ALLELE-SPECIFIC ENGINEERING OF METHYLYSINE WRITERS AND READERS FOR CONTROLLING CHROMATIN-DEPENDENT PROCESSES

by

Simran Arora

BS-MS Dual Degree, Indian Institute of Science Education and Research, 2014

Submitted to the Graduate Faculty of

The Dietrich School of Arts and Science in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Simran Arora

It was defended on August 20, 2019 and approved by

Prof. Dr. Stephen G. Weber, Professor, Department of ChemistryDr. Seth Horne, Associate Professor, Department ChemistryDr. Kabirul Islam, Assistant Professor, Department of ChemistryDr. Andrea Berman, Associate Professor, Department of Biology

Thesis Advisor: Dr. Kabirul Islam, Assistant Professor, Department of Chemistry

ALLELE-SPECIFIC ENGINEERING OF METHYLLYSINE WRITERS AND READERS FOR CONTROLLING CHROMATIN-DEPENDENT PROCESSES

Simran Arora, PhD

University of Pittsburgh, 2019

One of the key players in regulating the gene pattern is the post-translational modifications (PTMs) of histone proteins. Histone modifications regulate the transcriptional potential of genes by interacting with reader/effector protein domains. Post-translational modifications on methyllysine are ubiquitous in biological systems and critical for mammalian development. Specific perturbation of such interactions has remained a challenging endeavor. We hypothesized that incorporation of an unnatural modification with the aid of an engineered writer domain and its recognition by reader domain would regulate the downstream genes (epigenetic editing) leading to modification of the epigenetic landscape. The engineered orthogonal pairs together with catalytically inactive Cas9 would specifically modulate the expression of a gene of interest, thereby providing control on transcription machinery. We employed the allele-specific strategy towards engineering the epigenetic landscape and protein-protein interface orthogonal to the human proteome. We generated a hole-modified methyltransferase (writer) that would install an aryllysine moiety on histones in-cellulo. We established the orthogonality of the engineered system, overcame the permeability issue of SAM analogues, developed an antibody and established the applicability of the system in cells. Our data confirms successful benzylation of histone proteins in mammalian cells at sites known to be regulated by SUV39h2 (writer protein)

in cellulo. Further we engineered a chromodomain (reader) with a pocket to accommodate the bulky modifications. We established the biochemical integrity of the engineered interface, provided structural evidence for domain integrity, demonstrated the generality of the approach, and validated its applicability to identity transcriptional regulators. We have shown that the orthogonal reader domain on binding to the unnatural modification remains functionally intact.

The interactions of reader proteins with its binding partners are transient, weak and cell-cycle dependent thereby challenging to identify. We applied the interactome-based protein-profiling (IBPP) approach to the chromodomain *in cellulo* to identify its native binding partners. We confirmed established biochemical integrity of the mutant proteins, established their crosslinking efficiency *in vitro*, crosslinked them to their native binding partners *in vivo* and pulled them down. On LCMS/MS data validation, we envision translating this approach to other chromodomain containing proteins and identifying their binding partners.

TABLE OF CONTENTS

TAI	BLE OF	CONTENTS v
LIS	T OF TA	ABLES xi
LIS	T OF FI	GURES xii
PRI	EFACE	xvi
1.	INTRO	DUCTION TO EPIGENETIC REGULATORS 1
	1.1. (GENETICS AND EPIGENETICS1
	1.2. I	DNA MODIFICATION AS AN EPIGENETIC REGULATOR 2
	1.3. I	HISTONE MODIFICATIONS AS EPIGENETIC REGULATORS 4
	1.4. I	PTM- H3K9me3 IN EPIGENETIC REGULATION8
	1.5. I	EPIGENETIC REGULATORS IN DISEASES 11
	1.6.	APPROACHES TO STUDY THE PROTEINS AND THEIR LIMITATIONS
	•	
	1.7.	APPROACHES TO STUDY MEMBER SPECIFIC FUNCTIONS OF
	PROTI	EINS
2.	ENGIN	EERING METHYLLYSINE WRITERS TO CONTROL
CH	ROMAT	TIN-DEPENDENT PROCESSES17
	2.1. I	INTRODUCTION 17
	2.1	1.1. Role in Development and Diseases 19
	2.1	1.2. Efforts Towards Studying the Role of SUV39h2 19

	2.1.3.	Allele	Specific	Inhibitors	to	Study	the	Function	of
	Methylt	ransferas	es	••••••	•••••		•••••	•••••••••••	20
2.2.	RESU	LTS ANI) DISCUSS	ION	•••••		•••••	••••••	24
	2.2.1.	Identify	ing the Gat	ekeeper Sites o	on SU	V39h2	•••••	••••••	24
	2.2.2.	Library	of SAM An	alogues	•••••		•••••	••••••••••	24
	2.2.3.	Activity	of Enginee	red SUV39h2	Towai	rds Aroma	atic SA	M	25
	2.2.4.	Catalyti	c Efficiency	of the Engine	ered S	System	•••••	••••••	27
	2.2.5.	Enginee	ring MAT2	A to Generate	Benzy	yl SAM ir	Cells.	••••••	29
	2.2.6.	Benzyl A	Antibody G	eneration for	In-ce	llulo Dete	ection o	of Benzylated	I H3
		••••••		• • • • • • • • • • • • • • • • • • • •		••••••	•••••	••••••••••	30
	2.2.7.	In-cellul	o Benzylati	on, Detection	by WI	B with Be	nzyl An	tibody	31
	2.2.8.	Validati	on of <i>In-C</i>	ellulo Benzyl	ation	and Pres	ervatio	n of Engine	ered
	System	•••••		••••••		•••••	• • • • • • • • •	••••••	33
	2.2.9.	Establis	hing the Fu	nctional Relev	ance	of Engine	ered Sy	stem	39
	2.2.10.	Studying	g the Access	sibility of Benz	zylated	l Genome	•••••	••••••••••	41
2.3.	CONC	CLUSION	IS AND FU	TURE DIREC	CTION	IS	•••••	••••••••••	42
2.4.	MATE	ERIALS A	AND METH	IODS	•••••		•••••	••••••	43
	2.4.1.	Express	ion and Pur	ification of SU	J V39h	2 and Its	Mutan	ts	43
	2.4.2.	Express	ion and Pur	ification of M	AT2A	and Its N	Iutants	5	44
	2.4.3.	Benzyl S	SAM Synthe	esis	•••••		•••••	••••••••••	46
	2.4.4.	Methylt	ransferase A	Assay	•••••		•••••	••••••••••	47
	2.4.5.	MAT2A	and SUV3	9h2 Coupled A	Assay .		•••••	•••••	53
	2.4.6.	Synthesi	is of e-boc a	-N-Fmoc Lysi	ne		•••••	••••••	53
									vi

	2.4.7.	Synthesis of Benzyl Hcy 54
	2.4.8.	Peptide Purification for Methyltransferase Assay and Ab Generation. 55
	2.4.9.	MAT2A and SUV Mutant Expression Followed by Benzylation In-Cells
	2.4.10.	Nuclear Extraction
	2.4.11.	Western Blot 57
	2.4.12.	Imaging Experiments 58
	2.4.13.	qRT PCR 58
3. E	ENGINEER	ING METHYLLYSINE READERS TO CONTROL
CHRO	OMATIN-D	EPENDENT PROCESSES 60
3	8.1. INTR	ODUCTION
	3.1.1.	Role in Development and Disease
	3.1.2.	Efforts Towards Studying the Role of CBXs
	3.1.3.	Our Approach 63
	3.1.4.	Reported Applications of Bump-Hole to Protein-Protein Interface
	Enginee	ering 64
3	8.2. RESU	LTS AND DISCUSSION
	3.2.1.	Rationale for Generating the Mutants
	3.2.2.	Generation of Peptide Library
	3.2.3.	Screening and Validating the Binding of Bump-Hole Pairs
	3.2.4.	Determining the Structure of Bump-Hole Engineering
	3.2.5.	Generality of CBX1 and H3 Engineering Approach to Other Reader
	Protein	s
		vii

	3.2.6.	Probing the Binding of Engineered System in Biological Environment
	3.2.7.	Determining the Functionality of the Engineered System in Biological
	Enviror	1ment
3.3.	CON	CLUSIONS AND FUTURE DIRECTIONS
3.4.	MAT	ERIALS AND METHODS
	3.4.1.	CBX1, CBX3 and CBX5 Protein Expression and Mutagenesis
	3.4.2.	G9a Protein Expression and Mutagenesis
	3.4.3.	Peptide Purification
	3.4.4.	G9a Methyltransferase Assay91
	3.4.5.	FP Assay
	3.4.6.	ITC Assay
	3.4.7.	Crystal Structure, Protein Expression and Purification
	3.4.8.	Crystallization
	3.4.9.	Biotin Pull Down Experiments95
	3.4.10.	In-vivo Plasmid and Protein Expression
3.5.	FIGU	RES AND TABLES
SIT	E-SPEC	IFIC PHOTOCHEMISTRY ON EPIGENETIC READERS 113
4.1.	INTR	ODUCTION
	4.1.1.	Non-Histone Interacting Partners of Epigenetic Proteins 113
	4.1.2.	Tools Available to Identify the Binding Partners
	4.1.3.	Incorporation of Photo-crosslinkable Unnatural Amino Acid 115

4.

	4.1.4.	Background on Binding Partners of CBX1 Through its Chromodomain
4.2.	RESU	LTS AND DISCUSSION118
	4.2.1.	Rationale for Generating Mutants 118
	4.2.2.	Binding Affinity of Mutants119
	4.2.3.	Photo-crosslinking Efficiency of Mutants with Peptide 121
	4.2.4.	Crosslinking Efficiency with Histone Extracts
	4.2.5.	In Cell Crosslinking of CBX1 Mutant124
4.3.	CONC	CLUSIONS AND FUTURE DIRECTIONS 126
4.4.	MATI	ERIALS AND METHODS127
	4.4.1.	Protein Expression and Mutagenesis127
	4.4.2.	FP Assay 129
	4.4.3.	ITC Assay
	4.4.4.	Photo-crosslinking Experiment with Peptide and In-gel
	Fluores	cence130
	4.4.5.	Histone Extraction (HE)
	4.4.6.	HE Crosslinking132
	4.4.7.	In-Vivo Protein Expression
	4.4.8.	In-Vivo Crosslinking
	4.4.9.	Nuclear Extracts
	4.4.10.	Streptactin Pull Down
	4.4.11.	WB134
4.5.	FIGU	RES AND TABLES 135
		ix

5.	BROADER PERSPECTIVE	141
6.	APPENDIX A1	149
7.	REFERENCES	159

LIST OF TABLES

Table 1. Reader Proteins Decide the Fate of Chromatin	
Table 2. List of Genes Selected for qRT PCR.	
Table 3. LCMS Data.	75
Table 4. List of Interacting Partners of CBX1.	83
Table 5. List of Primers	
Table 6. X-Ray Data Collection and Refinement Statistics	112
Table 7. Fluorescence Polarization Data	120
Table 8. ITC Data.	121
Table 9. List of Primers for Generating TAG mutants on CBX1.	135

LIST OF FIGURES

Figure 1. Composition of Nucleosome.	5
Figure 2. Heterochromatin and Euchromatin Formation.	6
Figure 3. CBX1 and SUV39h2 Leading to PEV.	10
Figure 4. Structure Methyltransferases and their Mechanism.	18
Figure 5. Bump-Hole Approach.	21
Figure 6. Scheme for Generating Modified Histone3 in Mammalian Cells.	23
Figure 7. Establishing Orthogonality on Arylation with Gatekeeper Residues in SUV3	9h2.
	26
Figure 8. Kinetics of Native System and the Engineered System.	28
Figure 9. MAT2A Engineering to Incorporate Benzyl on H3.	29
Figure 10. Determining the Specificity of Antibody.	30
Figure 11. In-cellulo Benzylation.	32
Figure 12. Localization of Benzylation to Heterochromatin Regions in Nucleus.	33
Figure 13. Localization of Benzylation to H3K9me3 Rich Heterochromatin Regions	35
Figure 14. Minimum Time Required for Engineered System to Benzylate the Nucleus.	36
Figure 15. Localization of CBX1 Y26F/F50G in HE293T Cells	38
Figure 16. Gene Regulation by Benzylation on H3.	40
Figure 17. SUV39h2 MALDI Data.	48
Figure 18. SUV39h2-Y372A MALDI.	49
Figure 19. SUV39h2 Y372G MALDI.	50
	xii

Figure 20. MALDI Data for Kinetics of SUV39h2 with SAM.	51
Figure 21. MALDI Data for Kinetics of SUV39h2 Y372G with Benzyl-SAM.	52
Figure 22. Schematic for the Synthesis of Nα-Fmoc-Nε-(benzyl, Boc)-lysine	53
Figure 23. Schematic for the Synthesis of Benzyl Homocysteine.	54
Figure 24. Identification of Binding Site Residues on CBX1	65
Figure 25. G9a Methyltransferase	67
Figure 26. Schematic for Screening the Binding Affinity of CBX1* for H3K9-R	68
Figure 27. Schematic of FP Assay.	69
Figure 28. Screening the CBX1 Mutants Against the Bumped Peptides for Binding.	71
Figure 29. Validation of the Binding Data by ITC.	74
Figure 30. Crystallographic Evidence for the Structural Preservation of Engineered	System.
Figure 31. Transfer of CBX1*-H3K9benzyl to CBX3 and CBX5	80
Figure 32. Pull Down of Native or Engineered System from Nuclear Extracts	
Figure 33. Second Sphere of Interacting Partners Pull-Down.	85
Figure 34. Troubleshooting DNMT1 Pull-Down	
Figure 35. MALDI Spectra.	
Figure 36. MALDI Spectra.	
Figure 37. Dissociation Constant.	100
Figure 38. Dissociation Constant.	101
Figure 39. Dissociation Constants.	102
Figure 40. Dissociation Constant.	103
Figure 41. Dissociation Constant.	104
	xiii

Figure 42. Dissociation Constant.	105
Figure 43. Dissociation Constant.	106
Figure 44. Dissociation Constant.	107
Figure 45. Dissociation Constant.	108
Figure 46. Dissociation Constant.	109
Figure 47. Dissociation Constant.	110
Figure 48. Dissociation Constant.	
Figure 49. Illucidation of IBPP Approach	115
Figure 50. Identification of Novel Binding Partners of CBX1 through Cha	romodomain In-
Vivo.	117
Figure 51. Site Selection and Validation of Mutants.	
Figure 52. Crosslinking with Peptide followed by In-gel Fluorescence	
Figure 53. Crosslinking with Histone Extracts.	
Figure 54. In-Vivo Crosslinking.	126
Figure 55. Dissociation Constants.	
Figure 56. Dissociation Constants.	
Figure 57. Dissociation Constants.	
Figure 58. Dissociation Constants.	139
Figure 59. Dissociation Constants.	140
Figure 60. Schematic for Epigenetic Editing.	
Figure 61. Illustration of Recognition and Benzylation of Nucleosomes b	y SUV39h2 and
SUV39h2*	
Figure 62. Methyltransferase Assay of Mutants.	150
	xiv

Figure 63. Methyltransferase Assay of Mutants.	151
Figure 64. In Gel Fluorescence Crosslinking Assay.	152
Figure 65. Crosslinking Efficiency of Mutants Detected by Western Blot	153
Figure 66. Methyltransferase Activity of Mutants Generated with C-strep.	154
Figure 67. Crosslinking Efficiency of Mutants with Strep-Tag Insertion	155
Figure 68. Crosslinking Efficiency of Enzymes with Histone Extracts	157
Figure 69. Crosslinking Efficiency of Mutants to Full Length H3.	158

PREFACE

I would like to express my thanks to my research advisor, Dr. Kabirul Islam, for giving me the opportunity to perform my PhD under his supervision. Working in his lab under his supervision exposed me to the field of epigenetics and chemical biology. It is because of his belief in me that, me coming from a computational biophysics background as my masters' project, I am now able to express and purify proteins in bacterial and mammalian cells, I'm able to setup biochemical assays and able to perform techniques like western blot and qRT PCR. I got the opportunity to learn all these techniques in his lab.

I offer my sincere gratitude to my committee members, Prof. Dr. Steve Weber, Prof. Dr. Seth Horne, and Prof. Dr. Andrea Berman for their guidance and support throughout my PhD as well as their time and effort to read my thesis. I would like to thank Prof. Horne and Prof. Berman, for exposing me to the field of protein crystallography.

I acknowledge the support of past and current Islam lab members. I would specially, like to thank Dr. Babu Sudhamalla and Dr. Megan Breski, past lab members, for training me in basic biochemical techniques and protein purification. I also appreciate the support and brainstorming discussions with Valerie Snyder, Sam Kavoosi, Shana Wagner and Sushma Sappa in the lab.

Our alumni, from Dieters Lab, Dr. Yuta Naro and Dr. Nicholas Ankenbruck have always brainstormed in my struggle with molecular biology and cell biology techniques, specially when I started working with mammalian cells.

Lastly, I would like to thank my family, for their tremendous support and encouragement that has kept me motivated in my journey to graduate studies. I wish to acknowledge the support of my husband, Shivendra Tripathi, for understanding my aspirations and supporting me to fulfil them.

I am blessed to have met so many inspiring people to keep me determined and focused to reach my goals. I thank all of them.

1. INTRODUCTION TO EPIGENETIC REGULATORS

1.1. GENETICS AND EPIGENETICS

Genetics is the study of genes, heredity and genetic variations. Gregor Mendel's work on crosspollination between variants of peas gave birth to the concept of genes or traits of inheritance in the middle of the 19th century³. A gene is defined as a string of nucleotides on DNA that codes for functionally relevant biomolecules (protein/ RNA). In 1953 James D. Watson and Francis Crick explained the concept of inheritance of genes, by proposing the double helix structure of DNA⁴. Their model proposed that the two strands of DNA were complementary and therefore serve as independent templates leading to self-replication. This laid the basis for the transfer of genetic information from one generation to the next. There have been several discoveries in the field following their work leading to the discovery of codons for each amino acid⁵, DNA replication, transcription of genes to mRNA⁶ and many more. Genetic mechanisms and genetic factors, such as transcription factors, provide primary control over gene regulation, organismal development and cellular differentiation and genetic mutations have been linked to abnormal development and diseases⁷.

Genetics itself is insufficient to explain the onset of disease in adults due to environmental factors that do not include any DNA mutations. It is limited in explaining the plasticity of cells to respond to environmental factors such as nutrition, environmental compounds and stress⁸. It also does not

explain the mitotically or meiotically stable inheritance of genes independent of DNA sequence^{9,10}. These factors led to the discovery of the field of epigenetics.

The term epigenetics was first coined in 1940 by Conrad Waddington, a developmental biologist who studied the effect of environment-gene interactions on phenotypes ¹¹. Epigenetics spans molecular, cellular and environmental processes contributing to heredity of genes that do not involve an alteration in the DNA sequence. There are approximately 208 different types of cells in humans that differentiate from a single zygotic cell during development, carrying the same DNA¹². These cell types maintain their phenotypic identity and transfer their traits to the next generation almost identically. Epigenetic modifications are important for maintaining cellular integrity and aberrant epigenetic modifications lead to heritable diseases (imprinting disorders) and non-heritable diseases (most cancers). At the molecular level, epigenetic processes involve methylation of DNA¹³ and myriad of post-translational modifications (PTMs) on RNA¹⁴⁻¹⁶ as well as histone proteins^{17,18}.

1.2. DNA MODIFICATION AS AN EPIGENETIC REGULATOR

In the 1970s the first epigenetic molecular modification to be identified was DNA methylation, on 5th carbon of cytosine, leading to X-chromosome inactivation. 5-methylcytosine (5mC) is found at symmetrical CpG dinucleotides with an abundance of ~1% of all DNA bases¹⁹. It is a highly conserved epigenetic modification found in plants, animals and fungal models. There are broadly three types of epigenetic regulators for DNA methylation: writers (that incorporate the

modification), readers (that bind to the modification and recruit other factors) and erasers (that remove the modification).

In somatic cells methylated, CpG islands have been shown to be correlated to the repression of genes. In mammals, DNA methyltransferase (DNMT), comprise of family of writer enzymes, incorporates methylation on DNA strands. During mitosis, the hemimethylated DNA strands are recognized and methylated by DNMT1 and the ubiquitin-like plant homeodomain and RING finger domain 1 (UHRF1) reader proteins, to stably maintain the methylation pattern at CpG sites in daughter cells²⁰. Stable maintenance of 5mC regulates stable repression of genes and genomic imprinting.

Contrary to somatic cells, global demethylation is observed during early embryogenesis for setting up pluripotent states²¹. The ten eleven translocation (TET) family of enzymes, eraser enzymes, are involved in rapid demethylation of 5mC by oxidizing it to 5-hydroxymethylcytosine ¹⁹ (5hmC) and further to 5-formylcytosines (5fC) and 5-carboxycytosines (5caC) ²² that is converted to cytosine by a decarboxylase. Thus, TET enzymes, with specific recognition domains (example the CXXC domain) and along with reader proteins, can precisely regulate 5mC levels, thereby regulating the expression level of genes ²³. DNA methylation followed by demethylation is a reversible epigenetic process that regulates the expression of genes without altering the underlying information in the DNA.

1.3. HISTONE MODIFICATIONS AS EPIGENETIC REGULATORS

All eukaryotic cells package their DNA tightly into chromosomes built from nucleosomes, composed of four histone proteins around which 147 bps of DNA is wrapped²⁴. Each nucleosome is an octamer of two each of H2A, H2B, H3 and H4 histone proteins²⁵, as shown in figure 1. An array of nucleosomes connected through the DNA molecule and linker histones (H1) is termed chromatin. Chromatin is present in one of the two states: euchromatin-state where the gene is readily available for transcription as the nucleosomes are loosely packed or the heterochromatin state where the gene is silenced as the nucleosomes are in a condensed form²⁶. The chromatin state of different genes is an important factor in defining cell types in multicellular organisms.

Histones are basic proteins²⁷ that have disordered N-terminal regions protruding from the



Figure 1. Composition of Nucleosome. DNA (blue) is wrapped around the octamer of histone (orange), each dimer of H2A, H2B, H3 and H4 proteins. The histone tails protruding nucleosome are modified with PTMs.

nucleosomes. The covalent modifications on the amino acid residues of histones regulate the contact of histones with the underlying DNA, as well as other factors regulating the gene expression or compaction. These covalent modifications or post-translational modifications (PTMs) on histones include: acetylation, methylation, phosphorylation, ubiquitination, SUMOylation and glycosylation of lysine, arginine, serine, threonine, tyrosine, histidine and glutamic acid on histone tails²⁸. Methylation and acetylation of histones are the most well studied PTMs that occur at the lysine and arginine residues of the histone tails. All of these modifications are known as the histone code ²⁹.

The PTMs on histone tails effectively regulate the chromatin state and thereby cellular processes by acting as signals for gene regulation. Gene regulation via PTMs on histone tails is brought about by three types of epigenetic proteins namely writer, reader and eraser proteins, illustrated in figure 2. The writer and the eraser proteins add and remove the modification from the histone tails, respectively³⁰. The reader identifies and binds to a specific PTM and represses/activates genes through interacting proteins.



Figure 2. Heterochromatin and Euchromatin Formation. PTM H3K4me3 is shown in green, is incorporated by writer protein (MLL) is recognized by reader proteins (BPTF-PHD) that recruits downstream effector protein (NuRD) to form the euchromatin for gene transcription. On removal of these PTMs by eraser proteins (PHF8), reader and effector proteins corresponding to it form the heterochromatin again.

