, L., Walsh, M. J. and Zhou, M.-M. (2008) Structure and Hemimethylated CpG Binding of the SRA Domain from Human UHRF1. *Journal of Biological Chemistry* 283, 34490–34494.

## Appendix I

Antibiotics and other reagents:

Antibiotic/Reagents	Stock solution	Working solution
Kanamycin	50mg/ml in distilled water	50µg/ml
Chloramphenicol	35mg/ml in ethanol	35µg/ml
Ethidium Bromide	10mg/ml in distilled water	0.5µg/ml
IPTG	1.0M in distilled water	0.5mM
Lysozyme	10mg/ml in distilled water	As per requirement

Bacterial strains:

Bacterial strains	Purpose	Company
DH5α	Used for general cloning and plasmid propagation	Life technologies
Rosetta2(DE3)	Used for Protein expression	Novagen

Plasmid Vectors:

Vector	Expression host	Purpose	Antibiotic selection	Company
pET28-N-His-SUMO	Bacteria	Used for generating His-tagged recombinant proteins	Kanamycin	Novagen

Electrophoresis Running Buffer:

Buffer	Composition
DNA electrophoresis	
Tris Acetate EDTA (TAE) Stock Concentration - 50X Working concentration - 0.5X	For 1000ml – 242g of Tris-base, 57.1 ml of glacial acetic acid and 100ml of EDTA (pH:8) were dissolved in distilled water and made into 1000ml
Protein Electrophoresis	
Tris glycine SDS Stock Concentration – 10X Working Concentration – 1X	For 600ml – 18.2g of Tris-base, 86.5g of glycine and 6g of SDS were dissolved in distilled water and made into 600ml
Stacking gel -5% (For 5ml)	H2O - 3.4ml 30% acrylamide mix - 0.83 ml 1M Tris (PH 6.8) - 0.63 ml 10% SDS - 0.05 ml 10% APS - 0.05 ml TEMED -0.005 ml
Resolving gel -12% (For 10ml)	H2O - 3.3 ml 30% acrylamide mix - 4.0 ml 1.5M Tris (PH 8.8) - 2.5 ml 10% SDS - 0.1 ml 10% APS - 0.1 ml TEMED - 0.004 ml

Growth media and composition:

Growth Medium	Composition
Luria Bertani (LB) media	For 1000ml – 10g of tryptone, 5g of

	yeast extract and 10g of sodium chloride were dissolved in 1000ml of distilled water.
Luria Bertani (LB) agar	For 100ml – 1.0g of tryptone, 0.5g of yeast extract, 1.5g of agar and 1.0g of sodium chloride were dissolved in 100ml of distilled water.
All the growth media were sterilized by autoclaving at 120°C at 1.5bars for 20 min.	

#### Reagents for DNA and protein analysis

Reagent	Composition
DNA loading dye (6X)	For 10ml – 3ml of glycerol, 25mg of bromophenol blue and 25mg of Xylene cyanol were dissolved in dissolved in 10ml of distilled water.
SDS gel loading dye (6X)	For 100ml - 5.91g of Tris-HCl, 6g of SDS, 50ml of 100% glycerol 9ml of 14.7M $\beta$ -mercaptoethanol and 30mg of bromophenol blue were dissolved in 100ml of distilled water.
30% Acrylamide mix	For 100ml- 29g of acrylamide and 1g of Bis-acrylamide were dissolved in distilled water.
10% APS	For 10ml- 1g of Ammonium per sulfate dissolved in 10 ml of distilled water.
Staining solution for SDS-polyacrylamide gels	For 500ml- 1.25 g Coomassie R-250,225 mL methanol, and 50 mL glacial acetic acid were dissolved in distilled water

	and made the solution into 500ml.
Destaining solution for SDS-polyacrylamide gels	For 100ml – 40ml methanol, 10ml glacial acetic acid and 50 ml distilled water were mixed and used for SDS gel.

Reagents for protein purification:

Reagent	Composition
Ni-NTA affinity-based purification	
Lysis Buffer	25mM Tris base (pH-7.5) 700mM NaCl 3mM $\beta$ -mercaptoethanol 10mM Imidazole
Elution Buffer	25mM Tris base (pH-7.5) 700mM NaCl 3mM $\beta$ -mercaptoethanol 1M Imidazole
Ion exchange Purification	
Buffer A	15mM Tris base (pH-7.5) 100mM NaCl 3mM DTT
Buffer B	15mM Tris base (pH-7.5) 1M NaCl 3mM DTT
Gel filtration Purification	
Elution Buffer	15 mM Tris-HCl, pH 7.5 50 mM NaCl 3 mM DTT

List of Primer sequences:

Primer Name	Sequence 5' to 3'	Length	Tm
hUHRF1- 140-NdeI	CAACATATGTGGGATGAGACGGAATTGGG	29	60.5
hUHRF1- 427-NdeI	CAACATATGCCGTCCAACCACTACGGAC	28	61.6
hUHRF1- 295-BamHI	CAAGGATCCCTACCGCTCAATCTTGAAGAC TTCGT	35	62.9
hUHRF1- 380-BamHI	CAAGGATCCCTAGGCATCATTCCGGCACTC AG	32	62.5
hUHRF1- 630-BamHI	CAAGGATCCCTAGTTGGCCAGGGCTTCCA GG	31	63.6

## List of publications

1. Rajakumara E., Nakarakanti N.K.\*, Nivya MA\* & Satish M. (2016); Mechanistic insights into the recognition of 5-methylcytosine oxidation derivatives by the SUVH5 SRA domain. *Nature Scientific Reports* 6, 20161 (DOI: 10.1038/srep20161). \*Equal contribution
2. Satish M.\*, Nivya M.A.\*, Abhishek S., Nakarakanti N.K., Shivani D., M. Vani V., & Rajakumara E. (2018); Computational characterization of substrate and product specificities, and functionality of S-adenosylmethionine binding pocket in histone lysine methyltransferases from Arabidopsis, rice and maize. *Proteins* 86(1), 21-34 (DOI: 10.1002/prot.25399). \*Equal contribution
3. Abhishek S.\*, Nivya M.A.\*, Nakarakanti N.K.\*, Deeksha W., Khosla S. & Rajakumara, E. (2018); Biochemical and dynamic basis for combinatorial recognition of H3R2K9me2 by dual domains of UHRF1. *Biochimie* 149, 105-114 (DOI: 10.1016/j.biochi.2018.04.010). \*Equal contribution