Histone acetylation was first identified in the 1960s its link to DNA transcription was established in 1997 when the activity of a histone acetyltransferase (HAT) writer enzyme in *Tetrahymena* *thermophilia* was discovered. There are five different families of HATs with conserved structural similarity of the enzymatic core³¹. An acetyl group directly affects the charge state of the histone, neutralizing the charge on the lysine residue, impairs the binding between negatively charged DNA and the nucleosome composed of basic histones. An acetyl group effectively loosens the DNA wrapped around nucleosomes making it available for transcription, thus regulating the underlying gene. Histone deacetylases (HDACs), eraser enzymes, remove the acetyl group from the lysine residues, thereby restoring the positive charge on lysine groups and condensing the DNA. The bromodomain reader protein family of enzymes recognizes the acetyl-group on the lysine residues and interacts with transcription machinery to regulate the transcription of genes. Thus, histone acetylation leads to the activation of genes with the assistance of epigenetic regulators without altering the underlying DNA sequence.

SUV39h1, writer enzyme, was the first methyltransferase identified in *Drosophila melanogaster*^{32,33}. All methyltransferases are either members of SET domain-containing enzymes or non-SET containing enzymes. This modification is removed by demethylases, like LSD1 first identified in 2004³⁴. Family of demethylases include Jumonji C family, JHMD1, JMJD3, JMJD2D that target specific methyl-lysine groups. Histone methylation on lysine and arginine residues occurs to various degrees and indirectly regulates DNA transcription, as methylation does not change the charged state of the lysine residues. Thereby it is linked to both activation and repression of genes based on site and degree of methylation²⁸. Often the euchromatin sites are enriched in H3K4me3 and heterochromatin sites in H3K9me3 and H3K27me3, although during development both H3K4me3 and H3K27me3 are present on the promoters of the select genes. This indicates that it is not the modification that regulates activation or repression of genes, rather

it is the effector proteins (reader proteins) that bind these methyl marks and recruit downstream enzymes. Table 1 lists different levels of methylation on H3K4 groups leading to activation or repression and associated reader proteins without changing the genetic information.

Position of Lysine	Degree of Methylation	Reader Protein	Activation or Repression of Genes
Н3К4	Me3	BPTF-PHD	Activation
НЗК4	Me3	Double chromodomain	Activation
НЗК4	Me3	ING2 PHD	Repression
НЗК4	Me0	BHC80-PHD	Repression

Table 1. Reader Proteins Decide the Fate of Chromatin. Illustration of how methylation on

 H3K4 when bound to different reader protein decides the fate of gene being regulated by it

 differently.

1.4. PTM- H3K9me3 IN EPIGENETIC REGULATION

Establishment and maintenance of cellular identity are dependent on the expression of euchromatic genes as well as repression of heterochromatic genes. The mammalian genome consists of repeatrich sequences at centromeres and telomeres that may potentially lead to recombination³⁵. The cells regulate these regions by packaging them into heterochromatin regions, known as constitutive heterochromatin, remain silent across developmental lineage³⁶. The regions that dynamically form heterochromatin during development are called facultative heterochromatin regions³⁷. The PTM for constitutive heterochromatin region is H3K9me3^{38,39} and that for facultative heterochromatin region is H3K27me3⁴⁰. H3K9me3 represses facultative heterochromatin majorly, it has also been shown to suppress genes in a cell-type specific manner. Apart from constitutive heterochromatin H3K9me3 regulates expression of zinc finger transcription factors, olfactory receptors, and neurotransmitter-related genes³⁵.

The writer, reader and eraser proteins of H3K9me3 are SUV39, HP1 and KDM4. In humans, SUV39h2 (Suppressor of variegation 3-9 homologue 2) is primarily responsible for trimethylation on histone 3 at lysine 9. It has four major domains: chromodomain, pre-SET, SET and post-SET domain. The SET catalytic domain transfers the methyl group from SAM (S-adenosyl methionine) to the H3K9. Over-expression of SUV39h is implicated in several cancers including leukemia, lymphomas, lung cancer, breast cancer and colorectal cancer⁴¹. The first HP1 (heterochromatin protein 1) protein to be identified was HP1a in Drosophila melanogaster localized at the heterochromatin region with H3K9me3 mark. HP1 proteins were first identified as key players in heterochromatin-mediated gene silencing and have recently been recognized to be involved in several other processes including gene activation. CBX proteins are found in mammals and those belonging to the HP1 family are CBX1, CBX3 and CBX5. dHP1 based family of proteins contain three structural domains: N-terminal chromodomain (CD), a flexible linker region and a Cterminal chromoshadow domain (CSD)⁴². These proteins localize to the histones with the H3K9me3 modification through CD and they interact with the other complexes via CSD. The eraser protein KDM (lysine demethylase) belong to JmjC (Jumonji C) that employs α-ketoglutarate $(\alpha$ -KG), Fe (II) and oxygen as cofactors to catalytically remove the methyl groups from lysine residues. In humans, KDM4 (A-E) demethylate H3K9me3 along with other targets on non-histone proteins as well⁴³. Studies have established *KDM4* gene amplification and overexpression in tumors, including lung, breast, esophageal, prostate cancers and lymphoma⁴⁴.

The mechanism of heterochromatin formation by SUV39h and HP1 proteins is proposed via position effect variegation (PEV)⁴⁵ (Figure 3). PEV is the phenomenon where the repressive mark spreads thousands of kilobases from the site of nucleation on the euchromatic region. Spreading is initiated after the first mark is read by the HP1 protein that recruits Suv39h (a methyltransferase enzyme) that marks the nucleosome with H3K9me3. By dimerizing with the neighboring HP1 the nucleosomes to form a condensed chromatin state, silencing the gene for transcription.



Figure 3. CBX1 and SUV39h2 Leading to PEV. Top panel illustrates the cartoon for CBX1 and SUV39h2. The bottom panel illustrates the heterochromatin formation by SUV39h2 and CBX1.

H3K9me3 is a well-studied PTM on histones resulting in heterochromatin formation and regulating developmentally important genes and silencing repeat units at telomere and centromere of chromosomes. The mechanism of formation of heterochromatin is well studied and understood. Dysregulation of this mark on epigenetic landscape is implicated in several diseases. Therefore, H3K9me3 serves as a good system to test and generate new tools for understanding and studying the pathways regulated by specific PTM.

1.5. EPIGENETIC REGULATORS IN DISEASES

Aberrant changes in DNA methylation pattern are linked to several cancers. Studies have shown a link between changes in methylation pattern and inactivation of tumor suppressor genes and activation of oncogenes. This observation has been linked to decreased 5hmC levels thereby to TET's expression and activity²⁰.

The writer, reader and eraser proteins may play a direct or indirect role in several diseases like cancer, metabolic diseases, inflammation, and neuropsychiatric disorders. These proteins have been shown to drive the diseased state by two primary mechanisms. First, a mutation or altered expression of these proteins leads to a change in PTMs and an alteration in gene expression. For example, in several cancerous tissues N-lysine methyltransferase EZH2, writer enzyme, is overexpressed. EZH2 methylates lysine 27 of H3 and represses the differentiated genes in organisms. Studies show the over-expression of EZH2 in breast cancer and prostate cancer cell but a mechanistic basis is not well established yet⁴⁶. Second, reader, writer and eraser proteins can alter gene regulation by their interaction with upstream/downstream signaling factors. For example, the reader protein BRD4 that recognizes acetylated H4-lysine16, is translocated with testis-specific transcription factor NUT that drives carcinogenesis. A selective inhibitor of the BET (bromo and extra-terminal) domain, selectively killed the BRD4-NUT midline carcinoma xenograft. A more direct example is the interaction of BRD4 (bromodomain containing protein) with MYC (a transcription factor) that is activated in several cancers. BRD4 binds acetylated histones and regulates the transcription of MYC target loci. Inhibition of the bromodomain of BRD4 results in reduced activation of MYC as well as the transcription of *MYC* gene. An alteration of the PTM landscape on histone tails is the driving factor for many disease states.

It is evident that epigenetics is associated with several diseases- cancer, inflammation, diabetes and neuropsychiatric disorders³⁰. Therefore, techniques are required to probe and identify the targets of these epigenetic proteins for understanding the mechanisms underlying the diseased states as well as developing selective drugs/ therapies to cure them.

1.6. APPROACHES TO STUDY THE PROTEINS AND THEIR LIMITATIONS

Radioactive isotope labeling, chemical derivatization, antibody based pull down, and MSproteomics are a few techniques that enrich the PTMs. However, these techniques do not identify the enzymes responsible for incorporation of these PTMs and regulation of specific genes at certain stages of development of cells or responsible for certain diseased state are few⁴⁷.

There are a few broad approaches to study specific protein functions. The first approach is classical genetics, where the gene for the protein of interest (POI) is knocked out/down or mutated at the binding site such that it gains or loses its function. The downstream phenotypic changes suggest genes or pathways being regulated by POI. There are two major limitations to this approach first, it is difficult to assess whether the downstream effect observed is due to a modification at the catalytic/ binding site or some other structure-function of the protein and second, it takes hours/days for gene knockout/knockdown that makes study of these dynamic processes occurring in seconds to minutes (e.g. PTMs, transcription) challenging ^{48 49 50 51}.

The second approach is chemical genetics, where a small molecule (inhibitor or activator) binds a protein to modulate its activity thus facilitating downstream phenotypic changes. This approach provides temporal control but often lacks in specificity as the small molecule may potentially bind closely related member of a superfamily. More importantly, developing inhibitors for dynamic protein-protein interactions (PPIs) such as the interaction between CBX1 and H3K9me3 is more challenging than for enzymes with a well-defined small-molecule binding pocket ^{52 53}.

Chromatin Immuno-Precipitation followed by sequencing (ChIP-seq) is a technique that provides direct information of the binding partners of epigenetic proteins. It involves crosslinking between proteins by formaldehyde, immuno-precipitating POI and analyzing the binding partners by sequencing or proteomics. This technique lacks specificity as it crosslinks to any protein in its vicinity. Hence individually these approaches have limitations to probe the member-specific functions of reader proteins in highly dynamic biological processes.

1.7. APPROACHES TO STUDY MEMBER SPECIFIC FUNCTIONS OF PROTEINS

To elucidate the function of specific PTM (H3K9me3) and regulatory proteins, it is important to combine both genetics and small molecule approach to bring in spatial and temporal control. The toolkit involves modulating the genetics (specificity) of POI to express an engineered protein that becomes active only in the presence of temporally controlled stimuli like light or small-molecule. There are two parallel protein-protein interface engineering techniques that provide both spatial and temporal control. The first technique involves generating an extended pocket (hole) in reader protein via site-directed mutagenesis and incorporating a bulky (bump) modification on the histone 13

tail, commonly known as allele specific chemical genetics or the bump-hole method. The engineered pair interacts in a bio-orthogonal fashion in cells. The bump-hole technique enables us to regulate protein function by incorporating the bumped histone at the gene of interest (GOI). This technique was first successfully employed in 1987 by Miller on GTP-binding site of a GTPase⁵⁴. There are several examples to date for successful implementation of this technique on myriad systems like kinases⁵⁵, GPCRs⁵⁶, myosins⁵⁷, kinesins⁵⁸, methyltransferase, acetyltransferase⁵⁹, ADP ribosyltransferase⁶⁰, and demethylases⁶¹. The biggest limitation of this technique is in achieving orthogonality to the native system.

The second technique is photo-crosslinking of POI with an unnatural photo-crosslinkable amino acid (UAA) by shining UV light that would covalently bind to all the interacting partners. The specificity comes by incorporating UAA into the binding pocket of the reader/ writer protein. The regulator protein pulled down by IP/ tag would be crosslinked to nucleosomes interacting via the binding pocket of reader domain only. These techniques specifically identify binding partners (histone and non-histone proteins and associated genes) to study the difference in gene regulation pattern established by specific members of the writer and reader proteins. The advantage of this technique is the temporal control and identification of transient protein-protein interaction. There are more than 70 reported UAAs that have been incorporated in different systems for understanding their regulatory, structural and enzymatic roles⁶². Major limitation of this approach is expression (enrichment) of proteins with unnatural amino acid maintaining the structural and functional integrity.

In this thesis I aim to incorporate these techniques to writer and reader proteins that regulate H3K9me3 modification in cells. This thesis is a proof-of-concept that by engineering the binding/

enzymatic pocket of epigenetic regulator proteins (SUV39h2 and CBX1) we can edit the PTM landscape of histones/epigenome that can be employed to study the genes/ proteins being specifically regulated by them relevant to development and diseases.

In **Chapter 2** we have edited the epigenetic landscape by incorporating an unnatural modification on histone tail with the aid of a bio-orthogonal methyltransferase mutant of SUV39h2 (bump-hole approach). This chapter involves engineering the catalytic pocket of SUV39h2 as well as MAT2A (*in cellulo* generation of SAM). Engineering MAT2A overcomes the challenge of cellular permeability of SAM analogues. We have experimentally demonstrated that the bio-orthogonal system is able to successfully incorporate the unnatural modification in HEK293T cells through western blotting, imaging and qRT-PCR data.

In **Chapter 3** we engineered the protein-protein interface binding between the repressor protein CBX1 and its natural binding partner H3K9me3 by applying the bump-hole approach to develop a tool to identify member specific functions of CBX family of proteins. We identified the bioorthogonal pair of mCBX1 (mutant CBX1) and H3K9-R (R stands for unnatural modification) by employing a fluorescence polarization assay. We also confirmed the binding affinity and integrity of the binding pocket by isothermal titration calorimetry (ITC) and solved the crystal structure. We also generated similar mutants for CBX3 and CBX5 and show that they maintain the bioorthogonality of the system. By employing a biotin-peptide pull down assay we show specificity of the system in cellular environment and ability to interact specific transcriptional regulators.

In **Chapter 4** We developed a photo-crosslinkable CBX1 mutant that maintains the integrity of wild-type binding pocket. Here we validate the binding data through ITC experiments. Identified the mutants that strongly crosslink at peptide level, full length histone level, and validated the

crosslinking in the extracted histones from HEK293T cells. For the first time we show successful in-cell crosslinking with the reader proteins.

2. ENGINEERING METHYLLYSINE WRITERS TO CONTROL CHROMATIN-DEPENDENT PROCESSES

2.1. INTRODUCTION

Suppressor of variegation 3-9 homolog 2 (SUV39h2, also known as KMT1B) is a member of the SUV39 subfamily that consists of proteins, namely SUV39H1 (KMT1A), SUV39H2 (KMT1B), SETDB1 (KMT1E), SETDB2 (KMT1F), G9A (EHMT2) and G9a-like protein (GLP1), of lysine methyltransferases ⁴¹. *Suv(var)* genes were first identified in *Drosophila melanogaster*, in 1999 by Jenuwein *et al*, to suppress the position effect variegation (PEV) and thereby repress the genes. Later, the group identified the activity of SUV39h1 and SUV39h2 to transfer the methyl group from SAM to H3⁶³. SUV39h2 di- and tri-methylates H3K9, that is responsible for heterochromatin organization, transcriptional repression, and epigenetic silencing. All the SUV39 family members have a SAM binding domain and the target protein binding domain. Following a S_N2 mechanism the methyl group is directly transferred to target protein, figure 4a.

SUV39H2 consists of an N-terminal chromodomain that recognizes and binds to H3K9me3 and the C-terminal catalytic methyltransferase (SET domain). In 2010 Schapira *et al*, solved the structure of four enzymes that methylate H3K9 namely GLP, G9a, SUV39h2 and PRDM2 in their apo and complex form with the peptide and cofactor SAM. G9a, GLP and SUV39h2 have common

domains and are structurally similar. They consist of N-SET, Pre-SET, SET, I-SET, Post-SET and C-SET. Based on their structures it is concluded that firstly the peptide binding groove of the enzymes is negatively charged, secondly the I-SET domain consists of sites responsible for selective binding of the substrate, and finally the post-SET domain closes to form the catalytically



Figure 4. Structure Methyltransferases and their Mechanism. Fig a. Illustrates $S_N 2$ reaction mechanism for methyltransferase on H3 from SAM b. Illustrates favorable $S_N 2$ formation for b-unsaturated SAM analogues. Fig a and b are adapted from ref⁻¹. Fig c. and d. show the domain structure of overall conformation of G9a and SUV39h2 in their apo form (adapted from ref⁻²) respectively.

active conformation. Figure 4c and 4d show the structure of G9a and SUV39h2 in apo and complex forms displaying the structural similarity of the two enzymes ².

2.1.1. Role in Development and Diseases

SUV39h2 plays an important role in development as it is shown to be broadly expressed during mouse fetal development until day seventeen via RNA blot experiments. In adult mouse tissues it is significantly upregulated in testes tissues and significantly downregulated in ovaries compared to other tissues ⁶³. It also plays a crucial role in DNA damage repair (DDR) pathway and regulation of the cell cycle. It regulates the H2AX, known to be involved in DDR upon phosphorylation. SUV39h2 in association with the retinoblastoma tumor suppressor protein (pRb) represses cyclin E promoter which is involved in cell cycle regulation ⁶⁴. Consequently, dysregulation of SUV39h2 has been reported to be carcinogenic and its overexpression is reported in cancer tissues such as leukemia, lymphomas, lung cancer, breast cancer, colorectal cancer, gastric cancer, and hepatocellular cancer ⁴¹. SUV39h2 is considered a target for anti-cancer drug development and its regulatory mechanisms are not well understood.

2.1.2. Efforts Towards Studying the Role of SUV39h2

Reversibility of histone methylation makes it challenging to study the pathways for its involvement in diseases. There are only two well characterized non-histone binding partners and pathways elucidated for SUV39h2, LSD1 and H2AX. Methylation on both proteins was discovered by invitro methylation assay by various methyltransferases^{65,66}. LSD1 is a demethylase that acts on mono- and dimethylated H3K9. It is known to be over-expressed in several cancers. In 2015 L.
Piao *et al* showed that SUV39h2 tri-methylates LSD1 at lysine 322 stabilizing LSD1 suppressing the ubquitinylation at this site for LSD1 degradation and leading to cancer⁶⁶. In the study by K.Sone *et al*, SUV39h2 was shown to methylate H2AX at lysine 134, for H2AX to actively repair DNA as well as knockdown of SUV39h2 impaired the ability of H2AX to repair DNA⁶⁵.

Since all the methyltransferases that belong to SUV39 family have a similar binding pocket and similar mechanism, developing inhibitors specific towards a methyltransferase is challenging. Still there are inhibitors that have a higher potency towards one of the methyltransferase compared to the others example: pinometostat for DOT1L, BIX01294 for G9a/GLP and A-395 for PRC2 complex⁶⁷. There is a need to develop tools to study the function of SUV39h2 in an orthogonal fashion to deconvolute its function from the family of SUV39 methyltransferases. We plan to apply bump-hole or allele specific chemical genetics to decipher specific role of SUV39h2 in transcription.

2.1.3. Allele Specific Inhibitors to Study the Function of Methyltransferases

The Allele Specific Chemical Genetics (ASCG) approach also known as the Bump and Hole approach involves engineering at the interface of protein and ligand to study the role of a protein. It involves generating an extended pocket (hole) in the active site of the protein through site directed mutagenesis. The involvement of genetics approach brings specificity mutating only the protein of interest (POI), rendering it inactive towards its natural ligand or binding partners. The modified protein is expected to have low affinity for natural ligand but a high affinity for its complementary (bumped) modified ligand and vice versa for unmodified protein as shown in Figure 5. This makes the bump-hole system orthogonal to the wild type system. The temporally 20

controlled orthogonal system can be used to study the function of a designated protein belonging to large superfamily.



The first application of this method to a methyltransferase was in 2001 on yeast PRMT1 (Rmt1). Here in the enzyme pocket E117 was mutated to glycine generating a hole, and bulky (or bumped) N6-benzyl SAM was synthesized to fit in the hole. The E117G mutant was 500-fold less active towards SAM than the wild-type enzyme and 65-fold more active towards the bumped SAM analogue¹. Following this, multiple groups applied this technique on different methyltransferases with a range of SAM analogues. Luo and coworkers developed the tool "Bio-orthogonal Profiling of Protein Methylation" (BPPM), where they engineered the active site of methyltransferase to catalytically transfer the bulky sulfonium group from SAM to its substrate⁶⁸. They discovered that Y1211 for GLP and Y1154 for G9a act as gatekeeper residues that sterically hinder the bulky substituents on SAM. Also, the β -unsaturated SAM-substituents had higher k_{cat} as they stabilized the S_N2 reactions in the transition state shown in figure 4b. The stereo-electronic allele specific

pair led to the identification of novel non-histone partners of G9a and GLP *in-vitro*⁶⁹. A major challenge when applying this technique *in vivo* is the cell permeability of SAM. To overcome this challenge, they engineered MAT (methionine adenosyl transferase) to synthesize SAM-analogues in cells from cell permeable methionine analogue and ATP. They also synthesized Hey-SAM (hex-2-ene-5-yne SAM) with the terminal alkyne group that could be clicked (using Cu-azide alkyne cycloaddition chemistry) to a biotin probe and pulled down for genomic sequencing (CliEN-seq)⁷⁰.

In the study by Bothwell *et al*, they show that native G9a, GLP and SUV39h2 were able to transfer the propargylic group from ProSeAM (propargylic-Se-adenosyl L-selenomethionine is more stable than than the corresponding SAM analogue) to their respective substrates compared to other methyltransferases. Since all three methyltransferases were able to transfer this modification, they were not suitable to independently identify the binding partners of these methyltransferases using ProSeAM.

There is no additional literature on development of a tool to study the partners or genes being regulated by SUV39h2. Here we developed an allele-specific tool to identify the role of SUV39h2 *in-vivo*, a schematic is shown in figure 6.



Figure 6. Scheme for Generating Modified Histone3 in Mammalian Cells. Figure shows the general idea of introducing cell permeable methionine analogue in cells that is then converted to SAM analogue by MAT enzyme. SAM analogues then penetrate nuclear membrane to generate modified H3 by methyltransferase activity of SUV39h2 mutants. In Blue is the cell membrane and in red is the nuclear membrane.

2.2. RESULTS AND DISCUSSION

2.2.1. Identifying the Gatekeeper Sites on SUV39h2

To identify the similar residues as G9a Y1154 (PDB 5JJ0) in SUV39h2 (PDB 2R3A), we superimposed the two, and observed that SUV39h2 Y372 aligns to Y1154 in G9a figure 7. Based on this we generated SUV39h2-H327R, SUV39h2-H327R Y372A and SUV39h2-H327R Y372G. The H327R mutation on SUV39h2 is the same as H320R in SUV39h1 that is reported to be hyperactive with 20-fold more activity ⁷¹. Here onwards SUV39h2 H327R is treated as the wild-type enzyme.

2.2.2. Library of SAM Analogues

As positive controls for the wild-type enzyme we used SAM (1) and Allyl-SAM (2). We hypothesized that an aromatic substitution on SAM would accommodate better in the hydrophobic binding pocket than an alkyl chain, as we are mutating the aromatic residue of the enzyme. Towards this hypothesis we synthesized benzyl-SAM (B-SAM 3). As the synthesis of B-SAM is reported, we characterized the product by HRMS only. The library of SAM analogues tested in shown in figure 7.

2.2.3. Activity of Engineered SUV39h2 Towards Aromatic SAM

To confirm the activity of the mutants, we followed a similar assay condition as reported for G9a⁷². H3 (1-22 amino acids) peptide was obtained from the University of Pittburgh Proteomic Facility. The extent of H3 alkylation was monitored by MALDI-MS.

For all the enzymes the negative control was no enzyme. First, we established the assay for WT and SAM analogues, and then examined for the mutants towards bulky SAM analogues. Figure 7 shows the heat map for the screening. The WT is active towards SAM but inactive towards **3** or **4** benzyl-yne SAM (BY-SAM). SUV39h2 Y372A mutant is still active towards SAM along **3** and **4**. SUV39h2 Y372G is not active towards SAM and active towards B-SAM and BY-SAM. The SUV39h2 Y372G is more bio-orthogonal than the alanine mutation at 372 position. This heat map/ methyltransferase assay confirms our hypothesis of employing an aromatic residue on SAM for complementing the hole modified pocket of methyltransferase better than alkyl chains. Also transfer of bulky benzyl-yne group from SAM to H3-peptide confirms the hydrophobicity of the hole modified pocket and an extended hole in SUV39h2 Y372G/A to accommodate bulkier



Figure 7. Establishing Orthogonality on Arylation with Gatekeeper Residues in SUV39h2. The top left panel shows the superimposition of SUV39h2 (green) and G9a (cyan) binding pocket along with SAM. The top right panel shows the MALDI data for incorporation of benzyl on H3 peptide by SUV39h2 Y372G mutant. The bottom panel shows the heat map for incorporation of the modifications from respective SAM-analogue to H3. The darker the color the better the transfer of analogue from SAM to H3.

2.2.4. Catalytic Efficiency of the Engineered System

The kinetics analysis was performed of WT with 1 and hole modified Y372G with 3 to compare the catalytic efficiency of engineered system with the native system. Variable concentrations of the SAM or SAM-analogue 3 were incubated with the peptide and enzyme and spotted on the MALDI plate at three different time points with CHCA matrix. The percentage modification was evaluated by taking the ratio of area under the MALDI peak for modification by the area under the unmodified peptide plus area under modified peptide MALDI peak multiplied by 100. The % modification as a function of time for native system and for the engineered system is shown in figure 8. The slope for each concentration of SAM or B-SAM was calculated and the Michaelis-Menten curve was plotted for the native system and for the engineered system in figure 8. Here MALDI is not quantitative as we have not measured the ionization potential of benzylated peptide with respect to unmodified peptide. We have assumed similar ionization potential of benzylated peptide as that of unmodified peptide. K_{cat} value for native system is 12.4 min⁻¹ and for the engineered system is 1.3 min⁻¹, as the k_{cat} increases so does the catalytic efficiency, implying that the native system is 10-fold better than the engineered system. Whereas the K_m for native system is 11.7 μ M and for the engineered system is 32.2 μ M. Lower K_m values indicate stronger affinity towards cofactor, this implies the binding affinity of SUV39h2 towards SAM is stronger than the affinity of SUV39h2 Y372G towards benzyl-SAM. Although, benzyl group on SAM fits into the hole modified SUV39h2 and it is active, the activity is less compared to the native system. The bump-hole system for SUV39h2 and aromatic SAM needs more SAR (structure activity relationship) based optimization. Together this data suggests that the native system is faster at



incorporating methyl moiety on H3 than the engineered system at incorporating benzyl moiety on



H3. In cells, based on the sites being benzylated, to study the genes/ pathways being regulated by SUV39h2 we may have lose some important targets of SUV39h2 without knocking out/ knocking down native SUV39h2. Methylation is reversible but benzylation is not we believed that incubating the cells for longer duration of time would overcome the competition between the slow irreversible benzylation reaction to the faster and reversible methylation on H3. To establish proof of concept that the engineered system would work in cells we endeavored towards the cellular permeability of **3**. As discussed above for the permeability of **3** we had to engineer the catalytic pocket of S-adenosylmethionine synthetase (MAT).

2.2.5. Engineering MAT2A to Generate Benzyl SAM in Cells

Methionine and ATP act as substrates for MAT2A to yield SAM and phosphate and pyrophosphate as by products. The binding pocket for the transfer is shown in figure 9, with important residues. Luo *et al* identified residues V121, G120, I117, I322 to generate a hole and add bulky analogues to SAH. With MAT2A I117A mutant they generated Hey-SAM therefore we started with I117A mutant for generation of B-SAM. We obtained the plasmids for bacterial expression of protein from Dr M. Luo for MAT2A and MAT2A I117A. Benzyl-Hcy was synthesized and characterized via HRMS. Upon expression of the proteins we carried out the coupled MAT2A and SUV39h2 assay. The conversion of H3 peptide to H3K9benzyl peptide was low.

We hypothesized that generating an extended pocket for MAT2A by mutating I117 to glycine would improve its activity to generate benzyl SAM. Mutagenesis was performed and confirmed by sequencing. The protein was bacterially expressed and purified to conduct the assay for B-SAM



Adapted from Anal Biochem. 2014 Apr 1;450:11-9.

Figure 9. MAT2A Engineering to Incorporate Benzyl on H3. The left panel shows the binding pocket of MAT2A as well as the mechanism for generation of SAM from methionine and ATP with the potential hole modified neighboring residues. The right panel shows generation of benzylated H3K9 by coupled MAT2A and SUV39h2 assay.

generation. The assay showed that the enzyme efficiently generated B-SAM that was incorporated on H3 peptide as evident from the MALDI readout figure 9. We addressed cellular permeability of B-SAM with engineered MAT2A. The major challenge in employing the engineered systems in cells was the readout for benzylated H3. As antibodies are sensitive to detect nanogram levels of proteins we next pursued to generate an antibody for the same.

2.2.6. Benzyl Antibody Generation for In-cellulo Detection of Benzylated H3



The polyclonal benzyl antibody was obtained from Pocono Rabbit Farms and Laboratory (PRF&L). Benzyl lysine was synthesized that was incorporated in the peptide (H2N-

Figure 10. Determining the Specificity of Antibody. via western blot using anti-H3K9benzyl antibody. No signal was detected for lane 2 and 3 without benzylation on H3. Full-length H3K9-benzyl was generated by SUV39h2 Y372G assay and confirmed by LCMS shown on the right here. H3K9me3 was generated by SUV39h2 assay and similarly confirmed by LCMS (data not shown).

ARTKQTARK(benzyl)STGGKAPRKQLAC-CONH2) by Peptide Synthesis Core Facility at University of Pittsburgh. This peptide was used as an antigen in rabbit (host) for the generation of Ab that was then purified by passing the anti-serum through beads bound to the antigen peptide. The peptide has a terminal cysteine residue for its conjugation to maleimide activated KLH (keyhole limpet hemocyanin) protein. Peptides, small proteins and drugs require conjugation to proteins to generate an immune response. The specificity of the antibody towards benzylation was tested by performing the western blot against H3-full length, H3K9me3 (generated by SUV39h2 assay) and H3K9benzyl (generated by SUV39h2 Y372G assay), figure 10. The benzyl Ab was shown to be specific to H3K9benzyl as confirmed by WB against unmodified nuclear extracts and histone extracts. On confirmation of specificity of the antibody we tested our engineered system in cells.

2.2.7. In-cellulo Benzylation, Detection by WB with Benzyl Antibody

HEK293T cells were transfected with MAT2A I117G mutant and various SUV39h2 mutants with or without benzyl-Hcy. An important aspect of this experiment is the depletion of methionine from the media before addition of the unnatural methionine reagent. This was done by following the protocol by Wang *et al*, by treating the cells with methionine-free media before treating them with benzyl-Hcy⁷⁰. As negative control we used no transfection with and without benzyl-Hcy. The nuclear extracts from the cells were collected to test the benzylation of H3. Western Blot, shown in figure 11, with H3K9-benzyl Ab detected a band at expected molecular weight (15kDa) for H3 only when the cells were transfected with MAT2A I117G and SUV39h2 Y372A or SUV39h2

Y372G and treated with benzyl-Hcy. The negative controls do not have a band at that molecular weight. Negative controls transfected with MAT2A I117G and treated with benzyl-Hcy detected a band at 40kDa. As the cells not treated with benzyl-Hcy do not have this band this implies that a non-histone protein is benzylated by an endogenous methyltransferase in HEK293T cells. This experiment also confirms the specificity of the H3K9-benzyl antibody comparing the lanes from cells treated with and without benzyl-Hcy.





2.2.8. Validation of *In-Cellulo* Benzylation and Preservation of Engineered System

Figure 12. Localization of Benzylation to Heterochromatin Regions in Nucleus. HEK293T cells were transfected with MAT2A I117G and SUV39h2 Y372G followed by treatment with 1mM benzyl Hcy. After 24 hrs the cells were fixed and stained with H3K9me3 or H3K9benzyl antibodies. The figure shows heterochromatin regions in blue, H3K9me3 stained regions in green and H3K9benzyl stained regions in red. The imaging data was collected for more than 20 cells.

Upon confirmation of benzylation in cells by WB, we wanted to validate the result using a different technique and probe the system to determine the minimum duration it takes for the nucleus to be benzylated. We collaborated with Dr. Simon Watkins and Michael Calderon, at the Centre for Biologic Imaging, University of Pittsburgh to validate the data and determine the timespan to benzylate H3 in cells. Imaging visualizes single cell events compared to macroscopic event reflected in western blot. It also provides insight to localization of events (e.g. protein is localized in cytoplasm or nucleus).

We first established the benzylation in HEK293T cells transfected with MAT2A I117G and SUV39h2 Y372G by treating them with and without benzyl-Hcy figure 12. The fixed cells were stained with H3K9-benzyl Ab, H3K9me3 Ab and DAPI (dye that binds to AT rich regions in heterochromatin of the nucleus⁷³). The negative control shows that no signal for H3K9benzyl-Ab but an overlap of H3K9me3 on DAPI rich regions in the nucleus. The positive control shows a signal for benzylation in the nucleus of the cells that is tightly packed also to H3K9me3-Ab. This established the benzylation in cells using imaging-based technique. The heterogeneity in the staining of the cells with benzyl antibody being less than that of the nuclei present (stained with DAPI) may be attributed to the extent of transfection of both the MAT2A I117G plasmid as well

as SUV39h2 Y372G plasmid together in a cell. Next, we carried out experiments with different concentrations of benzyl-Hcy to determine the minimum time required for the nucleus to be benzylated. Here we used 0.5 mM and 2mM compound and fixed cells at 0, 3, 6, 8 and 24 hours. With 0.5mM concentration we saw the signal for antibody starting from 9 hours and with 2 mM



Figure 13. Localization of Benzylation to H3K9me3 Rich Heterochromatin Regions. HEK293T cells were transfected with MAT2A I117G and SUV39h2 Y372G followed by treatment with 1mM benzyl Hcy. After 24 hrs the cells were fixed and stained with H3K9me3 and H3K9benzyl antibodies. The figure shows heterochromatin regions in blue, H3K9me3 stained regions in green and H3K9benzyl stained regions in red. The imaging data was collected for more than 20 replicates.

concentration we saw the signal beginning from 8 hours of adding the benzyl-Hcy figure 13. The preliminary results implied that the MAT2A I117G along with SUV39h2 Y372G requires



minimum 8 hours to actively convert benzyl-Hcy to benzyl SAM that is then transferred to H3.

Figure 14. Minimum Time Required for Engineered System to Benzylate the Nucleus. HEK293T cells were transfected with MAT2A I117G and SUV39h2 Y372G followed by treatment with 0.5 mM or 2 mM benzyl Hcy. After specific time intervals reported to the right of each panel, the cells were fixed and stained with H3K9benzyl antibody. The figure shows heterochromatin regions in blue and H3K9benzyl stained regions in red. The imaging data was collected for more than 20 replicates.

We plan to conduct more rigorous time and concentration dependent experiments.

Further, we have a mutant CBX1 (reader protein), discussed in detail in Chapter 3, that binds invitro to benzylated H3 peptide and full-length protein. CBX1 protein is known to bind to H3K9me3 and form heterochromatin in cells. To explore the possibility of CBX1 Y26F/F50G mutant behaving similarly to wild-type system we conducted imaging-based experiments. We used the GFP-CBX1-Y26F/F50G mutant to visualize its localization. Cells were grown with the engineered proteins and after 24 hours of treatment with benzyl Hcy they were fixed on cover slips and stained with antibodies. The cells untreated with benzyl Hcy, CBX1 mutant did not localize to DAPI rich regions in the nucleus. Considering the lack of binding partner (benzylated H3) we proceeded to image the cells treated with benzyl Hcy, CBX1 mutant did not localize to benzylated regions in nucleus. This implied that our mutant CBX1 does not behave as the wild-type CBX1 in cells. Another mutant might be required to proceed with heterochromatin formation with CBX1 in cells.



Figure 15. Localization of CBX1 Y26F/F50G in HE293T Cells. HEK293T cells expressing GFP-CBX1 Y26F/F50G were fixed and stained with DAPI and Benzyl Ab either treated with or without benzyl Hcy. Merge for both treated and untreated cells show that CBX1 Y26F/F50G mutant does not localize to DAPI rich regions in the nucleus.

From our imaging experiments we have established that our antibody is specific towards benzylgroup as it stains only on cells treated with benzyl homocysteine. On applying our engineered system, we specifically benzylate DAPI rich heterochromatin regions in the nucleus of HEK293T cells, as expected from SUV39h2 histone methyltransferase. Also, our engineered system takes 8-9 hours to actively benzylate the nucleus of the cells, this may be attributed to lower catalytic activity of mutants towards benzyl Hcy and benzyl SAM. Finally, in figure 15 we show that H3K9me3 and H3K9benzyl overlap in their localization to DAPI rich heterochromatin regions in cells.

2.2.9. Establishing the Functional Relevance of Engineered System

qRT PCR, quantitative Reverse Transcriptase PCR, is the process of converting RNA to cDNA by

Gene	Coded Protein
AP3M2	Adapter Related Protein Complex 3 subunit Mu 2
ASIC1	Acid Sensing Ion Channel subunit 1
C17orf103	N-acetyltransferase domain containing 1
DHX37	DEAH Box Helicase 37
NRIP1	Nuclear Receptor Interacting Protein 1
NLRP1	NLR Family Pyrin Domain Containing 1
PTOV1	Prostate Tumor Overexpressed gene 1 protein
SLIT1	SLIT homologue 1 protein
ZFX	Zinc Finger protein X-linked

Table 2. List of Genes Selected for qRT PCR. The table lists the genes along with the coded

 protein selected based on published reports.

reverse transcriptase enzyme followed by PCR amplification of the gene of interest (using gene specific primers) and quantifying by fluorogenic probes.

SUV39h2 is reported to regulate several genes by methylating the H3 about which the DNA is wrapped. Recently, C.Chao *et al* reported several genes being dysregulated by SUV39h2 overexpression in nasopharyngeal carcinoma⁷⁴. We selected a few genes that were being down-regulated probably due to hyper-methylation at the H3 on the transcription start sites for these genes. If engineered system would benzylate the transcription start-site of these genes we would expect a dysregulation of these genes. These are as described in table 2. Upon extracting the RNA from the cells treated with or without the benzyl-homocysteine but expressing the hole modified enzymes we saw a downregulation for most of the genes, as shown in figure 16. Before analyzing the expression of targeted genes, the overexpression of SUV39h2 mutants was confirmed by using the primers for it as a positive control. We used Actin as our reference gene.



Figure 16. Gene Regulation by Benzylation on H3. HEK293T cells were either not transfected or transfected with MAT2A I117G and SUV39h2 Y372G and treated with or without benzyl Hcy. Actin was used as a control. The error bars represent indicate triplicate data a biological sample. The data was repeated for biological duplicates in two or more statistical replicates.

The downregulation or upregulation of the genes confirmed their regulation by the activity of SUV39h2 although it is difficult to predict whether the downregulation is due to the following downstream interacting partners not identifying these modifications or due to formation of heterochromatin. To gain a better understanding of the engineered system we went for ATAC-seq of our system.

2.2.10. Studying the Accessibility of Benzylated Genome

Assay for Transposase Accessible Chromatin followed by sequencing (ATAC-seq), is the technique that uses transposase Tn5 to fragment accessible regions of chromosome, followed by insertion of oligonucleotides at the ends of the fragments. These fragments are then amplified using appropriate primers followed by sequencing. The cleaved accessible sites on chromosomes would provide valuable information about the accessibility of heterochromatin and euchromatin regions in cells.

In RNA-seq total RNA including mRNA, tRNA, rRNA from samples is extracted, and only mRNA is pulled down for amplification and sequencing. This is possible due to the presence of poly-A tail at 3' end of mRNA. Sequencing mRNA from our samples would give us insight into the genes/ proteins being directly regulated by SUV39h2. Thus, providing direct evidence for functional relevance of the engineered system.

For this we collaborated with Dr. Janette Lamb at the Genomics Research Core, University of Pittsburgh, for preparation of the ATAC-seq and RNA-seq samples followed by analysis by Dr Uma Chandran at the Cancer Bioinformatics Services, UPMC Hillman Cancer Center.

Samples in HEK293T cells were prepared for this experiment expressing either the native system or SUV39h2 Y372G both with and without benzyl-Hcy. We do not expect any of the native enzymes (transcription factors or nucleosomal remodeling complexes) to identify and bind to bulky benzyl modification on H3, thereby melting the heterochromatin or increasing the accessibility of chromatin. To gain a better picture of the effect of modification on chromatin accessibility as well as genes being regulated by SUV39h2 we would have to wait for the data analysis by the facility.

2.3. CONCLUSIONS AND FUTURE DIRECTIONS

Past studies have demonstrated that engineering small-molecule and protein interface is a powerful tool to elucidate the function of a protein. As methyltransferases have similar catalytic pocket elucidating member specific function has been challenging. To our knowledge, for the first time we have generated an allele-specific pair for SUV39h2. We have demonstrated the orthogonality and catalytic efficiency for incorporation of benzyl-analogue on H3K9 by SUV39h2 Y372G mutant. As SAM analogues are not cell-permeable we engineered the catalytic pocket of MAT2A that converts cell permeable benzyl homocysteine to benzyl SAM. To validate our system in cells we generated H3K9-benzyl antibody. We have demonstrated the successful incorporation of benzyl on H3 in HEK293T cells via western blot and imaging experiments. To further understand the role of SUV39h2 in gene regulation as well as possibility of heterochromatin formation employing our engineered system, we have sent samples for ATAC-seq and RNA-seq. We anticipate identification of several genes being regulated by SUV39h2 that contribute to its

oncogenesis and whose detection was challenging due to reversibility of methylation on H3 and due to lack of a proper tool.

2.4. MATERIALS AND METHODS

2.4.1. Expression and Purification of SUV39h2 and Its Mutants

N-terminal 6xHis-tagged SET domain of Suv39H2 (112-410 amino acids, PDB code: 2R3A) in pET28a-LIC vector was obtained from Prof. Jinrong Min at the University of Toronto. The SUV39h2 construct contained H327R mutation for improved methyltransferase activity. The plasmid was transformed in *E.coli* One Shot BL21-star (DE3), chemically competent cells (Invitrogen) for expression. A single colony was picked and grown overnight in 10 mL of Luria-Bertani (LB) at 37°C in presence of 50 µg/mL of kanamycin. The 10 mL culture was diluted to a 100-fold and the cells were grown at 37°C in presence of 50 µg/mL of kanamycin till they reached an optical density (O.D) of 0.8. Protein over-expression was induced by adding 0.5 mM of IPTG, in the presence of 25 µM ZnCl₂ and grown at 17°C overnight. The cells were harvested and resuspended in buffer containing 50 mM Tris (pH 8.0), 150 mM NaCl, 10 % glycerol, 5 mM βmercaptoethanol, 25 mM imidazole, DNase, lysozyme and protease inhibitor tablets. The cells were lysed by pulsed sonication (Qsonica-Q700) and centrifuged at 13,000 rpm for 50 mins at 4°C. The soluble extracts were incubated with Ni-NTA agarose resin (Thermo) according to manufacturer's protocol. The beads were washed with 20 column volumes of wash buffer (50 mM Tris pH 8.0, 150 mM NaCl, 10 % glycerol, 5 mM β-mercaptoethanol and 25 mM imidazole). The protein was eluted with 50 mM Tris pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol and 400 mM imidazole. Eluted protein was subjected to further purification by gel filtration chromatography (Superdex-75) using AKTA pure FPLC system (GE healthcare) with buffer containing 50 mM Tris pH 8.0, 150 mM NaCl and 10 % glycerol. The purified protein was concentrated using Amicon-Ultra 10K centrifugal filter device (Merck Millipore Ltd.). The concentration of the protein was determined using Bradford assay kit (BioRad Laboratories) with BSA as standard. SUV39h2 Y372G/A mutants were generated by the QuickChange site-directed mutagenesis method (Stratagene) by following manufacturer's instructions. The resulting mutant plasmid was confirmed by DNA sequencing. The mutants were expressed and purified as described above for wild type SUV39h2.

2.4.2. Expression and Purification of MAT2A and Its Mutants

MAT2A and MAT2A I117A mutant were a kind gift from Dr. M. Luo at the Memorial Sloan-Kettering Cancer Center. The plasmids were transformed in *E.coli* One Shot BL21-star (DE3) chemically competent cells (Invitrogen) for expression. Single colony was picked and grown overnight in 10 mL of Luria-Bertani (LB) at 37°C in presence of 50 μ g/mL of kanamycin. The 10 mL culture was diluted to a 100-fold and the cells were grown at 37°C in presence of 50 μ g/mL of kanamycin till it reached an optical density (O.D) of 0.8. The protein over-expression was induced by adding 0.5 mM of IPTG to cultures and grown at 17°C overnight. The cells were harvested and resuspended in buffer containing 50 mM Tris (pH 8.0), 150 mM NaCl, 10 % glycerol, 5 mM β -mercaptoethanol, 25 mM imidazole, DNase, lysozyme and protease inhibitor tablets. The cells

were lysed by pulsed sonication (Qsonica-Q700) and centrifuged at 13,000 rpm for 50 mins at 4°C. The soluble extracts were incubated with Ni-NTA agarose resin (Thermo) according to manufacturer's protocol. The beads were washed with 20 column volumes of wash buffer (50 mM Tris pH 8.0, 150 mM NaCl, 10 % glycerol, 5 mM β-mercaptoethanol and 25 mM imidazole). The protein was eluted with 50 mM Tris pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol and 400 mM imidazole. Eluted protein was subjected to further purification by gel filtration chromatography (Superdex-75) using AKTA pure FPLC system (GE healthcare) with buffer containing 50 mM Tris pH 8.0, 150 mM NaCl and 10 % glycerol. The purified protein was concentrated using Amicon-Ultra 10K centrifugal filter device (Merck Millipore Ltd.). The concentration of the protein was determined using Bradford assay kit (BioRad Laboratories) with BSA as standard. The concentrated protein was stored at -80°C before use. MAT2A II17G mutant was generated by the QuickChange site-directed mutagenesis method (Stratagene) by following manufacturer's instruction. The resulting mutant plasmid was confirmed by DNA sequencing. The mutant was expressed and purified as described above for wild type MAT2A.

2.4.3. Benzyl SAM Synthesis



Native SAM was purchased from Sigma-Aldrich (cat# A4377) and used in the methylation assay without further purification. Briefly, 50-100 equivalents of each activated electrophile, benzyl bromide, was dissolved into a freshly prepared mixture of formic acid and acetic acid (1:1, 2 mL) and placed in an ice bath. S-adenosyl-L-homocysteine (10 mg, 26 µM, 1 equivalent) was added to the solution. AgClO₄ (5.4 mg, 1 equivalent) was used to drive the reaction to completion. The reaction was warmed to ambient temperature and continued until completion (typically 5-8 hrs.). The resultant reaction mixture was quenched by adding 10 mL of distilled water containing 0.01% TFA (v/v). The aqueous phase was washed three times with diethyl ether (3×10 mL) and then passed through Nalgene 0.2 µM syringe filter. SAM analogues were purified with semi-preparative reversed-phase HPLC (XBridge BEH C18 OBD Prep Column, 5µm, 10×250mm) eluting at a flow rate of 4 mL/min with acetonitrile (linear gradients to 10% in 30 min and then to 70% in 5 min) in aqueous trifluoroacetic acid (0.01%). Since the stereochemistry at sulfonium center could not be assigned unambiguously, diastereomeric mixtures of these compounds were collected. The mixture was concentrated by SpeedVac for 2 hr, followed by lyophilization overnight. The dried product was re-dissolved in water containing 0.01% TFA (v/v)

and stored at -80° C before use. The concentrations of the SAM analogues were determined by UV absorption with $\varepsilon_{260}=15400$ L mol-1 cm-1. Concentrations of the bioactive epimers were estimated to 50% of the total diastereomeric mixture. Integrity of the compounds was confirmed by ESI-HRMS and comparing the chemical shifts in the ¹H NMR spectra to those of the reported values for the indicated compound.

2.4.4. Methyltransferase Assay

Trimethylated H3K9 peptide was generated on (ARTKQTARKSTGGKAPRKQLAK(Biotin)-CONH2) by incubating 2 μ M of wild-type enzyme with 20 μ M of peptide and 80 μ M SAM in 50 mM Tris (pH 8.0) for 30-40 mins at room temperature. For the 'bumped' modifications 5 μ M G9a-Y1154A or G9a-Y1154G mutant was incubated with 20 μ M peptide and 200 μ M SAM-analogue in 50 mM Tris (pH 8.0) for 150-180 mins at room temperature. The methyltransferase activity was confirmed using MALDI mass spectrometry. 1 μ L of the assay volume was mixed with 1 μ L of saturated solution of α -cyano-hydroxy-cinnamic acid (CHCA) on the MALDI target plate. The sample was analyzed using Flex analysis for Bruker Daltonics UltrafleXtreme MALDI TOF-TOF on the Reflector Positive mode. The negative control included all components of the assay except for the SAM analogue. For SUV39h2 kinetics, 0.5 μ M of SUV39h2 or 2 μ M of SUV39h2 Y372G, enzyme was incubated with 10 μ M peptide and varying concentrations of SAM (0-50 μ M) and benzyl SAM (0-200 μ M) respectively in 50 mM tris pH 8.0. The reaction was spotted on the MALDI plate with CHCA matrix at varying time points. The area under each modified/ unmodified peak was integrated to calculate the percentage modification. Sigma Plot was used for calculating the slope of each concentration, that was then employed to calculate the kinetics using (Michaelis-Menten equation).



Figure 17. SUV39h2 MALDI Data. The data was obtained from MALDI after methyltransferase assay on H3 peptide by SUV39h2 with different SAM analogues (**A.** for SAM, **B**. for Allyl SAM, **C**. for Benzyl SAM, and **D**. for Benzyl-yne SAM). The unmodified peptide corresponds to a peak at 2607 m/z. Trimethylation would correspond to a peak at 2649 m/z. Benzylated peptide would correspond to 2697 m/z and benzyl-yne would correspond to 2723 m/z.



Figure 18. SUV39h2-Y372A MALDI. The data was obtained from MALDI after methyltransferase assay on H3 peptide by SUV39h2-Y372A with different SAM analogues (**A**. for SAM, **B**. for Allyl SAM, **C**. for Benzyl SAM, and **D**. for Benzyl-yne SAM). The unmodified peptide corresponds to a peak at 2607 m/z. Momo, Di, and trimethylation would correspond to peaks at 2621, 2635 and 2649 m/z. Benzylated peptide would correspond to 2697 m/z and benzyl-yne would correspond to 2723 m/z.



Figure 19. SUV39h2 Y372G MALDI. The data was obtained from MALDI after methyltransferase assay on H3 peptide by SUV39h2-Y372G with different SAM analogues (**A.** for SAM, **B.** for Allyl SAM, **C.** for Benzyl SAM, and **D.** for Benzyl-yne SAM). The unmodified peptide corresponds to a peak at 2607 m/z. Momo, Di, and trimethylation would correspond to peaks at 2621, 2635 and 2649 m/z. Benzylated peptide would correspond to 2697 m/z and benzyl-yne would correspond to 2723 m/z.



Figure 20. MALDI Data for Kinetics of SUV39h2 with SAM. The MALDI data for methyltransferase assay on SUV39h2 with SAM at different concentrations shown here. The unmodified peptide corresponds to the peak at 2607 m/z, mono, di, and trimethylation at 2622, 2336 and 2650 Da.



Figure 21. MALDI Data for Kinetics of SUV39h2 Y372G with Benzyl-SAM. The MALDI data for methyltransferase assay on SUV39h2 Y372G with Benzyl-SAM at different concentrations is shown here. The unmodified peptide corresponds to the peak at 2607 m/z and benzylated peptide to 2697 m/z.

2.4.5. MAT2A and SUV39h2 Coupled Assay

The assay was conducted with 10 μ M H3-peptide was added into 100 μ L mixture containing 50 mM Tris-HCl pH 8.0, 100 mM KCl, 2 mM MgCl₂, 2.5 mM ATP, 2.5 mM a methionine analogue, 7.5 μ M MAT2A mutant, 1 μ M SUV39h2 or 5 μ M SUV39h2 mutant. The resultant reaction mixture was incubated for 30 min at ambient temperature (22°C). After 30 mins the assay was spotted on MALDI plate and analyzed as stated above.



2.4.6. Synthesis of e-boc a-N-Fmoc Lysine

The compound was synthesized following the reported method. Benzaldehyde (262 μ L, 2.97 mmol) was added to a stirred solution of N α -Fmoc-lysine hydrochloride (1 g, 2.97 mmol) in 20mL of anhydrous ethanol in the presence of 3 Å molecular sieves at room temperature. After stirring for 1 h, sodium cyanoborohydride (186 mg, 2.97 mmol) was added. The reaction was stirred overnight. The reaction mixture was then acidified by dropwise addition of 0.2 M hydrochloric acid, which led to a fine white precipitate. The reaction mixture was filtered, and the filtrate was concentrated to afford crude mixture. Without purification, the crude mixture was subject to Boc-

carbamate protection as the following. Into a stirred solution of the crude mixture, 10 mL of tetrahydrofuran, sodium bicarbonate till the solution reached a pH 8, and Boc anhydride (640 mg, 2.96 mmol) were added. Upon completing the reaction, which was monitored by TLC, the reaction mixture was washed with diethyl ether. The aqueous layer was brought to a pH of 6 and then extracted with dichloromethane and washed sequentially with 0.2 M HCl (5 mL × 3), saturated sodium bicarbonate (5 mL × 3), and brine (5 mL), dried over anhydrous magnesium sulfate and concentrated to give crude product. The crude product was purified by silica gel column chromatography with ethyl acetate and hexane (v/v of 50:50) in 1% acetic acid as eluent to afford pure N α -Fmoc-N ϵ -(benzyl, Boc)-lysine with 31% yield (0.52 g, 0.93 mmol) after two steps. Integrity of the compound was confirmed by ESI-HRMS and comparing the chemical shifts in the 'H NMR spectra to those of the reported values.

2.4.7. Synthesis of Benzyl Hcy



The compound was synthesized following ref⁷⁵. Methionine (1 g, 6.7 mmol) was dissolved in cc HCl (~3 mL) by stirring and heating the flask to 60°C. The flask was refluxed over-night after addition of benzyl chloride (770 μ L, 6.7 mmol). Benzyl homocysteine was washed with diethyl 54

ether, water was evaporated to dryness and benzyl homocysteine was re-dissolved in hot water. Integrity of the compound was confirmed by ESI-HRMS and comparing the chemical shifts in the 'H NMR spectra to those of the reported values.

2.4.8. Peptide Purification for Methyltransferase Assay and Ab Generation

All the H3 peptides were synthesized at the University of Pittsburgh Peptide Synthesis facility. Analytical scale separation was performed using Zorbax reversed phase C18 (5 μ m, 4.6 x 250 mm) column with UV detection at 220 nm. The column was equilibrated with 0.1% aqueous trifluoroacetic acidic (TFA) and the peptide was separated with the linear gradient of acetonitrile to 10% in 15 min and then to 70% in 5 min with a flow rate of 1 mL/min. The crude peptides were purified using semi-preparative reversed phase HPLC column Xbridge C18 (5 μ m, 10 x 250 mm) eluting with a flow rate of 5 mL/min and an acetonitrile gradient from 0% to 90% in 15 mins and then to 100 % in 18 mins in 0.1% aqueous TFA. The purified peptides were concentrated on speedvac and dried in the lyophilizer. The dried peptides were resuspended in 0.1% aqueous TFA and the concentration was measured based on the observation that 1 mg/mL of peptide generates an absorbance of 30 at a wavelength of 205 nm. The integrity of the peptides was confirmed by MALDI mass spectrometry.
2.4.9. MAT2A and SUV Mutant Expression Followed by Benzylation In-Cells

Human embryonic kidney (HEK) 293T cells were grown in Dulbecco modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum in a humidified atmosphere containing 5% CO₂. MAT2A and SUV39h2 mutants were transfected in the cells at 60% confluency with the aid of lipofectamine 2000. 14 hours post transfection, the media was exchanged with methionine-free DMEM media supplemented with 10% FBS (Gibco) for 20 mins to deplete the methionine. The media was again exchanged with methionine free DMEM media supplemented with 10% FBS (and 1mM benzyl homocysteine. The cells were grown for 24 in 5% CO2 at 37°C and then uplifted using trypsin and collected for nuclear extraction/ imaging/ qRT PCR.

2.4.10. Nuclear Extraction

Nuclear extracts were generated by incubating the cells with NIB (Nuclear Isolation Buffer) on ice for 15 mins. NIB buffer contains 15 mM Tris pH 8.0, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 1 mM CaCl2, 1 mM DTT, 2 mM sodium orthovanadate, 250 mM sucrose, 1x Protease Inhibitor and 0.4% NP-40. The sample was centrifuged at 2000g for 5 mins at 4°C, following which the pellet is resuspended in NEB (Nuclear Extraction Buffer). NEB contains 25 mM Tris pH 8.0, 250 mM NaCl, 1mM EDTA, 10% glycerol, 0.2% NP-40 and 1x Protease Inhibitor (Roche). The nuclear extracts are then pulse sonicated at 100 Amps for 10 mins with each pulse lasting for 1 min ON and 20 secs OFF duration. The sheared DNA was centrifuged at 2000g for 5 mins at 4°C. The concentration of nuclear extracts was measured using Bradford assay.

2.4.11. Western Blot

Equal volumes of samples were separated on SDS-PAGE and transferred onto 0.45 µm PVDF membrane at a constant voltage of 80V for an hour at 4°C. The membrane was washed once in TBST buffer (50 mM Tris pH 7.4, 200 mM NaCl, and 0.01% Tween) for 5mins at RT. Then blocked for an hour at room temperature (RT) in 5% milk buffer prepared in TBST. Immunoblotting was performed with primary antibodies with dilutions as per manufacturer's protocol for H3K9bn antibody (PRF&L) overnight at 4°C. The membranes were washed with TBST buffer thrice at RT for five minutes each. The blots were then incubated with the HRP conjugated secondary antibodies Goat anti-Rabbit IgG (Active Motif cat# 15015) with 5% nonfat dry milk, dilution 1:5000 in TBST. The membranes were washed again with TBST buffer thrice at RT for five in bands were visualized by chemiluminescence using VISIGLO HRP Chemiluminescent substrates A and B (cat# N252-120ML and N253-120ML, aMReSCO) following manufacturer's protocol.

2.4.12. Imaging Experiments

The cells were fixed in 2% paraformaldehyde made in PBS for 15 minutes at room temperature. The monolayer was washed twice in PBS to remove residual paraformaldehyde, The samples were stored in PBS at 4°C until use. The cells were permeabilized with 0.1% Triton X-100 made in PBS solution for 15 min then washed 3 times with PBS followed by washing 5 times with a PBS+ 0.5% BSA (PBB). The cells were blocked with 2% BSA for 45 minutes at room temperature. The monolayer was washed 5 times with PBB. The primary antibody was diluted to the desired concentration in PBB and 70-100 μ l antibody solution was dropped over the section for overnight incubation at 4°C. The monolayer of cells was washed 5 times with PBB. The secondary antibody was diluted in PBB and the cells were incubated in it for 60 minutes followed by washing 5 times with PBB. Cells were incubated in Hoechst (1 mg/100 mL) for 30 seconds followed by washing thrice with PBS. The coverslips were adhered to slides with gelvatol and stored at 4°C for imaging.

2.4.13. qRT PCR

Total RNA was extracted from the samples using E.Z.N.A HP Total RNA Kit (Omega Bio-Tek). 1ug of RNA from each sample was used for reverse transcriptase reaction using qScript cDNA SuperMix (Quantabio). For cDNA formation, reaction was conducted at 42°C for 30 minutes followed by 85°C for 5 minutes following manufacturer's protocol. 1 µL of the cDNA reaction mix was added to a mix of 5 µM reverse and forward primers along with PerfeCTa SYBR Green Supermix Reagent (Quantabio). The qRT-PCR was setup on Bio-Rad CFX Manager 3.1 for 45 58 cycles with each cycle denaturing at 95°C for 10 secs and binding at 55°C for 30 secs. As the number of cycles in PCR increase the intensity of the probe increases exponentially. C_t is set as the threshold value of fluorescence intensity above which the signal crosses the background value, the cycle number is referred to as C_q . More the template mRNA in the sample lower is the C_q value. Housekeeping genes such as GAPDH and Actin are used as reference genes to relatively analyze the upregulation or downregulation of target genes in a sample, this value is known as ΔC_q . When comparing the target genes from two different samples the relative expression is known as $\Delta \Delta C_q$, also referred to as relative normalized expression. Results were analyzed using the Data Analysis of the Bio-Rad CX Manager in delta delta Cq mode, relative to control (control is setup as empty).

3. ENGINEERING METHYLLYSINE READERS TO CONTROL CHROMATIN-DEPENDENT PROCESSES

3.1. INTRODUCTION

Trimethylation at lysine 9 on histone 3 (H3K9) is an important epigenetic mark for heterochromatin formation, SUV39h2 is a key player for incorporating this mark. Other key player in maintenance and formation of heterochromatin are HP1 proteins. HP1 protein was discovered in 1986 in *Drosophila*, as a non-histone chromosomal protein associated with heterochromatin ⁷⁶. In 1930 Muller *et al* translocated genes in flies by X-rays. Upon translocation of genes from euchromatic region to the vicinity of heterochromatin region, they acquired a variegated expression ⁷⁷ called Position Effect Variegation (PEV). Through mutations in HP1 gene, the protein was identified as a suppressor of PEV ⁷⁸ although the mechanism of PEV was not determined.

The chromodomain (chromatin organization modifier) is identified as a stretch of around 50 amino acids that binds to methylated lysine, such as K4, K9, K27 on H3 and K20 on H4. There are 29 mammalian proteins that contain chromodomain ⁷⁹ that includes : HP1-like proteins, Polycomb (Pc)-like proteins and CHD tandem proteins. There are three homologues of HP1 proteins in humans coded by *chromobox gene*, namely CBX1, CBX3 and CBX5. The importance of

chromodomain containing proteins in epigenetic regulation is reflected by the fact that it is conserved across different species of animals as well as plants ⁸⁰.

Biophysical studies have shed light on the binding of CBXs to H3K9me3. The CBX proteins belonging to HP1 family have a conserved chromodomain (CD) at the N-terminal, a chromoshadow domain (CSD) at C-terminal that are linked by the hinge region (HR). The chromodomain of CBXs recognize the methylated lysine in TARKmeS motif present on histone (H3⁸¹) and non-histone proteins (G9a⁸²). Phosphorylation of S10 adjacent to K9me3 hinders the interaction between CBX1 and H3K9me3⁸³. The binding of CBX1, CBX3 and CBX5 to H3K9me3 is in the range of μ M, at 5 μ M, 15 μ M and 30 μ M respectively as determined by fluorescence polarization assay⁸⁴. Mutating H3T6 to alanine decreases the binding of CBXs to H3K9me3 by 10-fold, demonstrating the importance of (-3) amino acid on H3 in its binding to CBX⁸⁴. The NMR structure of homologue of CBX1 in mouse, MOD1 or M31, with H3K9me2 provided insights into the residues involved in their binding ⁴². It revealed that the aromatic cage formed by Tyr 21, Trp 42 and Tyr 45 interacts with the positively charged trimethylated lysine moiety via cation-pi interactions. Their study predicted the induced-fit mechanism for binding of peptide to the CD of M31, as the β -stranded structure of the protein and peptide extend upon binding. The structure also shows that QTAR residues preceding K9 on H3 peptide are involved in binding to the protein. Together these studies frame the molecular picture of interaction and binding of H3K9me3 and CBX1.

3.1.1. Role in Development and Disease

Dialynas *et al* in 2007⁸⁵ studied the distribution of CBX proteins in ten different mammalian cell lines (both undifferentiated and partially differentiated ES cells) and established the heterogeneity of the distribution pattern of these proteins in a cell-dependent manner. Through Fluorescence Recovery After Photobleaching (FRAP) based experiments they established that CBX1 and CBX5 mostly localize to heterochromatin region whereas CBX3 is associated to euchromatic region in a distinct fashion. Aucott R, *et al* in 2008⁸⁶ carried out the gene knock-out (KO) experiment and showed CBX1 is essential for mice neonates development as CBX-/- homozygotes died within few hours of birth. This suggested that CBXs have non-redundant functions in cellular development as other CBX proteins were not able to compensate the loss of CBX1. In 2015 Mattout *et al*⁸⁷ showed that CBX1 KO embryonic stem cells (ESCs) differentiated spontaneously compared to the wild type (WT) ESC, suggesting that CBX1 also plays an important role in maintaining the pluripotency of stem cells. Their data confirmed that CBX1 regulates developmental genes and pluripotent factors in ESCs. Similar studies have also confirmed the importance of CBX5 and CBX3 in development.

CBXs are also involved in dynamic cellular processes like transcriptional activation and elongation, sister chromatid cohesion, chromosome segregation, telomere maintenance, DNA repair, and RNA splicing. Despite having similar structural and biochemical similarities, HP1 variants have overlapping as well as distinct functions and their nuclear localization is also dissimilar in different cells. The role of each HP1 family member has had limited investigation.

3.1.2. Efforts Towards Studying the Role of CBXs

To study the member specific role of CBXs (CBX1, CBX3 and CBX5) the genetics approach is not facile as the CBXs have redundant or overlapping functions. The chemical approach of developing the inhibitors faces two major challenges; firstly, all methyl-lysine reader proteins have a similar aromatic binding pocket thereby compromising the specificity for target proteins, and secondly, as it is protein-protein interaction the inhibitor should mimic all the surface interactions and binding pocket interactions to compete out the native protein. Despite these challenges there are methyl-lysine specific inhibitors such as UNC1215 for L3MBTL3 (MBT domain), UNC2170 for 53BP1 (tudor domain) and Ac-FAYkme3S-NH2 for CBX7 (chromodomain)⁸⁸. For chromodomain containing proteins the efforts are employed towards developing an inhibitor for CBX7 and CBX6, as they are involved in several diseases⁸⁸.

Presently there is no inhibitor for CBX protein that belongs to HP1 family, therefore studying member specific function of these proteins is challenging.

3.1.3. Our Approach

The objective of my work is to elucidate the function of CBXs in member-specific manner through protein-protein interface engineering. To accomplish the goal, I would like to employ two complementary chemical genetic approaches: (1) Bump-hole approach to develop orthogonal CBX1 and H3K9Me3 pairs and (2) Unnatural amino acid (UAA) mutagenesis to furnish photocross-linkable CBXs to identify their interacting partners.

3.1.4. Reported Applications of Bump-Hole to Protein-Protein Interface Engineering

Bump-Hole approach or the Allele Specific Chemical Genetics approach has been successfully employed to enzymes that bind to small molecules. There are very few examples that employ this technique to investigate protein-protein interactions. Engineering of PPI based interaction by Crick to explain the packaging by amino acids in helices was known as "knobs into holes". To understand the role of PPI, allele based chemical genetics has been successfully employed at the interface of calmodulin and calmodulin binding protein and homodimer of chorismate mutase ⁸⁹.

Here I wanted to employ this technique at the epigenetic reader and modified H3K9 interface to generate their orthogonal pair and then translate it to other CBX members to study their member specific functions.

3.2. RESULTS AND DISCUSSION

Successful implementation of bump-hole strategy on a system is based on: 1) High-resolution structural information about the protein-protein interface, 2) the structural and functional integrity of mutated proteins and 3) biocompatibility of mutant allele and the ligands for *in vivo* applications ¹. Based on the structure available at PDB and the binding affinity of CBX1 for H3K9me3 we endeavored to generate an orthogonal bump-hole pair for this system.

3.2.1. Rationale for Generating the Mutants

All HP1 reader proteins have a N-terminal acidic patch. The Addgene construct for bacterially expressing CBX1 lacked the N-terminal acidic patch present in most of the HP1-family chromodomains. We believed this sequence to be important for binding to the histone tail as histone proteins are composed of several basic residues to interact with the acidic phosphate group on DNA. We inserted the acidic LEEEEE patch in the construct by site-directed mutagenesis and confirmed by sequencing. This construct was used for further studies and called wild-type CBX1.



Figure 24. Identification of Binding Site Residues on CBX1. Left Panel shows the nmr structure of CBX1 bound to H3K9me2 (**PDB 1GUW**). In blue is the CBX1 protein and in orange is the H3K9me2 peptide. The panel on right illustrates the residues in the binding site of chromodomain of CBX1.

Based on the NMR structure available at protein data bank (PDB) of CBX1 with H3K9me2 peptide (PDB ID: 1GUW), figure 24, five residues that form the binding pocket (called aromatic cage) suited the purpose of expanding the binding pocket. These residues were mutated to alanine and glycine residues. The hole-modified constructs were generated by site-directed mutagenesis,

confirmed by sequencing and expressed in *E.coli*. The proteins were purified using affinity chromatography followed by size exclusion chromatography. All the proteins thus generated were verified by LCMS.

3.2.2. Generation of Peptide Library

A small peptide containing H3K9 sequence with the TAMRA tagged on its N-terminal (TAMRA-GGARTKQTARKSTGGKAPR) was synthesized and purified to incorporate the bumped analogues at lysine 9 using G9a. G9a is a methyltransferase that transfers the methyl group from S-Adenosyl Methionine (SAM) to H3 at K9 position specifically. Islam et al. in 2011 employed the bump and hole technique on G9a and its ligand SAM to incorporate bumped modification on H3 at K9. Based on this prior work, we mutated F1152 and Y1154 sites to alanine and glycine (holes) on G9a, sequenced the plasmids, expressed the protein in *E.coli* and purified them. We synthesized a series of SAM analogues following reported methods. These analogues were purified by HPLC and characterized by LCMS. The methyltransferase assay was conducted on all the SAM analogues using the two G9a* (hole modified G9a) to incorporate the bumped modification on TAMRA labeled H3K9 peptide. The full conversion of the peptide was confirmed by MALDI mass spectrometry. G9a-F1152 mutant only transferred allyl-peptide efficiently while the remaining peptides were generated using G9a-Y1154 mutant. A schematic of the assay is shown in figure 25. Next, we screened the hole-modified CBX1 mutants against the bumped H3K9R peptides to find orthogonal pairs with strong binding affinity.



Figure 25. G9a Methyltransferase. The top panel shows the activity of G9a or the mutant G9a* to transfer the R-group from SAM to Histone3 at lysine 9. The bottom panel shows the MALDI for TAMRA labeled peptide on the bottom left panel to successful transfer of benzyl group by G9a Y1154G mutant on the bottom right panel.

3.2.3. Screening and Validating the Binding of Bump-Hole Pairs

Based on our dependence on G9a methyltransferase assay for the generation of peptide we preferred using fluorescence polarization (FP) assay for the purpose of screening, as it requires low concentration of peptide. FP assay is based on the principles of optics. Light is an electromagnetic radiation, in an unpolarized form vibrates in all directions perpendicular to the direction of propagation. If the electric field vector is restricted to a single plane, then the light is plane polarized. A fluorescently labelled peptide, small-molecule or oligonucleotide polarizes the

light emitted on it equally in all the directions (unpolarized light), as the fluorophore reorients during the lifetime of its excited state. On binding to a heavy bio-macromolecule the rate of reorientation reduces as compared to the lifetime of the excited state. Thus, the emitted light becomes polarized upon the binding of fluorophore attached small molecule to the bio-macromolecule. On titration of fluorophore with the protein the measure of polarization increases giving a sigmoidal binding curve, that gives the K_d value for the system.



Consideration the requirement of fluorophore on the peptide, we employed TAMRA labeled peptide in the G9a methyltransferase assay. For FP assay we followed conditions same as those reported by Seeliger *et al*⁹⁰. They used 100nM peptide in their assay but we got better signal on two-fold increase in the peptide concentration (200nM), this concentration was kept consistent in the screening assays. The semi-log binding curve produced K_d value of 1.7μ M for CBX1-H3K9me3 system that was close to that reported by other groups^{84,90}, as shown in figure 27. The curves were fitted to the equation Fx= Fmin + (Fmax – Fmin)*x / (K_d + x). Fx is the fluorescence polarization signal at concentration x; Fmin is the fluorescence polarization in the unbound state; Fmax is the fluorescence polarization in the bound state; K_d is the apparent molar dissociation constant. Hence the assay conditions were set for screening the CBX1-hole modified proteins and G9a-SAM based TAMRA labeled H3K9R peptide.



Figure 27. Schematic of FP Assay. The top left panel illustrates that polarized light gets unpolarized by fast rotating small probes but not by heavy and slowly rotating probes in the bottom left panel. As more protein binds to the fluorescent probe the rotational dynamics of the probe slows down thereby generating a high FP signal that is then plotted to give a binding curve.

We started the screening by conducting the binding assay for all the hole-modified CBX1 against H3K9me3. A high K_d would signify lack of binding, thus orthogonality to the wild-type substrate. The dissociation constants (K_d) for D54A and T56A was 8.3 μ M and 20 μ M respectively from our assay conditions. The low K_d values implied that the hole-modified mutant had a strong affinity for native peptide and hence would not be bio-orthogonal. Several papers suggest that W47 is an important residue for CBX1's binding to trimethylated peptide⁴². We also observed no binding for W47A mutant even at 1mM concentration for the H3K9me3-TAMRA peptide. We generated a few more mutants of W47 as W47G, W47V and W47F to determine whether increasing the hydrophobicity or the aromaticity would restore the binding affinity. We observed no binding for W47V and W47G mutant but a K_d of 120 μ M for W47F. This suggested that aromaticity plays an important role in this binding event. This result inspired us to include benzyl H3K9 peptide in our screening for CBX1 bump-hole system as well, to restore the cation-pi interactions.

The data for wild type CBX1 suggests that on introducing the bulky groups on lysine residue, the steric clash between CBX1 and modified H3K9R peptides increases which in turn increases the K_d values (lowering the binding affinity). Decrease in binding affinity of mutant CBX1-F50A compared to WT CBX1 with H3K9me3 was measured through FP assay. Next, the remaining hole-modified CBX1 proteins were screened against all the modified H3K9R peptides generated by G9a, heat map shown in figure 28. The data for F50A shows strong binding affinity for H3K9-pentene and H3K9-benzyl (K_{dS} of F50A are 14µM and 19µM respectively). We believed that extending the hole slightly in CBX1 might lower the steric clash between the CBX1 F50 mutant and benzyl peptide. We generated F50G mutant and screened it against our library of peptides. The binding data has improved for all the bulky peptides including the native trimethyl

modification. The binding for F50G with H3K9benzyl is the strongest (K_d of 7.2 µM) and closest to the native system (1.7 µM). The Y26A mutation did not bind to the bumped peptide modifications like F50A and F50G. The data may be explained by the orientation of F50 and Y26 in the binding pocket with respect to H3K9-R as shown in figure 24. The F50 is oriented on the top of H3K9-R whereas Y26 is oriented parallel in the aromatic cage. Therefore, introducing a hole on the top allows the bulky modifications to protrude out from the cage.

	N+	NH		NH	NH	NH	NH
	Me3	Allyl		Butene	Butyne	Pentene	Benzyl
WТ		2	120	132	350	450	350
Y26A		73	200	1000	1000	1000	1000
W47A	1	000	1000	1000	1000	1000	1000
F50A		63	120	180	75	14	19
D54A		9.1	180	200	300	1000	1000
T56A		17	140	500	500	1000	1000
F50G		53	147	55	46	14	7.2
Y26FF50G		74	160	19	68	11	5.5
Y26SF50G		550	430	230	1000	1000	150
Y26TF50G	1	000	1000	250	1000	250	200
					1	000 K _o	_d in μM 1

Figure 28. Screening the CBX1 Mutants Against the Bumped Peptides for Binding. Heat map of the dissociation constant generated by conducting the FP assay CBX1 mutants with the bumped peptides (generated from G9a methyltransferase assay). The higher the K_d the weaker the binding is reflected by the darker blue color in the heat map diagram.

Based on the primary screening results and data presented by Baril *et al*⁹¹, where they demonstrate that Y26F contributes more towards cation-pi interaction than F50, we envision getting better binding constant and orthogonality by further extending the CBX1 binding pocket. Hence, we constructed three more mutants all carrying F50G mutant along with slight restructuring of the Y26 residue. We generated Y26S/F50G, Y26F/F50G, and Y26T/F50G double mutants by site directed mutagenesis and confirmed by sequencing. It was followed by over-expression and purification of protein via affinity chromatography and size-exclusion chromatography. All the proteins were confirmed by LCMS.

The newly expressed and purified proteins were screened in FP based binding assay against the H3K9R peptides. We observed that the binding affinities were improved for most of the H3K9R analogues with Y26F/F50G but the orthogonality was improved from 53 μ M for F50G to 74 μ M for Y26F/F50G. Similarly, for H3K9-benzyl peptide, the binding affinity shifted to 5.5 μ M for Y26F/F50G from 7.2 μ m for the F50G mutant. This validated our hypothesis that extending the binding pocket of CBX1 would accommodate the bulkier modifications peptide mostly through improved steric complementarity.

Isothermal titration calorimetry (ITC) measures the heat being released or absorbed during a binding event, independent of fluorophore. The instrument consists of two cells, one being the sample cell that contains the protein (in our case) and the other cell is known as the reference cell (contains the buffer). Same temperature for both the cells is maintained by the instrument. The ligand or the binding partner (peptide) is injected into the sample cell, on binding of the ligand to the protein there are heat changes in the sample cell. The difference in the temperature of the two cells is measured for each injection to generate a binding isotherm. This technique provides

thermodynamic parameters (entropy, enthalpy and free energy of binding) as well as binding constant and stoichiometry of reaction. We used this technique to validate the binding data from our FP assay. We were not able to use this technique for screening as it requires a high amount and concentration of ligand that is difficult to get through the G9a methyltransferase assay.

For ITC validation, Dr Sinan Wang, a Post-doctoral fellow in Islam lab, synthesized the N-ε-(boc,benzyl)-N-α-Fmoc lysine. This was used to synthesize the H2N-ARTKQTARK(Benzyl)STGGKA-CONH2, for our assay. We also synthesized H3K9me3 H2N-ARTKQTARK(Me3)STGGKA-CONH2 as control.

We first measured the K_d for the native system (CBX1 and H3K9me3) and compared it to K_d value from FP assay. The K_d for the native system was almost 5.3 µM, the difference in binding can be attributed to the difference in the buffers used in two different techniques. We then conducted the assay with CBX1- H3K9benzyl, CBX1-Y26F/F50G- H3K9me3 and CBX1-Y26F/F50G-H3K9benzyl pairs. The K_d values as well as the thermodynamic parameters are reported in figure 29. The dissociation constant for the engineered system and the native system is close to the values generated by FP assay, thereby validating our binding data.

Apart from the CBX1 Y26F/F50G we also validated the data for Y26F/F50A for H3K9me3 and H3K9-benzyl. The thermodynamic parameters are reported in figure 29. This data was generated to gain more confidence/ reliability in the binding data for engineered system.



Figure 29. Validation of the Binding Data by ITC. Top panel illustrates the schematic of generating the binding isotherm from an ITC experiment. The bottom panel gives the binding constant for each CBX1- H3K9-R pair along with the thermodynamic parameters generated. N is the number of binding sites.

From the ITC data we gained confidence in the binding affinity of our engineered system of CBX1-Y26F/F50G and H3K9-benzyl peptide to be equivalent to CBX1- H3K9me3 system. Our next concern was whether the binding of H3K9-benzyl in CBX1* affected the conformation of binding pocket as well as the overall folding/ structure of CBX1. We sought to solve the crystal structure to address these concerns.

3.2.4. Determining the Structure of Bump-Hole Engineering

Protein structure can be solved by nuclear magnetic resonance (NMR) spectroscopy, X-ray diffraction (XRD) and cryogenic electron microscopy (cryo-EM). The resolution of structure solved by cryo-EM is low so we were inclined towards NMR and XRD. As the size of CBX1 protein is low (10 kDa) both NMR and XRD were equally favorable for solving the structure of our system. We gathered that the crystal structure for the wild type CBX1 with H3K9me3 was not solved previously and only the NMR structure has been solved. In collaboration with Dr. Prof. Seth Horne, Department of Chemistry, University of Pittsburgh, we solved the structure of protein-peptide complex for the wild-type CBX1-H3K9me3 and CBX1*- H3K9benzyl.

Protein	Expected M. Wt	LCMS Determined mass
Uncleaved CBX1-WT	8624.58	8624.18
Cleaved CBX1-WT	6488.29	6487.51
Uncleaved CBX1- Y26FF50G	8518.46	8518.17
Cleaved CBX1-Y26FF50G	6382.16	6381.23

Table 3. LCMS Data. Table shows the LCMS data collected for TEV cleaved and uncleaved proteins compared to the expected LC-HRMS data.

The first step towards XRD is generation of protein crystals of a purified protein. A crystal consists of atoms, group of atoms, molecules, or a group of molecules arranged in a periodically repetitive pattern in three dimensions. As proteins have both structured and disordered regions, obtaining crystals of proteins with disordered regions is difficult. The N-terminal LEEEEE patch and the 6-His tag are not structured and thus bring disorder to our system, to address that we used the original

Addgene plasmid (without LEEEEE) and cleaved the 6-His at the TEV-cleavage site. The TEV cleavage of both the wild-type and engineered construct was monitored through LCMS. Any uncleaved protein was removed through Ni-NTA beads, followed by purification of protein (9kDa) from TEV (27kDa) by size exclusion chromatography. The purity was re-confirmed by LC-HRMS table 4.

Protein crystals are grown by slow and controlled precipitation from an aqueous solution, to maintain the structure of proteins. A precipitant is added to the aqueous protein solution to precipitate the proteins. On evaporation of water from the protein and precipitant solution, their concentration reaches the supersaturated condition where nucleation is followed by crystal growth. The precipitant can be a salt like sodium chloride or ammonium sulfate or a hydrophobic molecule such as polyethylene glycol. Slow precipitation from few nucleation produce large crystals whereas fast precipitation or many nucleation produce results in small crystals or amorphous solids.

We employed hanging drop method to generate the crystals for CBX1 wild-type following the reported crystallization conditions for CBX3-H3K9me3⁹². Upon optimizing around the crystallization conditions reported and changing the protein concentration in the drops, we saw precipitation of the protein. We reasoned that the formation of complete structure on binding to the H3K9me3 is generating the crystals and the disorderness in the protein is leading to it crashing out of the drops. The drops with the peptide that formed crystals were reproducible. Similar steps were followed for the CBX1 Y26F/F50G with H3K9benzyl. The crystals were cryo-protected and shot by the X-ray beams for diffraction data, analyzed and solved by Dr Horne. The native structure (PDB 6D07) was solved by molecular replacement with a model derived from a structure of CBX5

bound to H3K9me3. The structure of mutant (PDB 6D08) was solved by using the refined model of the native CBX1 complex.

On superimposing the native CBX1 with H3K9me3 on CBX5 we observed that overall fold and details of the binding interactions were virtually identical figure 30. Conserved structural features include an extended β -strand conformation adopted by the ligand, two 'polar fingers' (E25 and D41) that envelop the histone tail around Thr6, and the aromatic cage around the substrate - methylated Lys side chain. Superimposition of the backbone coordinates for the bump-hole mutant complex on the native complex gave an RMSD of 0.6 Å C α , confirming the overall structural integrity of the engineered system. Further, figure 30 revealed that F50 of wild type CBX1 undergoes steric interference with the benzyl group of H3K9bn, explaining the observed orthogonality. The aromatic binding pocket of the engineered system shows that the benzyl bump effectively fills the hole created by the F50G mutation (figure 30) and that the removal of the side chain in CBX1 has no measurable effect on local backbone conformation. Finally, it can be concluded that the engineered system preserves the overall fold of the chromodomain only perturbing the binding pocket by generating a hole to accommodate the bulky aromatic residue.



Figure 30. Crystallographic Evidence for the Structural Preservation of Engineered System. The topmost panel is the superimposition of CBX1 (PDB 6D07) in blue and CBX5 (PDB 3FDT) in pink with the zoomed-in view of the polar fingers. The middle panel is a superimposition of wild-type CBX1-H3K9me3 (orange) interaction and engineered CBX1 Y26F/F50G (in yellow)-H3K9benzyl (in green) interaction with the individual binding pockets shown in the bottom panel.

3.2.5. Generality of CBX1 and H3 Engineering Approach to Other Reader Proteins

We next sought to expand the engineering strategy to other methyllysine readers. Sequence and structure analyses of methyllysine readers revealed an aromatic residue, equivalent to F50 in CBX1, is present in all the readers analyzed (figure 31). We compared the structure of CBX3 and CBX5 to CBX1 and realized that F50 residue in CBX1 is comparable to F44 in CBX3 and F45 in CBX5. Conveniently, CBX3 already carries a phenylalanine at the position equivalent to Y26 in CXB1 but we had to mutate Y21 to phenylalanine in CBX5 to generate equivalent hole to CBX1. FP assay was employed to measure the dissociation constant of CBX3 and CBX5 for wild-type and engineered system. We measured the binding affinity of CBX3 and CBX5 to H3K9me3 and compared them to the reported values ⁸⁴. Wild-type CBX3 recognized H3K9me3 peptide effectively but no binding was observed towards the benzylated analogue. In contrast, CBX3 F44G exhibited enhanced affinity for 'bumped' histone peptide but insignificant towards the trimethyl mark, figure 31. Similar results were observed for CBX5 as well, figure 31. Collectively, these results demonstrate that multiple histone writers and readers can be engineered at potential gatekeeper sites to establish orthogonal protein-protein interface suitable for studying methyllysine-dependent processes.



Figure 31. Transfer of CBX1*-H3K9benzyl to CBX3 and CBX5. The top left panel shows the binding affinity of CBX3 for H3K9me3 (red) and H3K9benzyl in blue compared to the hole modified CBX3-F44G on the top right panel. In the bottom left panel shows the binding of CBX5 to H3K9me3 in red and H3K9benzyl in blue compared to CBX5 Y21F F45G mutant in the bottom right panel. The biding data was collected by conducting the FP assay on TAMRA labeled peptide carrying the modifications generated by G9a/G9a* methyltransferase assay.

3.2.6. Probing the Binding of Engineered System in Biological Environment

To further investigate whether the engineered protein-protein interface is functional in a complex biological environment we designed the pull-down experiment from cell-lysates employing modified peptide with biotin as bait. A schematic of the experiment is shown in figure 32. Human embryonic kidney HEK293T cells were cultured to express full-length CBX1 variant carrying the engineered chromodomain and a self-aggregating chromo-shadow domain. Biotinylated histone peptides, each carrying either unmodified, methylated or benzylated lysine, were first generated with G9a assay. The peptides were incubated with avidin beads and then incubated with the cellular extracts. The most critical step in this pull-down is washing with appropriate concentration of salt (NaCl). The salt concentration differentiates between different binding affinities for non-covalent interactions. Western blot of the elution revealed a highly specific interaction between the engineered pair. The trimethylated peptide pulled-down all the three wildtype CBX proteins and the benzylated peptide was specific for hole-modified CBX1 figure 32. To establish the specificity of our system we employed biotinylated peptides carrying methyllysine at a site different from H3K9. These peptides failed to interact with the CBX1 mutant, confirming the ability of the

engineered system to mediate allele-specific histone-reader interactions under biologically relevant conditions figure 32.



Figure 32. Pull Down of Native or Engineered System from Nuclear Extracts. The top panel illustrates the experimental design for the pull-down of CBX1 Y26F/F50G (CBX1*) or CBX1 by using biotinylated peptide with trimethyl or benzyl modification as bait. The bottom panel shows the western blot data for the pull down from cellular extracts of HEK293T cells either w/wo CBX1* transfection with designated antibodies.

			CBX binding	
CBX Protein	Binding Protein	Model/cell lines/	domain	Functional Implication
CBX1/CBX3/CBX5	H3-K9me2/3		CD	Repression
CBX1/CBX3/CBX5	SUV39		CSD	Repression
		Drosophilla		
CBX1/CBX5	CAF1 (chromatin assembly factor1)	melanogaster	CSD	Repression
CBX1/CBX5	CHAF1A (chromatin Assembly factor 1)	Homo sapiens	CSD	Repression
		Drosophilla		
HP1	Vig(vasa intronic gene)	melanogaster		RISC
CBX1	DNMT3a	Homo sapiens		Repression
CBX5	PEP (peptide on Ecdysone Puffs),DDP1, HRB87F	Drosophilla melanogaster		Ribonucleoproteins (RNA processing)
CBX1/CBX3/CBX5	МКІ67	Homo sapiens		ATP binding, nucleotide binding
				Townes-Brocks syndrome 1 (TBS1); DNA
CBX1 (M31)	SALL1	mouse		binding, Repression
CBX1/CBX3/CBX5	TRIM28 (tripartite motif containing 28), also known as :KAP1; TF1B; RNF96; TIF1B; PPP1R157	Homo sapiens	CSD	
CBX1/CBX3/CBX5	καρ1/τιγβ	Homo sapiens		co-repressor for the large family of KRAB domain-containing zinc finger proteins
				member of the SNF2 family of
CBX1(M31)	ATRX/HP1-BP38	mouse		helicase/ATPases th
	zinc-finger proteins WOC (without children)	Drosophilla		
CBX1/CBX3/CBX5	and Relative-of-WOC (ROW)	melanogaster		gene expression
HP1	BMI1	Mice		PRC1
CBX5	ORC1. ORC3			ORC
		Drosophilla		
HP1	HOAP(HP1 ORC associated Protein)	melanogaster		ORC
CBX1	PRDM6	melanogaster		
00/12	BNA hinding protein-2 (Bhp2)			
	Fragile X mental retardation syndrome	-		
CBX1	related protein 1 (fmr1, a component of RISC)	Drosophilla melanogaster		
	TCTP (Translationally controlled tumor protein)			
	ADNP(Activity-dependent neuroprotector homeobox protein)	Human		
CBX1	BARDI (BRCAI-associated RING domain protein 1) CHD3/4 (Chromodomain-helicase-DNA-			
	binding protein 3) IF2P HUMAN Eukaryotic translation			
	initiation factor 5B SIRT7 HUMAN NAD-dependent			
	deacetylase sirtuin-7 SMCA4 HUMAN Transcription activator			
	BRG1 TIF1B HUMAN Transcription intermedian			
	factor 1-beta			
	ZN828 HUMAN Zinc finger protein 828			
	Ber protein 020			

Table 4. List of Interacting Partners of CBX1. To select the protein to be pulled downby CBX1- chromoshadow domain we generated the table of known interacting partnersfrom reported references.

3.2.7. Determining the Functionality of the Engineered System in Biological Environment

Although we solved the crystal structure with chromodomain of CBX1 and established the overall conformation of the domain is unperturbed on engineering the binding pocket, we did not establish the functionality of full-length CBX1 including chromoshadow domain is unaffected by the engineering. To address this, we employed the biotin pull-down assay to pull down the proteins that interact with the chromodomain of CBX1. As the interaction between biotin peptide with modification and CBX1 is primary interaction (first sphere of interaction) the interaction between chromo-shadow domain of CBX1 and its binding partners is the second sphere of interaction. Identification of the binding partners have been taken up by several groups earlier ^{81, 63, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104}. Based on present literature table 4 we identified three binding partners of CBX1, expected to bind via the chromoshadow domain, namely TIF1b, CAF1a and DNMT3a.

Nuclear receptors are transcription factors which mediate functions like development, differentiation and homeostasis upon binding to respective ligand like the intermediary factors (mediators). Transcription intermediary factor 1-beta (TIF-1 β) also known as KAP1 or TRIM28 is a transcription corepressor. It associates with Kruppel-associated box domain-zinc finger proteins (KRAB-ZFPs) to repress genes. It is also known to associate with SETDB1 (histone methyltransferase), HP1 or the NuRD histone deacetylase complex ¹⁰⁵. Its binding to HP1 was first identified in 1996 through the chromoshadow domain of CBX1 ¹⁰⁶. It was pulled down with CBX1 by Rosnoblet *et al* ¹⁰⁴ as well. Based on literature we speculated that our engineered system should pull-down TIF1 as an interacting partner via its chromoshadow domain same as the native system conserving the functionality of the protein. To our biotin pull-down experiment, we loaded more nuclear extracts and observed the pull-down of TIF1b with H3K9me3 and H3K9benzyl peptides,

figure 33b. This concluded that the functionality of CBX1 towards TIF1 β is unaffected by the engineering in the binding pocket at chromodomain of CBX1. To show robustness of our conclusion, we set out to pull down another binding partner of CBX1 with our engineered system, CAF1a.

Chromatin assembly factor 1 alpha (CAF1a) is another binding partner of CBX1. During the cell cycle, DNA is replicated and equally distributed amongst the daughter cells similarly faithful transfer of epigenetic component of chromatin state is also necessary. CAF1a is involved in establishment and maintenance of these states. p150, a subunit of CAF1a, interacts with PCNA, HP1, methyl binding protein 1 (MBP-1) and Bloom's syndrome DNA helicase (BLM). We wanted to establish the preservation of functionality of engineered system by pulling down CAF1a with



Figure 33. Second Sphere of Interacting Partners Pull-Down. The top panel a. illustrates the experiment for the pull-down of second sphere of interacting partner pull-down. The bottom panel is the western blot data for pull down of TIF1b, CAF1-p150 and DNMT1 confirmed by respective antibodies. Same membrane was blotted for anti-CBX1 as a control for successful pull-down of first sphere of interacting partners pull-down.

both native and engineered system. Figure 33c shows that H3K9me3 with native CBX1 and H3K9benzyl with CBX1 Y26F/F50G were able to pull-down the CAF1a compared to our negative controls. This further enhanced the confidence in employing our system to the cells without perturbing the functionality of the protein.

DNMT1 is also known as a binding partner of CBX1 important for its function, but has never been directly pulled down with CBX1 *in vivo*. Either the binding is too weak and transient to be pulled down or the it is indirect. To assess the binding using our pull-down experiment, we first tried to pull down DNMT1 with H3K9me3-biotin peptide and CBX1 system. We did not pull down DNMT1 although input (nuclear extracts without any pull down) blotted against DNMT1 antibody, figure 33d. The experiment was repeated to be confident of the results. To address the issue of weak interaction between DNMT1 and CBX1 we loaded the nuclear extracts on the



Figure 34. Troubleshooting DNMT1 Pull-Down. An increasing amount of nuclear extracts were loaded on the biotinylated peptide beads to pulldown DNMT1 as determined by western blotting. Anti-DNMT1 blot shows presence of DNMT1 in input as a control and an increase in CBX1 being pulled-down with increasing nuclear extracts. The blot with anti-CAF1 confirms the pull-down of second-sphere of interacting partners with CBX1.

H3K9me3-biotin beads as an increasing gradient. The modified peptide was still not able to pulldown DNMT1, the input showed a signal for the presence of protein in the nuclear extracts. Also, the elution from the same experiment were blotted against CAF1 to confirm the pull-down was working. Figure 34 shows the data, confirming that the pull-down was working, the nuclear extracts contained the DNMT1 protein but CBX1 was not able to pull-down DNMT1. This was most probably due to transient and weak nature of interaction between DNMT1 and CBX1 or indirect interaction between them.

As we already had the engineered system to benzylate H3 in cells, we wanted to establish the binding of engineered CBX1 with H3K9benzyl *in-vivo*.

3.3. CONCLUSIONS AND FUTURE DIRECTIONS

Generating an allele specific chemical genetics at protein-protein interface is challenging endeavor as it requires engineering a dynamic and expansive region as well as modifying one of the proteins with a bulky orthogonal group *in vivo*. Here we developed the bump-hole pair for choromodomain containing CBX1 protein and H3K9. We have established biochemical integrity of binding for the engineered interface, along with structural rationale for orthogonality. We have also demonstrated the potential generality of the approach to other methyllysine writers and validated functional compatibility of the synthetic interface in recognizing transcriptional regulators.

3.4. MATERIALS AND METHODS

3.4.1. CBX1, CBX3 and CBX5 Protein Expression and Mutagenesis

Human CBXs (chromodomain) for bacterial expression were bought from Addgene and transformed in E.coli One Shot BL21-star (DE3) chemically competent cells (Invitrogen) for expression. A single colony was picked and grown overnight in 10 mL of Luria-Bertani (LB) at 37°C in presence of 50 µg/mL of kanamycin. The 10 mL culture was diluted to a 100-fold and the cells were grown at 37°C in presence of 50 µg/mL of kanamycin till it reached an optical density (O.D) of 0.8. The protein over-expression was induced by adding 0.5 mM of IPTG to cultures and grown at 17°C overnight. The cells were harvested and resuspended in buffer containing 50 mM Tris (pH 8.0), 150 mM NaCl, 10 % glycerol, 5 mM β-mercaptoethanol, 25 mM imidazole, DNase, lysozyme and protease inhibitor tablets. The cells were lysed by pulsed sonication (Qsonica-Q700) and centrifuged at 13,000 rpm for 50 mins at 4°C. The soluble extracts were incubated with Ni-NTA agarose resin (Thermo) according to manufacturer's protocol. The beads were washed with 20 column volumes of wash buffer (50 mM Tris pH 8.0, 150 mM NaCl, 10 % glycerol, 5 mM βmercaptoethanol and 25 mM imidazole). The protein was eluted with 50 mM Tris pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol and 400 mM imidazole. Eluted protein was subjected to further purification by gel filtration chromatography (Superdex-75) using AKTA pure FPLC system (GE healthcare) with buffer containing 50 mM Tris pH 8.0, 150 mM NaCl and 10 % glycerol. The purified protein was concentrated using Amicon-Ultra 3K centrifugal filter device (Merck Millipore Ltd.). The concentration of the protein was determined using Bradford assay kit (BioRad Laboratories) with BSA as standard.

Mutagenesis was performed using QuikChange Lightning Site Directed Mutagenesis Kit (catalog # 210519) following manufacturers protocol. The PCR amplified DNA was transformed in XL-10 Gold cells. Single colonies were picked for mini-prep using GeneJET Plasmid Mini-Prep Kit (Catalog # K0503) and sent for sequencing to Genewiz. All mutagenesis was carried out following the same method.

3.4.2. G9a Protein Expression and Mutagenesis

The N-terminal 6-His tagged methyltransferase SET domain of G9a-WT and mutant plasmids as well as SUV39h2-WT and its mutants were a gift from Dr. M. Luo's lab at Memorial Sloan-Kettering Cancer Center. The plasmids were expressed in *E.coli* cells. In general the plasmid was transformed in *E.coli* One Shot BL21-star (DE3) chemically competent cells (Invitrogen) for expression. Single colony was picked and grown overnight in 10 mL of Luria-Bertani (LB) at 37°C in presence of 50 µg/mL of kanamycin. The 10 mL culture was diluted to a 100-fold and the cells were grown at 37°C in presence of 50 µg/mL of kanamycin till it reached an optical density (O.D) of 0.8. The protein over-expression was induced by adding 0.5 mM of IPTG, in the presence of 25 µM ZnCl₂, and grown at 17°C overnight. The cells were harvested and resuspended in buffer containing 50 mM Tris (pH 8.0), 150 mM NaCl, 10 % glycerol, 5 mM β-mercaptoethanol, 25 mM imidazole, DNase, lysozyme and protease inhibitor tablets. The cells were lysed by pulsed sonication (Qsonica-Q700) and centrifuged at 13,000 rpm for 50 mins at 4°C. The soluble extracts were incubated with Ni-NTA agarose resin (Thermo) according to manufacturer's protocol. The beads were washed with 20 column volumes of wash buffer (50 mM Tris pH 8.0, 150 mM NaCl, 10 % glycerol, 5 mM β -mercaptoethanol and 25 mM imidazole). The protein was eluted with 50 mM Tris pH 8.0, 150 mM NaCl, 5 mM β -mercaptoethanol and 400 mM imidazole. Eluted protein was subjected to further purification by gel filtration chromatography (Superdex-75) using AKTA pure FPLC system (GE healthcare) with buffer containing 50 mM Tris pH 8.0, 150 mM NaCl and 10 % glycerol. The purified protein was concentrated using Amicon-Ultra 10 K centrifugal filter device (Merck Millipore Ltd.). The concentration of the protein was determined using Bradford assay kit (BioRad Laboratories) with BSA as standard.

3.4.3. Peptide Purification

All the H3 peptides were synthesized at the University of Pittsburgh Peptide Synthesis facility. Analytical scale separation was performed using Zorbax reversed phase C18 (5 μ m, 4.6 x 250 mm) column with UV detection at 220 nm and 550 nm (for TAMRA labelled peptide). The column was equilibrated with 0.1% aqueous trifluoroacetic acidic (TFA) and the peptide was separated with the linear gradient of acetonitrile to 10% in 15 min and then to 70% in 5 min with a flow rate of 1 mL/min. The crude peptides were purified using preparative reversed phase HPLC column Xbridge C18 (5 μ m, 10 x 250 mm) eluting with a flow rate of 5 mL/min and a acetonitrile gradient from 0% to 90% in 15 mins and then to 100 % in 18 mins in 0.1% aqueous TFA. The purified peptides were concentrated on speedvac and dried in the lyophilizer. The dried peptides were resuspended in 0.1% aqueous TFA and the concentration was measured based on the observation that 1 mg/mL of peptide generates an absorbance of 30 at a wavelength of 205 nm. The concentration of TAMRA-labelled peptide was determined by measuring the absorbance at a

wavelength of 555nm with an extinction coefficient of 65000 Lcm⁻¹mol⁻¹. The integrity of the peptides was confirmed by MALDI mass spectrometry.

3.4.4. G9a Methyltransferase Assay

Trimethylated H3K9 peptide was generated by incubating 2µM of wild-type enzyme with 20 µM of peptide and 80 µM SAM in 50 mM Tris (pH8.0) for 30-40 mins at room temperature. For the bumped modifications 5 µM G9a-Y1154A or G9a-Y1154G mutant was incubated with 20 µM peptide and 200 µM SAM-analogue in 50 mM Tris (pH8.0) for 30-60 mins at room temperature. The methyltransferase activity was observed using MALDI mass spectrometry. 1 µL of the assay volume was mixed with 1 µL of saturated solution of α -cyano-hydroxy-cinnamic acid CHCA (Fischer # AAJ67635EXK) on the MALDI target plate. The sample was analyzed using ultrafleXtreme (Bruker) and analyzed on flexAnalysis. The negative control included all components of the assay except the SAM analogue. Upon confirmation of the assay completion the peptides were purified using the C18 Sep-pak column (Waters, cat# WAT054955). The peptides were concentrated on speedvac and resuspended in 0.1% aqueous TFA.

3.4.5. FP Assay

Fluorescence polarization assay was employed to screen the binding affinity between the holemodified CBX1 mutants and the bumped H3 peptides generated by G9a methyltransferase assay. The assay was performed in a 384 well small volume black/clear microtiter plates (Falcon) with
200 nM TAMRA labelled peptides and varying concentrations of protein (0.3-750 μ M) in 10 mM Tris pH8.0, 150 mM NaCl, 0.05% Tween 20 and 0.5 mM of Tris (2-carboxyethyl) phosphine. The plate was centrifuged at 3000 rpm for 3mins and the polarization was read for each well on TECAN M1000 plate reader with an excitation at 530 nm and emission at 570 nm. For determination of dissociation constants (K_d), the background corrected fluorescence polarization was plotted against the concentrations in μ M. The data was fitted to single site binding equation Y= $B_{max}*X/(K_d+X)$, where Y is the specific binding, B_{max} is the maximal binding and X is the concentration of ligand, using SigmaPlot software. The mP values for various concentrations were then divided by the highest mP value to get fraction bound as plotted in the graph.

3.4.6. ITC Assay

The ITC measurements were carried out at 25°C using ITC₂₀₀ (Microcal, Malvern). Both protein and peptide were dissolved in 50 mM Tris-HCl pH 8.0, 200 mM NaCl and 10% glycerol. The sample cell contained the protein and the syringe had a 10-fold higher peptide concentration, individual concentrations for each experiment are given in table #. Each experiment comprised of 0.4 μ L of first injection followed by 19 injections of 2 μ L at a spacing of 120s with a stirring speed of 750 rpm. The data points were fitted to a single binding site model using the Microcal ITC₂₀₀ Origin data analysis software.

3.4.7. Crystal Structure, Protein Expression and Purification

We used the original construct from Addgene of CBX1 chromodomain to express the wild-type protein for determination of their crystal structure as the disordered acidic patch would hinder the generation of ordered crystals. The wild type protein was mutated to the double mutant CBX1 Y26FF50G, expressed and purified following the same protocol as described above. After expression the proteins were incubated with TEV protease to cleave the 6His-tag from the N-terminal. The SELECTEV Protease from VWR (cat# 30810-1) and the assay was conducted as per manufacturer's protocol. The reaction progress was monitored visa LCMS. Ni-NTA agarose beads were used to remove the cleaved His-tag as well as the TEV protease. The cleaved protein in the unbound supernatant was further purified through size-exclusion chromatography (Superdex-75) with 50 mM Tris (pH 8.0) and 200 mM NaCl. The purified protein was concentrated to 8 mg/mL for setting up the crystal trays.

3.4.8. Crystallization

Crystals of the native CBX1 / H3 ligand complex and of the bump-hole mutant complex were grown by hanging drop vapor diffusion. For the native complex, a stock solution consisting of 6.4 mg mL⁻¹ CBX1 and 7.2 mg mL⁻¹ H3K9me3 peptide ligand (ARTKQTARXSTGGKA-NH₂; X = N_{ϵ} , N_{ϵ} -trimethyl-lysine) in 0.2 M NaCl, 0.05 M Tris pH 8.0 was mixed (0.7µL + 2.0µL) with a well buffer composed of 0.2 M NaCl, 0.1 M Tris pH 8.0, 30% w/v PEG 3350, and the resulting drop allowed to equilibrate at room temperature over a well containing 0.7 mL crystallization 93 buffer. For the bump-hole mutant complex, a stock solution consisting of 6.3 mg mL⁻¹ CBX1(Y26F/F50G) and 7.6 mg mL⁻¹ H3K9bn peptide ligand (ARTKQTARXSTGGKA-NH₂; X = N_{ϵ} -benzyl-lysine) in 0.2 M NaCl, 0.05 M Tris pH 8.0 was mixed (0.7µL + 1.2µL) with a well buffer composed of 0.1 M citrate pH 5, 3M (NH₄)₂SO₄, and the resulting drop allowed to equilibrate at room temperature over a well containing 0.7 mL crystallization buffer.

A single crystal of each complex was flash frozen in liquid nitrogen after cryoprotection in well buffer containing 30% v/v glycerol. X-ray diffraction data were collected using Cu/K_{α} radiation on a Rigaku/MSC diffractometer (FR-E generator, VariMax optics). A Saturn 944 CCD detector was used for data collection on the native complex and a RAXIS HTC image plate detector for the bump-hole mutant complex. Crystals were maintained at 100 K during data collection with an X-Stream 2000 system. Raw diffraction data were integrated, merged, and scaled with d*TREK.

Both structures were solved by molecular replacement using the *Phenix* software suite. The search model for the native complex was derived from a published structure of the related protein CBX5 in complex with a H3K9me3 peptide ligand (PDB 3FDT). The resulting refined coordinates for native CBX1 were used to solve the structure of the bump-hole mutant pair. Model refinement was accomplished using *Phenix* and *Coot*. Final data collection and refinement statistics are given in Table S##. Coordinates and reflection files are deposited in the PDB under accession codes 6D07 (native) and 6D08 (bump-hole mutant). Analysis and validation of the refined coordinates using the MolProbity server indicated both structures scored in the 100th percentile based on comparison to representative structures of comparable resolution in the PDB.

3.4.9. Biotin Pull Down Experiments

Biotin conjugated histone3 peptides with various modifications (un-methylated, tri-methylated, pentenylated and benzylated) were incubated with streptavidin beads (20μ L), equilibrated in 50 mM Tris pH 8.0, at 4°C for an hour. The excess peptide was washed-off with buffer-I (five times, 100uL each), 150 mM NaCl, 50 mM Tris pH 8.0, 0.1% Triton X-100 and 1x-protease inhibitor tablet. The streptavidin-biotin peptide was incubated with 200 µg of nuclear extract overnight at 4°C. Non-specific interactions to the peptide were removed by washing the samples five times with 100 µL of buffer-II, 400 mM NaCl, 50 mM Tris pH 8.0, 0.1% Triton X-100, and 1x-protease inhibitor tablet. Samples were eluted with 4x Dye and heating at 95°C for 5 mins.

For the second sphere of interacting partner pull-down 500 ug of nuclear extracts were used to incubate with the streptavidin-biotin peptide samples and the buffer-II was 400 mM NaCl, 50 mM Tris pH 8.0, 0.1% Triton X-100, and 1x-protease inhibitor tablet, to wash-off non-specific interactions.

3.4.10. In-vivo Plasmid and Protein Expression

For mammalian expression the plasmid was bought from Addgene (plasmid # 24079). It was mutated to CBX1-Y26FF50G following the protocol described earlier. Human embryonic kidney HEK293T cells were grown in Dulbecco modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum in a humidified atmosphere containing 5% CO₂. The HEK293T cells express high levels of endogenous CBX1 so we did not have to transfect the cells with CBX1-WT plasmid for its over-expression. CBX1 mutant was transfected in the cells at 80% confluency with the aid of lipofectamine 2000. 24 hours post transfection, the cells were uplifted using trypsin and collected for lysis with cold RIPA buffer (Sigma) supplemented with 1X Roche protease inhibitor cocktail and 5mM TCEP. The cells were further lysed by sonicating for 15 mins at an amplitude of 100Amps, with a repeating 1 min pulse cycle at 4°C. To remove cell debris the lysates were centrifuged at 12,000 rpm for 30 mins at 4°C. The supernatant was passed through detergent removal spin column (Pierce # 87778) and eluted with Tris buffer (50 mM Tris-HCl pH 8.0, 10% glycerol, 2mM TCEP and 1X Roche protease inhibitor cocktail) following manufacturer's protocol. Protein concentration was measured using Bradford Assay (Bio-Rad Laboratories).

3.5. FIGURES AND TABLES

Mutation	Primer (5'-3')			
Bacterial Construct CBX1				
LEEEEE_Fwd	CTT GTA TTT CCA GGG CCT GGA GGA GGA GGA GGA GGA ATA TGT GG			
LEEEEE_Rev	CCA CAT ATT CCT CCT CCT CCT CCA GGC CCT GGA AAT ACA AG			
Y26A_Fwd	GGA GGA GGA AGC CGT GGT GGA AAA AGT TCT CG			
Y26A_Rev	CGA GAA CTT TTT CCA CCA CGG CTT CCT CCT CCT CC			
W47A_Fwd	GAG TAC CTC CTA AAG GCC AAG GGA TTC TCA GAT			
W47A_Rev	ATC TGA GAA TCC CTT GGC CTT TAG GAG GTA CTC			
F50A_Fwd	CTA AAG TGG AAG GGA GCC TCA GAT GAG GAC AAC			
F50A_Rev	GTT GTC CTC ATC TGA GGC TCC CTT CCA CTT TAG			
D54A_Fwd	GGA TTC TCA GAT GAG GCC AAC ACA TGG GAG CCA			
D54A_Rev	TGG CTC CCA TGT GTT GGC CTC ATC TGA GAA TCC			
T56A_Fwd	CTC AGA TGA GGA CAA CGC CTG GGA GCC AGA AGA			
T56A_Rev	TCT TCT GGC TCC CAG GCG TTG TCC TCA TCT GAG			
F50G_Fwd	CTA AAG TGG AAG GGA GGC TCA GAT GAG GAC AAC			
F50G_Rev	GTT GTC CTC ATC TGA GCC TCC CTT CCA CTT TAG			
Y26S_Fwd	GGA GGA GGA GGA AAG CGT GGT GGA AAA AGT TCT CG			
Y26S_Rev	CGA GAA CTT TTT CCA CCA CGC TTT CCT CCT CCT CC			
Y26T_Fwd	GGA GGA GGA GGA AAC CGT GGT GGA AAA AGT TCT CG			
Y26T_Rev	CGA GAA CTT TTT CCA CCA CGG TTT CCT CCT CCT CC			
Mammalian Construct CB	X1			
mCBX1_Y26F_F	GAA GAG GAG GAA GAG GAA TTC GTG GTG GAA AAA GTT CTC GAC CG			
mCBX1_Y26F_R	CGG TCG AGA ACT TTT TCC ACC ACG AAT TCC TCT TCC TCC TCT TC			
mCBX1_F50G_F	CCT CCT AAA GTG GAA GGG AGG CTC AGA TGA GGA CAA C			
mCBX1_F50G_R	GTT GTC CTC ATC TGA GCC TCC CTT CCA CTT TAG GAG G			
Bacterial Construct CBX3				
CBX3 F44G F	CCT GAA GTG GAA GGG AGG CAC AGA TGC TGA			
CBX3_F44G_R	TCA GCA TCT GTG CCT CCC TTC CAC TTC AGG			
Bacterial Construct CBX5				
CBX5_Y21F_F	CCA GGG CGA GGA GTT CGT TGT GGA GAA GG			
CBX5_Y21F_R	CCT TCT CCA CAA CGA ACT CCT CGC CCT GG			
CBX5_F45G_F	CTG AAG TGG AAA GGC GGC TCT GAG GAG CAC			
CBX5_F45G_R	GTG CTC CTC AGA GCC GCC TTT CCA CTT CAG			
Table 5. List of Prin	ners. Primers generated for constructing the mutants on CBX1 (bacterial			

and mammalian), CBX3 and CBX5 plasmids.





Figure 36. MALDI Spectra. Spectra of biotinylated peptides for pull-down experiments. The H3K9me3-biotin and H3K9benzyl-biotin were generated by G9a/G9a* methyltransferase assay.



Figure 37. Dissociation Constant. Dissociation constants of wild type CBX1 towards the TAMRA- H3K9-R peptides carrying various modifications, one at a time, determined by fluorescence polarization values. Error bars represent standard deviation from two independent measurements. For weakly binding pairs we did not observe saturation in mP, the K_d values were determined by Bmax.



Figure 38. **Dissociation Constant.** Dissociation constants of CBX1-Y26A towards the TAMRA- H3K9-R peptides carrying various modifications, one at a time, determined by fluorescence polarization values. Error bars represent standard deviation from two independent measurements. For weakly binding pairs we did not observe saturation in mP, the K_d values were determined by Bmax.



Figure 39. Dissociation Constants. CBX1-W47A towards the TAMRA- H3K9-R peptides carrying various modifications, one at a time, determined by fluorescence polarization values. Error bars represent standard deviation from two independent measurements. For weakly binding pairs we did not observe saturation in mP, the K_d values were determined by Bmax.



Figure 40. Dissociation Constant. Dissociation constants of CBX1-F50G towards the TAMRA- H3K9-R peptides carrying various modifications, one at a time, determined by fluorescence polarization values. Error bars represent standard deviation from two independent measurements. For weakly binding pairs we did not observe saturation in mP, the K_d values were determined by Bmax.



Figure 41. Dissociation Constant. Dissociation constants of CBX1-F50A towards the TAMRA- H3K9-R peptides carrying various modifications, one at a time, determined by fluorescence polarization values. Error bars represent standard deviation from two independent measurements. For weakly binding pairs we did not observe saturation in mP, the K_d values were determined by Bmax.



Figure 42. Dissociation Constant. Dissociation constants of CBX1-D54A towards the TAMRA- H3K9-R peptides carrying various modifications, one at a time, determined by fluorescence polarization values. Error bars represent standard deviation from two independent measurements. For weakly binding pairs we did not observe saturation in mP, the K_d values were determined by Bmax.



Figure 43. Dissociation Constant. Dissociation constants of CBX1-T56A towards the TAMRA- H3K9-R peptides carrying various modifications, one at a time, determined by fluorescence polarization values. Error bars represent standard deviation from two independent measurements. For weakly binding pairs we did not observe saturation in mP, the K_d values were determined by Bmax.



Figure 44. Dissociation Constant. Dissociation constants of CBX1-Y26F/F50G towards the TAMRA- H3K9-R peptides carrying various modifications, one at a time, determined by fluorescence polarization values. Error bars represent standard deviation from two independent measurements. For weakly binding pairs we did not observe saturation in mP, the K_d values were determined by Bmax.



Figure 45. Dissociation Constant. Dissociation constants of CBX1-Y26S/F50G towards the TAMRA- H3K9-R peptides carrying various modifications, one at a time, determined by fluorescence polarization values. Error bars represent standard deviation from two independent measurements. For weakly binding pairs we did not observe saturation in mP, the K_d values were determined by Bmax.



Figure 46. Dissociation Constant. Dissociation constants of wild-type CBX1 towards the H3K9me3 and H3K9benzyl peptides, determined by ITC experiments.



Figure 47. Dissociation Constant. Dissociation constants of CBX1 Y26F/F50G towards the H3K9me3 and H3K9benzyl peptides, determined by ITC experiments.



	Native CBX1-ligand pair	Bump-Hole Mutant CBX1- ligand pair	
Data Collection			
Unit cell	a = 32.0, b = 32.1, c = 34.1	a = 33.2, b = 34.1, c = 33.3	
dimensions (Å, °)	α = 99, β = 106, γ = 102	α = 77, β = 90, γ = 65	
Space group	P1	P1	
Resolution (Å)	26.34–2.10 (2.18–2.10)	25.95–2.10 (2.18–2.10)	
Total observations	12,512	25,664	
Unique observations	6,737	7,149	
Redundancy	1.9 (1.8)	3.6 (3.6)	
Completeness (%)	93.5 (86.7)	90.4 (87.4)	
I/σ	8.5 (2.6)	16.2 (3.8)	
R _{merge} (%)	7.3 (20.8)	4.7 (14.2)	
Refinement			
Resolution (Å)	26.34–2.10	23.28-2.10	
R (%)	21.4	22.9	
R _{free} (%)	25.7	25.2	
Avg. B factor (Å ²)	22.8	38.7	
RMSD			
Bonds (Å)	0.003	0.002	
Angles (°)	0.53	0.45	

Table 6. X-Ray Data Collection and Refinement Statistics.

4. SITE-SPECIFIC PHOTOCHEMISTRY ON EPIGENETIC READERS

4.1. INTRODUCTION

4.1.1. Non-Histone Interacting Partners of Epigenetic Proteins

Epigenetic proteins play a myriad of biologically important roles by writing, reading and erasing post-translational modifications on histone and non-histone proteins. PTMs have profound effects on cellular processes like apoptosis, cell cycle progression, DNA damage and DNA repair apart from transcriptional regulation. There are several reports that show epigenetic proteins to have a broad range of substrate affinity including histones, transcription factors, coactivators, chaperones and hormone receptors, to list a few. For example: p53, a transcription factor, when acetylated at residue 373 by p300/CBP (writer protein with acetyltransferase activity) leads to apoptosis whereas acetylation at 320 position by PCAF (another acetyltransferase enzyme) promotes cell survival after DNA damage ¹⁰⁷. This exemplifies the importance of PTMs as well as the epigenetic regulator in controlling the fate of the cells. Interestingly, demethylase LSD1 has been shown to interact with Rb1 (retinoblastoma protein 1), tumor suppressor protein, to control the stages of cell cycle by immunoprecipitation (IP) experiment. However, IP could not determine the domain of LSD1 that interacts specifically with Rb1 ¹⁰⁸. Novel tools are required to capture role substrates and interacting partners of epigenetic writers, readers and eraser proteins.

4.1.2. Tools Available to Identify the Binding Partners

Traditional approaches to identify the binding partners of reader proteins include peptide based binding assay or pull-down from cellular lysates or co-immunoprecipitation. The reversible nature of post translational modifications together with the transient nature of their interaction with the reader proteins demands new techniques to capture them in their constitutional (cellular) environment. Sudhamalla et. al, devised the IBPP (Interactome Based Protein Profiling) approach to capture the transient interactions between the PTM (acetylated lysine) and reader proteins (BRD4) with temporal precision¹⁰⁹. Their approach involved incorporation of an unnatural photocrosslinkable amino acid in the binding pocket of the reader protein that would form an irreversible covalent bond with its interacting partner specifically on illumination with UV light, figure 49. They were able to establish binding efficiency, substrate specificity and crosslinking efficiency of BRD4 mutants in their biochemical assays. Further, they were able to identify novel binding partners of BRD4 with LCMS/MS from HEK293T cellular lysates. Successful application of a similar strategy was reported in 2011, by Krishnamurthy et. al, where they identified transient interaction between Swi/Snf (nucleosome remodeling complex) and VP16 (a transcriptional activator)¹¹⁰. Here, the incorporation of photo-labile unnatural modification was on VP16 and not on the epigenetic regulator protein. Together IBPP approach enables identification of transient interactions in a temporal manner in native environment. Successful application of IBPP requires incorporation of unnatural photo-crosslinkable amino acid in the protein of interest.

4.1.3. Incorporation of Photo-crosslinkable Unnatural Amino Acid



Efficient, orthogonal and site-specific incorporation of unnatural amino acid (UAA) using the cellular translational machinery in *E.coli* was first achieved by P Schultz^{111,112}. Orthogonality was attained by generating suppressor transfer-RNA (t-RNA) and aminoacyl t-RNA synthetase (aaRS) pair against UAG (amber stop codon) from the archea *Methanococcus jannaschii (Mj)*. The frequency of occurrence of amber stop codon in *E.coli* is about 9% (in non-essential genes) and in humans is about 20% impacting the cellular proteome minimally¹¹³. *Mj*t-RNA^{Tyr} was mutated in the anti-codon loop to CUA to bind to UAG codon on the mRNA followed by two rounds of directed evolution to generate an orthogonal tRNA^{Tyr}/aaRS pair. There are more than 70 UAAs reported now for incorporation of in *E.coli*. p-azido-phenylalanine is a photoreactive UAA. Its crosslinking ability has been utilized for the purpose of crosslinking proteins to study their binding partners^{114,115}. At a wavelength of 254-400 nm the azide group loses the molecular nitrogen

generating the reactive triplet nitrene that reacts with neighboring molecules to form a covalent bond.

4.1.4. Background on Binding Partners of CBX1 Through its Chromodomain

Several groups have reported identification of binding partners of CBX1/ HP1 β and their role in cellular processes. Ryu *et. al* ¹⁰⁰, performed a comprehensive proteomic analysis of CBX1, CBX3 and CBX5 in *Drosophilla melanogaster*, by double affinity (HA-tag and Flag-tag) purification followed by proteomics. Their experiment identified CBX1's interaction with Tudor-SN (endonuclease), fmr1 (fragile X mental retardation 1, involved in connectivity of neurons in brain) and TCTP (translationally controlled tumor protein, regulator of cellular growth and proliferation). These examples illustrate the role of CBX1 in cellular processes apart from transcription regulation. Rosnoblet *et. al* ¹⁰⁴, applied tandem affinity purification followed by mass spectrometry of mammalian HP1 proteins to identify their binding partners from HeLa cells and HEK293T cells. They identified several zinc finger proteins (ADNP2, ZNF8 and ZN552) to be pulled down by their method, with CBX1 protein. The pull-down technique lacks the discrimination of binding between chromodomain and chromoshadow domain.

Liu *et al* ¹¹⁶, developed a method for identification of binding partners of CBX1 by its chromodomain by conducting a pull-down of nuclear extracts from HEK293T cells with the chromodomain of CBX1 covalently linked to the magnetic beads. The binding partners were eluted with H3K9me3 peptide. The samples were analyzed by LC-MS/MS, upon trypsinization. Their technique revealed CBX1 chromodomain to associate with DNA damage repair pathway proteins, 116

proteins that interact with ribosomes, and RNA splicing. Although their method specifically identified the binding partners of CBX1 through its chromodomain it did not identify the binding partners in living cells.

With the aim of identifying the interacting partners of chromodomain of CBX1 in living cells, we applied the IBPP approach to the protein. We would incorporate a photo-crosslinkable UAA in the



binding pocket of chromodomain of CBX1, express the mutant in HEK293T cells, crosslink the

protein to its native binding partners in live cells and then enrich them for proteomic analysis.

4.2. RESULTS AND DISCUSSION

4.2.1. Rationale for Generating Mutants

We selected five residues in the binding pocket of CBX1 to incorporate pAzF that would specifically interact through chromodomain with H3K9me3, figure 51a. We also selected four residues on the surface of CBX1 interacting with H3K9me2 peptide for pAzF incorporation, based on nmr structure PDB 1GUW, as the crosslinking is affected by positioning of pAzF. The selected sites were mutated to TAG amber suppressor codon and confirmed by sequencing. The protein



Figure 51. Site Selection and Validation of Mutants. A. shows the sites selected for UAA incorporation PDB(1GUW). **B** shows the Coomassie gel for expression of proteins. **C.** tabulates the LCMS data for expression of mutants.

was expressed in *E.coli* by co-transforming with CBX1-TAG plasmid and pEVOLpAzF plasmid (tRNA^{Tyr}_{CUA}/*M.j.*pAzF-RS) and over-expressed in the presence of pAzF. The proteins were purified through affinity chromatography and size-exclusion chromatography. The expression of protein was confirmed by running them on SDS-PAGE gel figure 51b. and incorporation of UAA was verified by LCMS figure 51c.

4.2.2. Binding Affinity of Mutants

To identify the mutants that bind to H3K9me3 with an affinity similar to wild-type CBX1, FP based binding assay was conducted on all mutants. Again, G9a methyltransferase assay was conducted to generate trimethyl peptide (TAMRA-GGARTKQTARK(me3)STGGKAPR-CONH2). Mutants binding with K_d close to the wild-type CBX1 (1.7 μ M) would also interact with the native binding partners in cellular environment with the same affinity as the wild-type CBX1, thereby reducing the false positive hits in proteomics data. As expected W47pAzF did not bind strongly to H3K9me3-peptide since any mutation to W47 compromises the structure of the binding pocket. Also, residues outside of the binding pocket show reduced binding probably because these residues point toward the peptide and caused steric clash due to mutation to a bulky aromatic residue. Y26pAzF, F50pAzF and D54pAzF mutants recapitulated the binding affinity of wild type CBX1 suggesting that incorporating aromatic pAzF group does not affect the structural integrity of the aromatic cage.

CBX1- Protein	Κ _d (μM)	
WT	1.7	
Y26pAzF	7.7 ± 2.3	
V28pAzF	>300	
W47pAzF	>260	
F50pAzF	11.5 ± 2.7	
D54pAzF	11.5 ± 2.1	
T56pAzF	>300	
E58pAzF	>300	
L63pAzF	>150	
C65pAzF	19.2 ± 1.3	
Y26/F50pAzF	48.9 ± 8.7	
Y26/D54pAzF	50.8 ± 5.7	
D54/T56pAzF	42.5 ± 9.0	
Table 7.	Fluorescence	

Polarization Data. Table shows the binding affinity of mutants towards TAMRA-H3K9me3 peptide. Error bars generated from duplicates of data.

To validate the binding data, we conducted ITC experiments with trimethylated peptide (H2N-ARTKQTARK(Me3)STGGKA-CO2H). The difference in buffer for ITC assay compared to FP assay accounts for the difference in K_d values. The K_d of wild-type CBX1 with H3K9me3 as measured by ITC was 5 μ M. The K_d values for the mutants in the binding pocket measured via ITC assay are shown in table 8, validate the binding data generated for the mutants. The K_d

measured for D54pAzF by FP assay is 11.5 μ M and that measured by ITC is 10.9 μ M, exemplify that the binding data generated for the mutants by two different techniques is reliable and valid. Based on the K_d values we identified Y26pAzF, F50pAzF and D54pAzF as the mutants that bound strongly to H3K9me3 peptide without perturbing interactions in the binding pocket of chromodomain.

CBX1- Protein	K _d (μM)	N	ΔG (cal/mol)	ΔH (cal/mol)	ΔS (cal/mol/deg)
WT	5.3 ± 0.4	0.985 ±0.01	-7200	-7990 ± 71.92	-2.65
Y26pAzF	11.3 ± 1.1	0.905 ± 0.01	-6777	-7653 ± 99	-3.04
W47pAzF	160 ± 9	0.830 ± 0.02	-5221	-6532 ± 340	-4.55
F50pAzF	33 ± 5	0.952 ± 0.03	-6204	-8730 ± 383	-8.77
D54pAzF	10.9 ± 1.3	0.806 ± 0.02	-6843	-9058 ± 191	-7.69
T56pAzF	290 ± 80	0.243 ± 0.174	-5200	-15770 ± 13340	-36.7

Table 8. ITC Data. Table shows the binding constant for each CBX1-pAzF mutant with H3K9me3 peptide along with the thermodynamic parameters. N here represents the number of binding sites.

4.2.3. Photo-crosslinking Efficiency of Mutants with Peptide

Next, we wanted to assess the crosslinking ability of these mutants to its substrates. For strong crosslinking, the photo-crosslinking group (azido group) should orient towards the substrate in its most populated conformation. To identify the mutants that crosslink to H3K9me3 we employed UV crosslinking of proteins with TAMRA-H3K9me3 (generated by G9a methyltransferase assay) followed by in gel fluorescence detection. The bands of the protein crosslinked to TAMRA-peptide would have a higher molecular weight and run slowly on SDS-PAGE gel. As the molecular weight

of CBX1 is 9.3 kDa and the peptide was 2 kDa the difference in the mass of peptide bound to protein and unbound was not reliable to conclude crosslinking efficiency from these experiments



Figure 52. Crosslinking with Peptide followed by In-gel Fluorescence. Gels scanned at 570 nm for TAMRA-H3K9me3 detection after crosslinking with the peptide. The site mutated to pAzF is depicted at the top followed by +/- for with or without illumination of UV light. The bottommost band is degraded TAMRA from the peptide, the band above it present in all the lanes is TAMRA-peptide and the band present in lanes illuminated with UV light is the protein and peptide crosslinked band.

figure 52. The T56pAzF mutant consistently showed a strong crosslinked band in this experiment.

We generated five double mutants based on the binding data and the preliminary crosslinking data to improve the binding affinity and the crosslinking. These mutants are Y26pAzF/F50pAzF, Y26pAzF/D54pAzF, F50pAzF/D54pAzF, F50pAzF/T56pAzF and D54pAzF/T56pAzF. Upon purification all mutants were verified via LCMS for the incorporation of two pAzF amino acids figure 51c.

As the expression of double TAG mutants was less compared to single mutants, we carried out FP assay on three of the mutants to assess the binding affinity/ structural integrity of these mutants towards H3K9me3. The binding affinity for Y26pAzF/D54pAzF and Y26pAzF/F50pAzF has decreased suggesting an increase in the steric clash for H3K9me3 in the binding pocket. Unexpectedly, the binding affinity of D54pAzF/T56pAzF increased from T56pAzF single mutant, this could possibly be attributed to increase in the pi-charge in the binding pocket. The double mutants did not seem to bind the native substrate of CBX1-chromodomain better than the single mutants.

4.2.4. Crosslinking Efficiency with Histone Extracts

As the molecular weight of full-length H3 is higher than the peptide we expected that crosslinking to full-length H3K9me3 would give us more reliable data (with an appreciable difference between crosslinked and non-crosslinked bands) in a biologically relevant environment. The histones were extracted from HEK293T cells by acid precipitation. The H3 from all the histones was quantified by running histone extracts along with BSA standards on SDS-PAGE gel. H3 was then crosslinked to the bacterially expressed unnatural CBX1 proteins. The crosslinking samples were western blotted against H3 antibody and His-tag antibody.

Interestingly, our negative control, wild-type CBX1 formed dimer and trimer. Both the antibodies confirmed a crosslinked band for all the single and double mutants, figure 53, except D54pAzF. H3 antibody does show a crosslinked band for D54pAzF but the His-tag antibody does not recognize the uncrosslinked protein either. Although it binds strongly to H3K9me3, our concern 123

with the structural integrity of D54pAzF mutant led to the exclusion of this mutant in our future studies. Also, the double mutants have low expression and do not crosslink as strongly as the Y26pAzF single mutant, so we excluded these from the future studies too. Based on our binding data as well as crosslinking data it seemed Y26pAzF would strongly compete against wild-type CBX1 to bind and crosslink to its native binding partners. Next, we pursued to express this mutant in HEK293T cells and crosslink it to its native binding partners in cells too.



Figure 53. Crosslinking with Histone Extracts. The figure shows crosslinking of all the pAzF mutants with histone extracts followed by western blot with H3 or His-tag antibody. The sites mutated to pAzF are depicted on the top followed by illumination with or without the UV light.

4.2.5. In Cell Crosslinking of CBX1 Mutant

To express the protein and crosslink them in cells, we first inserted a strep-tag at the c-terminal of the construct. This is important to pull-down the protein and its crosslinked binding partners from nuclear extracts. We then generated the Y26TAG mutant and together with plasmid for tRNA and tRNA-RS^{117,118,119}we were able to successfully express it in HEK293T cells, figure 54, as detected by HA-tag and CBX1 antibodies. The cells were grown on 10 cm petri-dish and exposed to the

UV light from the top, when expressing Y26pAzF. The nuclear extracts were first blotted to confirm the expression of protein. Interestingly, the samples exposed to UV light always showed bands of higher molecular weight. To enrich the crosslinked samples, we incubated them with strep-tactin beads, washed and eluted. The elutions were coomassie stained and western blotted against HA-tag and CBX1 antibodies. The elutions from UV exposed and unexposed samples on coomassie staining did not show enrichment of bands. Based on the sensitivity of antibodies the samples exposed to UV did show enrichment of crosslinked bands in the +UV elutions. In order to quantify the enrichment and identify the CBX1 interactome we plan to send the elutions to the proteomics facility. Western blot on elutions with H3 antibody was performed to confirm the crosslinked band was histone 3. The western blot with H3 antibody was not conclusive as several bands close to H3 molecular weight. This may be attributed to presence of polynucleosomes in the

sample instead of mononucleosomes. To overcome this the samples should be sonicated for a longer duration of time or treated with MNase.



Figure 54. *In-Vivo* **Crosslinking**. On the left is the western blot with anti-HA for CBX1 Y26pAzF expression from nuclear extracts generated after crosslinking and before the pull-down. On the right are the western blot with anti-HA, anti-CBX1 and anti-H3 and the Coomassie stain after pull-down.

4.3. CONCLUSIONS AND FUTURE DIRECTIONS

We employed IBPP approach to chromodomain containing CBX1 proteins and successfully crosslinked it to its binding partners *in cellulo*. We first established the biochemical integrity of mutants expressing UAA followed by their crosslinking efficiency in biochemical assays. Finally, we expressed full length CBX1Y26pAzF protein in HEK293T cells and crosslinked them by exposing them to UV light. We anticipate getting the proteomics data for the pull-down samples

and validate the crosslinked proteins by biochemical assays. We envision extending the IBPP approach to other chromodomain containing proteins to shed light on their role in chromatin dependent processes.

4.4. MATERIALS AND METHODS

4.4.1. Protein Expression and Mutagenesis

All mutagenesis (insertions and site-directed mutagenesis) were done using QuickChange Lightening site-directed mutagenesis kit (Agilent Technologies). PCR protocol was 18 cycles of 20 seconds denaturing at 95°C, annealing of primers at 60°C for 10 seconds, and polymerization at 68°C for 5minutes. For insertion of residues the number of cycles was increased to 28. The PCR amplified gene was transformed in BL21 XL10 Gold cells and colonies are grown in presence of antibiotics (Kan-50 ug/uL). The plasmids were extracted using GeneJET. Plasmid Miniprep Kit (Thermo Scientific) and sequenced by Genewiz.

All TAG (amber suppressor) mutants were generated following the same protocol as described for chapter 3 and confirmed by sequencing. The mutated plasmids were co-transformed with evolved *Methanococcus jannaschii* p-Azido-L-phenylalanine RS (2copies + tRNA) expression vector (Addgene ID: 31186) in BL21 star (DE3) cells (Invitrogen). 15uL of cells were incubated with 10 ng of both the plasmids on ice for 30 minutes followed by heat-shock at 42°C for 25-30 seconds and immediately cooling to 4°C for 2 minutes. The cells were first grown in SOC media for an hour at 37°C and then on Kan 50 μ g/mL and Chl 35 μ g/mL LB agar plates overnight at 37°C.
Single colony from the plate was grown overnight in 10 mL LB with 50 µg/mL Kan and 35µg/mL Chl at 37°C. This culture was centrifuged at 1000 g for 10 minutes, 8 mL of supernatant LB was removed. The pelleted cells were re-suspended in 2 mL LB to inoculate 1L of GMML medium containing 50 µg/mL Kan and 35 µg/mL Chl. The cultures shake at 37°C for 8-9 hours till they reach an OD₆₀₀ of 0.7-0.8. The cells were introduced to p- AzF compound to a final concentration of 1mM and cooled by shaking for 30 minutes at 17°C. After the cultures cooled to 17°C they were induced with 0.05% arabinose and after half an hour of addition the cells are induced with 0.25 mM IPTG. The cells shake at 17°C for 20 h at 225 rpm. The cultures were centrifuged at 4000 rpm for 20 minutes. The pellet was re-suspended in 15 mL lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β- mercaptoethanol, 10% glycerol, 25 mM imidazole, Lysozyme, DNase, and Roche (protease inhibitor cocktail), followed by sonication (Qsonica-Q700) at 60 Amp for 2 minutes with pulse on/off for 10 seconds. The lysed cells were centrifuged at 13000 rpm for 50 minutes the supernatant is loaded on equilibrated Ni-NTA agarose resin (Thermo) beads. The supernatant shakes at 4°C with beads for 45 minutes. After 45 minutes of incubation with beads the supernatant is let to flow-through the column followed by washing with 20 mL of wash buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol and 25 mM imidazole). The protein was eluted with 5 mL of elution buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol and 400 mM imidazole) in 5X 1mL fractions. The fractions were then injected in AKTA pure FPLC system (GE healthcare), Superdex 75 column. The FPLC buffer contained 10% glycerol, 50 mM Tris-HCl pH 8.0 and 150 mM NaCl. The FPLC fractions were concentrated using Amicon Ultra-3K centrifugal filter device (Merck Millipore Ltd.). The concentration of protein was measured using Bradford assay kit (BioRad Laboratories) with BSA as standard. The protein was stored in -80°C as aliquots.

4.4.2. FP Assay

Fluorescence polarization assay was employed to screen the binding affinity between the CBX1 mutants expressing unnatural amino acid and H3K9me3 peptide generated by G9a methyltransferase assay. The assay was performed in a 384 well small volume black/clear microtiter plates (Falcon) with 200 nM TAMRA labelled peptides and varying concentrations of protein (0.3-750 μ M) in 10 mM Tris pH 8.0, 150 mM NaCl, 0.05% Tween 20 and 0.5 mM of Tris (2-carboxyethyl) phosphine. The plate was centrifuged at 3000 rpm for 3 mins and the polarization was read for each well on TECAN M 1000 plate reader with an excitation at 530 nm and emission at 570 nm. For determination of dissociation constants (Kd), the background corrected fluorescence polarization was plotted against the concentrations in μ M. The data was fitted to single site binding equation Y= B_{max}*X/(Kd+X), where Y is the specific binding, B_{max} is the maximal binding and X is the concentration of ligand, using SigmaPlot software. The mP values for various concentrations were then divided by the highest mP value to get fraction bound as plotted in the graph

4.4.3. ITC Assay

The ITC measurements were carried out at 25°C using ITC₂₀₀ (Microcal, Malvern). Both protein and peptide were dissolved in 50 mM Tris-HCl pH 8.0, 200 mM NaCl and 10% glycerol. The sample cell contained the protein and the syringe had a 10-fold higher peptide concentration, individual concentrations for each experiment are given in table #. Each experiment comprised of 0.4 μ L of first injection followed by 19 injections of 2 μ L at a spacing of 120 s with a stirring speed of 750 rpm. The data points were fitted to a single binding site model using the Microcal ITC₂₀₀ using Origin Software.

4.4.4. Photo-crosslinking Experiment with Peptide and In-gel Fluorescence

1uM TAMRA labeled H3K9me3 peptide was incubated with 25 uM CBX1-WT or mutants in binding buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Tween 20, and 0.5 mM TCEP). They were irradiated with UV (365 nm) for 30 mins at 4°C using Transilluminator 2040 EV (Stratagene). Subsequently samples were heated at 95oC for ten minutes with the loading dye and ran on a 15% SDS-PAGE gel. The samples were visualized for TAMRA peptide using the ChemiDoc MP Imaging System (excitation filter 605/50).

4.4.5. Histone Extraction (HE)

Histone extraction was carried out following protocol the protocol by D. Shechter et.al¹²⁰. The HEK293T cells are grown in 10% FBS and DMEM media to a 100% confluency and collected by trypsinization. The cells are washed twice with PBS buffer and incubated in hypotonic solution (10 mM Tris–Cl pH 8.0, 1 mM KCl, 1.5 mM MgCl₂, 1 mM DTT and 1X protease inhibitor) for 30 mins at 4°C. The intact nuclei were pelleted by centrifuging at 4°C, 10,000 g for 10 mins. The supernatant was discarded, and the nuclei was resuspended in 400 μ L of 0.4 N H₂SO₄, overnight. The samples were the centrifuged at 16000 g, 4°C for 10 mins to remove nuclear debris. The supernatant was transferred to a fresh 1.5 mL centrifuge tube, 132 μ L of TCA was added dropwise to precipitate the histones and incubated on ice for 30 mins. The histones were pelleted by centrifuging at 16,000 g, 4°C for 4 mins. The supernatant was removed and the pelleted histones were washed thrice with ice-cold acetone, centrifuging at 16,000g, 4°C for 5 mins after every wash. Finally, the histone pellet was air dried for 20 mins and re-suspended in dI H₂O. 1uL of Histone extracts was loaded on a 15% gel with the BSA standards to quantify the amount of H3 for the assay.

4.4.6. HE Crosslinking

10 μ g histone H3 from histone extracts was incubated with 45 μ M of L92AzF in the binding buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Tween 20, and 0.5 mM TCEP). The samples were subjected to UV irradiation at 365 nm for 30 min at 4°C. Negative controls were not subjected to UV irradiation. The samples were loaded on 15% gel followed by western blot protocol.

4.4.7. In-Vivo Protein Expression

Human embryonic kidney (HEK) 293T cells were grown in Dulbecco modified Eagle medium (DMEM) (Gibco) supplemented with 10% fetal calf serum in a humidified atmosphere containing 5% CO2. At 60-70% confluency, cells were transfected with CBX1-Y26TAG plasmid, TPS-136 tRNA synthetase plasmid and TPS-192 suppressor tRNA plasmid in a 1:0.1:1 ratio¹²¹ with Lipofectamine 2000. 6hrs post transfection the media of the cells was changed to 20% FBS in DMEM and the cells were treated with 1 mM pAzF. After 24 hours the media was changed back to 10% FBS with 0.5 mM pAzF for another 24 hours.

4.4.8. In-Vivo Crosslinking

For *in-vivo* crosslinking the media of the cells was changed to cold PBS (3 mLs for 10-cm pertridish) and the cells were exposed to UV (using Spectroline Maxima ML-3500S) at 365nm for 30 mins at 4°C held at a distance of 3 inches from the petridish¹²¹. After 30 minutes the cells were collected by trypsinization followed by nuclear extraction.

4.4.9. Nuclear Extracts

The nuclear extracts were generated by following the protocol by Nielsen *et. al*⁹⁸. Briefly cellular pellet washed twice with PBS was resuspended in nuclear isolation buffer (15 mM Tris-HCl pH 7.5, 60 mM KCl, 15 mM NaCl, 5 mM MgCl₂, 1mM CaCl₂, 1mM DTT, 2 mM sodium vanadate, 250 mM sucrose, 1X protease inhibitor cocktail and 0.4% NP-40). The cells were incubated with NIB on ice for 5 mins followed by centrifugation at 2000g for 5 mins at 4°C. The pelleted nuclei were resuspended in nuclear extraction buffer (25 mM Tris pH 8.0, 250 mM NaCl, 1mM EDTA, 10% glycerol, 0.2% NP-40 and protease inhibitor cocktail). The nuclear extracts were then sonicated at 100Amps for 10 mins with 1min pulse on and 20 secs of pulse off duration. The extracts were then centrifuged at 2000g for 5 mins at 4°C followed by quantification using Bradford assay.

4.4.10. Streptactin Pull Down

5 mg of nuclear extract from crosslinked and non-crosslinked samples were incubated with streptactin beads (pre-washed with 50 mM Tris pH 8.0) for one-hour at room temperature. The supernatant was pulled off and the beads were washed thrice with 50 mM Tris pH 8.0, 200 mM NaCl, 0.1% Tween and 1x protease inhibitor cocktail. The protein was eluted with 40 μ L of 0.8% SDS twice. The elutions were heated to 95°C with 1X loading dye and ran on 8% gel for WB.

4.4.11. WB

Equal volumes of the pulled-down, photo-crosslinked samples were separated on SDS-PAGE and transferred onto a 0.45 µm supported PVDF membrane (Bio-Rad Laboratories) at a constant voltage of 80 for 1 h at 4°C. Membranes were blocked with 5% non-fat milk in TBST buffer (50 mM Tris HCl pH 7.4, 150 mM NaCl, 0.01% Tween-20) for 1 hr at room temperature with gentle shaking. Immunoblotting was performed with 1:500 diluted HA-tag or 1:1000 diluted for CBX1 antibody or 1: 2000 for H3 antibody for 12 hrs at 4°C. The antibody solution were removed and membranes were washed three times with TBST buffer. The blots were then incubated with HRP-conjugated secondary antibody Goat anti-Mouse IgG (cat#15014, Active motif) or with anti-Rabbit IgG () in TBST for 1 h at room temperature. After similar washing, protein bands were visualized by chemiluminescence using VISIGLO HRP Chemiluminescent substrates A and B (cat# N252-120ML and N253-120ML, aMReSCO) following manufacturer's protocol.

4.5. FIGURES AND TABLES

Mutant	Sequence
Y26TAG_F	GGA GGA GGA GGA ATA GGT GGT GGA AAA AGT TCT CG
Y26TAG_R	CGA GAA CTT TTT CCA CCA CCT ATT CCT CCT CCT C
V28TAG_F	GAG GAG GAA TAT GTG TAG GAA AAA GTT CTC GAC CGT
V28TAG_R	ACG GTC GAG AAC TTT TTC CTA CAC ATA TTC CTC CTC
W47TAG_F	GGA GTA CCT CCT AAA GTA GAA GGG ATT CTC AGA TGA GG
W47TAG_R	CCT CAT CTG AGA ATC CCT TCT ACT TTA GGA GGT ACT CC
F50TAG_F	CTA AAG TGG AAG GGA TAG TCA GAT GAG GAC AAC
F50TAG_R	GTT GTC CTC ATC TGA CTA TCC CTT CCA CTT TAG
D54TAG_F	GGA TTC TCA GAT GAG TAG AAC ACA TGG GAG CCA
D54TAG_R	TGG CTC CCA TGT GTT CTA CTC ATC TGA GAA TCC
T56TAG_F	CTC AGA TGA GGA CAA CTA GTG GGA GCC AGA AGA
T56TAG_R	TCT TCT GGC TCC CAC TAG TTG TCC TCA TCT GAG
E58TAG_F	GAG GAC AAC ACA TGG TAG CCA GAA GAG AAC CTG
E58TAG_R	CAG GTT CTC TTC TGG CTA CCA TGT GTT GTC CTC
N62TAG_F	TGG GAG CCA GAA GAG TAG CTG GAT TGC CCC GAC
N62TAG_R	GTC GGG GCA ATC CAG CTA CTC TTC TGG CTC CCA
L63TAG_F	CCA GAA GAG AAC TAG GAT TGC CCC GAC
L63TAG_R	GTC GGG GCA ATC CTA GTT CTC TTC TGG
C65TAG_F	GAG AAC CTG GAT TAG CCC GAC CTC ATT GCT
C65TAG_R	AGC AAT GAG GTC GGG CTA ATC CAG GTT CTC
D54TAG_F50TAG(t)_Fwd	GAT AGT CAG ATG AGG ACA ACA CAT GGG AGC
D54TAG_F50TAG(t)_Rev	GCT CCC ATG TGT TGT CCT CAT CTG ACT ATC
D54TAGF50TAGt_F2	GAT AGT CAG ATG AGT AGA ACA CAT GGG AGC
D54TAGF50TAGt_R2	GCT CCC ATG TGT TCT ACT CAT CTG ACT ATC
D54TAGT56TAG F	GGA TTC TCA GAT GAG TAG AAC TAG TGG GAG CCA
D54TAGT56TAG R	TGG CTC CCA CTA GTT CTA CTC ATC TGA GAA TCC
mCBX1_Y26TAG_F	5'- GAA GAG GAG GAA GAG GAA TAG GTG GTG GAA AAA GTT CTC GAC CG -3'
mCBX1 Y26TAG R	5'- CGG TCG AGA ACT TTT TCC ACC ACC TAT TCC TCT TCC TCC
mCBX1 F50TAG F	5'- CCT CCT AAA GTG GAA GGG ATA GTC AGA TGA GGA CAA CAC ATG GG -3'
mCBX1_F50TAG_R	5'- CCC ATG TGT TGT CCT CAT CTG ACT ATC CCT TCC ACT TTA GGA GG -3'
mCBX1 T56TAG F	5'- GGG ATT CTC AGA TGA GGA CAA CTA GTG GGA GCC AGA AGA G -3
mCBX1 T56TAG R	5'- CTC TTC TGG CTC CCA CTA GTT GTC CTC ATC TGA GAA TCC C -3'
mCBX1_D54TAG_F	GGG ATT CTC AGA TGA GTA GAA CAC ATG GGA GCC AGA AGA G
mCBX1_D54TAG_R	CTC TTC TGG CTC CCA TGT GTT CTA CTC ATC TGA GAA TCC C
mCBX1-F50TAG 2 F	CCT CCT AAA GTG GAA GGG ATA GTC AGA TGA GGA CAA C
mCBX1-F50TAG 2 R	GTT GTC CTC ATC TGA CTA TCC CTT CCA CTT TAG GAG G
mCBX1 D54T56TAG F	GGG ATT CTC AGA TGA GTAGAA CTA GTG GGA GCC AGA AGA G
mCBX1_D54T56TAG_R	CTC TTC TGG CTC CCA CTA GTT CTA CTC ATC TGA GAA TCC C

Table 9. List of Primers for Generating TAG Mutants on CBX1. All the primers with

mCBX1 were generated for mammalian construct and the rest are for bacterial plasmids.



Figure 55. Dissociation Constants. Of CBX1-pAzF mutants, towards the TAMRA-H3K9me3 peptide, determined by fluorescence polarization values. Error bars represent standard deviation from two independent measurements. For weakly binding pairs we did not observe saturation in mP, the K_d values were determined by Bmax.



Figure 56. Dissociation Constants. Of wild type CBX1pAzF mutants, towards the TAMRA-H3K9me3 peptides, determined by fluorescence polarization values. Error bars represent standard deviation from two independent measurements. For weakly binding pairs we did not observe saturation in mP, the K_d values were determined by Bmax.







5. BROADER PERSPECTIVE

Epigenetics spans all the mechanisms that are concerned with heritable, reversible processes that change the gene expression without modulating the underlying genetic code¹²². These mechanisms are important for embryonic development, changes and propagation of cell type specific gene expression patterns providing cellular stability¹²³. The fundamental mechanisms by which these changes are brought about are DNA methylation, and several chemical modifications on RNA and histone proteins by epigenetic proteins (writer, reader and eraser proteins). Alterations in epigenetic state (aberrant expression of epigenetic protein or disruption in modifications on epigenetic landscape) are associated with several diseases. Understanding the role and function of each epigenetic modification may unravel new insights for diagnosis, treatment, and even prevention of diseases. Several inhibitors of epigenetic enzymes have been approved for use in the clinic to treat cancer patients with hematological malignancies but they lack locus-selective specificity, that may result in the expression or repression of undesirable parts of the genome¹²⁴. The plasticity (reversibility) of these modifications has laid the basis for the field of epigenetic editing that would improve epigenetic therapy by reducing off-target effects. Alterations in epigenetic landscape that are gene-specifically manipulated using epigenetic regulatory proteins is termed as epigenetic editing¹²⁵. Present approaches to rewrite a gene's epigenetic signature fuse the catalytic domain of writer or an eraser protein to a programmable gene-specific DNA-binding domain (DBD) that targets it to a specifc-locus ^{122, 124}. We proposed employing allele-specific chemical genetics on the writer and reader proteins to manipulate the epigenetic landscape, figure 60. ASCG or bump-hole approach had previously been successfully employed to several proteinsmall molecule interactions. Here, an extended pocket (hole) is generated in the binding/ catalytic domain of the protein that is active in the presence of the bumped analogue and not the native cofactor.



to control the formation of heterochromatin.

Here we applied this approach to histone methyltransferase SUV39h2 that trimethylates H3K9 and recruits CBX1 to form the heterochromatin. We engineered SUV39h2 at position Y372 to generate an extended pocket by mutating it to alanine or glycine residue. Based on MALDI-methyltransferase assay we observed that the glycine mutant was more orthogonal than the alanine mutant towards the native SAM at trimethylating H3K9. Lower binding affinity to native substrate on extending the binding pocket maybe attributed to poor accommodation of the substrate. Both the mutants were able to transfer the aromatic (benzyl group) from B-SAM to H3K9. We measured the catalytic efficiency of the transfer for Y372G mutant for benzyl SAM. The K_m and k_{cat} values for the engineered system were 32.2 μ M and 1.3 min⁻¹ respectively compared to 11.7 μ M and 12.4 142

min⁻¹ for the wild-type system respectively. The data suggested that the engineered system is less efficient at incorporating the benzyl on H3K9 compared to trimethylation on H3K9 by WT SUV39h2. As benzylation is irreversible with respect to methylation, we expected our engineered system to confidently identify its target genes *in-cellulo*. The study also involved engineering MAT (methionine adenosyl transferase) because B-SAM is not cell-permeable, but benzyl-methionine is cell permeable. MAT2A-I117G employs ATP and benzyl-methionine to generate benzyl-SAM, that acts as a cofactor for SUV39h2 Y372G to benzylate H3K9. Most importantly, for epigenetic editing we generated the benzyl-H3K9 antibody that confirmed the incorporation of benzyl on H3K9 in mammalian cells (HEK293T) by western blot and fixed cell imaging experiments. To identify the targets of SUV39h2 with our ASCG system, we first verified the downregulation of reported genes regulated by SUV39h2 (by qRT PCR), followed by sending the nucleosomal samples for ATAC-seq and RNA samples for RNA-seq.

Establishing the process of writing/ incorporating an unnatural modification on H3K9 in cells with an engineered methyl-lysine writer protein doesn't necessarily imply heterochromatin formation. The target genes not being regulated in cells may be attributed to orthogonality of the modification towards transcription factors and other regulatory factors. To gain control over heterochromatin formation of genes by introduction of benzyl group at H3K9 we had to engineer the reader protein CBX1 to identify and bind to benzylated H3K9. The engineered CBX1* that would bind to H3K9benzyl would ensure heterochromatin formation at the H3K9-benzylated sites. We mutated the residues in the binding pocket of CBX1 to smaller residue (alanine) to generate the hole and screened the library of H3K9-analogues through fluorescence polarization assay. The Y26F/F50G mutant of CBX1 bound most strongly to benzylated H3K9 (5.5 μ M) and with least affinity to trimethylated H3K9 (74 µM). These binding affinities were confirmed by isothermal caloriemetric titration experiments. As there are very few reports on application of bump-hole approach to protein-protein interface the next investigation was the structural integrity of the engineered system. For the first time we solved the crystal structure of CBX1 with H3K9me3 (PDB 6D07) as well as of CBX1 Y26F/F50G with benzyl H3K9 (PDB 6D08). The rmsd of the backbone atoms for the two structures was 0.6 A°, showing the binding of benzylated H3K9 to engineered CBX1 did not perturb the overall conformation of the chromodomain. We addressed the member-specific role of CBXs that bind to H3K9 by mutating similar gatekeeper residues of CBX3 and CBX5 and measuring their binding affinity to H3K9-benzyl. Identification of member-specific role for PPI based interactions by application of ASCG approach is important as the development of inhibitors for such interactions is challenging, since it involves mimicking several surface interactions to compete the binding interaction of PPI. Our data suggests that engineering similar sites on other members of the reader protein leads to similar accommodation (binding) of these members to benzylated H3K9, displaying the generality of our approach. To confirm the applicability of our approach in cellular context we expressed the engineered CBX1 in mammalian cells (HEK293T) generated the nuclear extracts from them and incubated them with beads bound to biotinylated peptide carrying methyl or benzyl modification on them. After vigorous washing the protein was eluted from the peptide and quantified through western blot. Our data confirmed binding of CBX1, CBX3 and CBX5 to H3K9me3 only and not to H3K9-benzyl whereas CBX1-Y26F/F50G eluted from the benzylated H3K9 beads only and not from H3K9me3 or other trimethylated lysines on histone 3 (H3K27me3 or H3K4me3 or H3K36me3), under the experimental conditions. This confirmed that the engineered system remains orthogonal as well as specific under cellular conditions. Following these experiments, we also confirmed that the binding of chromoshadow

domain with downstream interacting partners remains unperturbed by our engineering. Similar pull-down with H3K9, H3K9me3 and H3K9benzyl beads was carried out under less vigorous washing conditions. Our results confirmed that the binding of regulatory proteins to chromoshadow domain of CBX1* is intact as the wild-type system. This result was important to confirm the heterochromatin in cells with our engineered system.

Finally, we expressed the MAT2A*, SUV39h2* and CBX1* in HEK23T cells and conducted fixed cell imaging experiments to overlay benzylation with CBX1*. CBX1* did not merge on top of benzylated nuclear regions. The chromodomain of SUV39h2 may be responsible for this result as it binds and recruits SUV39h2 to locations that are trimethylated at H3K9, shown in figure 61. The chromodomain of SUV39h2 in our system benzylates the H3K9 that is neighbouring to the H3K9me3 nucleosomes which would bring together CBX1* and CBX1 instead of CBX1*-CBX1*. Engineering the chromodomain of SUV39h2 may overcome this challenge and lead to heterochromatin formation in cells.



So far, we have edited the epigenetic landscape by benzylating the H3K9 in mammalian cells and we have also engineered the reader protein to identify and bind the unnatural modification without perturbing its structure and function in a cellular environment. We still need to demonstrate that our system localizes to benzylated histones and forms the heterochromatin in cells. The complete information of the targets of CBX1 is required to validate our system. A gap in knowledge lies here as the approaches taken so far to identify the genes regulated by CBX1 do not discriminate between the binding partners of chromodomain and chromoshadow domain. If they do discriminate between the two the experiments are not conducted in cells. We applied the IBPP approach to CBX1 to identify its binding partners in live mammalian cells. IBPP approach includes incorporating an unnatural amino acid in the binding pocket (chromodomain) of the protein that would photo-crosslink upon illumination of UV light. This would ensure that the proteins pulleddown with CBX1* are the targets of chromodomain only. To apply this approach, we first identified the site that would bind to H3K9me3 as the wild-type protein on mutation to photocrosslinkable unnatural amino acid (pAzF). We also verified that the mutant would crosslink invivo to the H3K9me3 peptide followed by histone extracts before taking the system to cells. Upon identification of the site that best represents the wild-type system in terms of binding and crosslinking to its substrate we expressed the mutant in mammalian cells and crosslinked them by illumination with UV light before collecting their nuclear extracts. Western blot of the nuclear extract confirmed crosslinking of CBX1* to its targets only upon illumination of UV light as the higher molecular weight bands did not appear in negative control. The proteomics/ genomics analysis of the samples would provide a more comprehensive picture of the roles/ binding partners/ targets of CBX1.

To conclude, to accomplish epigenetic editing through ASCG approach we engineered MAT2A, SUV39h2, and CBX1. We have established benzylation in mammalian cells with our engineered MAT2A and SUV39h2 system. We have also engineered protein-protein interaction of an epigenetic reader protein to bind to an unnatural protein interface orthogonally and specifically. We hope that our effort would lead to the formation of heterochromatin in cells and finally when fused to dCas9 it would silence genes locus-specifically. Also, to identify the binding partners/ targets of reader proteins we applied an unbiased approach, IBPP approach to CBX1. The proteomics/ genomics data would probably shed more light on the binding partners of CBX1.

6. APPENDIX A1

To establish the reliability of our IBPP tool, we applied it on a system whose non-histone substrate were already reported. Previously Islam *et. al*, identified non-histone substrates of G9a by applying their bump-hole pair of G9a Y1154A and hex-2-ene-5-yne SAM in cellular extracts followed by click chemistry with a biotin probe to pull-down its substrates. We applied the IBPP to G9a to authenticate the tool as well as to identify its transient interacting partners. First the sites are identified for incorporation of UAA (pAzF) followed by identification of mutant that would have the activity same as that of wild-type protein and strong substrate crosslinking efficiency. The crosslinking efficiency would be validated in more biologically relevant context followed by crosslinking in cellular extracts followed LCMS/MS to identify the binding partners.

Previous lab members, Corrin Durham and Amy Ryan, identified the sites in the catalytic pocket of G9a for crosslinking followed by expression and purification of UAA proteins. To crosslink the substrates of G9a we expected the UAA containing mutant to have the same activity as wild-type G9a. The mutants with less or no activity would suggest a change in the catalytic pocket of G9a thereby introducing false positive data in our further studies. Methyltransferase assay was applied to test the activity of the mutants shown in figure 62 and figure 63. The peptide is dimethylated by WT G9a under the assay conditions. Mutants F1087pAzF, F1158pAzF, Y1165pAzF and Y1166pAzF closely resemble WT in their activity whereas Y1067pAzF, Y1085pAF, F1152pAF, Y1154pAzF and W1159pAzF are less active than the WT suggesting a structural change in the catalytic pocket. Interestingly, F1138pAzF is more active than the WT suggesting a structural or



electronic change in the binding of substrates or an overall conformational change in the enzyme

Figure 62. Methyltransferase Assay of Mutants. The unmodified peptide has a mass of 2339 Da, monomethylated peptide is 2353Da, demethylated peptide is 2367Da and trimethylated peptide is 2381Da. The name of the mutant is given on the top right corner of each spectra. Each assay was conducted on 20 μ M of peptide with 60 μ M SAM and 2 μ M enzyme in 50mM Tris pH8.0.

favoring the catalytic efficiency.



Figure 63. Methyltransferase Assay of Mutants. The unmodified peptide has a mass of 2339 Da, monomethylated peptide is 2353Da, demethylated peptide is 2367Da and trimethylated peptide is 2381Da. The name of the mutant is given on the top right corner of each spectra. Each assay was conducted on 20 μ M of peptide with 60 μ M SAM and 2 μ M enzyme in 50mM Tris pH8.0.

Next, we crosslinked the TAMRA-H3K9 with the mutants to assess their crosslinking efficiency, figure 64. For the WT protein we see a signal indicating the presence of TAMRA-peptide in the binding pocket. We also see higher molecular bands suggesting oligomerization of G9a protein.



Figure 64. In Gel Fluorescence Crosslinking Assay. The numbers represent the residue numbers that are mutated to pAzF. The +/- correspond to presence or absence of U.V. light respectively. Each assay was carried out in 50 mM Tris pH 8.0, with 20 μ M peptide, 60 μ M SAM and 2 μ M enzyme followed by crosslinking for 30 mins at 4°C. The signal for TAMRA was visualized by absorption at 530nm and emission at 570 nm.

Y1067pAzF, Y1085pAzF, F1087pAzF and F1152pAzF crosslinked most strongly amongst the other mutants. A lower activity than the WT protein for these mutants along with the crosslinked data implies longer residence time of substrates (higher binding affinity) in the catalytic pocket compared to other mutants.

To confirm our data, we crosslinked the mutants to biotin-H3K9 peptide along with SAM in the assay followed by western blot with anti-biotin antibody, figure 65. This confirms Y1067pAzF and F1152pAzF to crosslink strongly. Y1085pAzF forms more dimer on crosslinking and Y1165pAzF has also emerged as a strong crosslinking mutant.

Figure 65. Crosslinking Efficiency of Mutants Detected by Western Blot. The numbers represent the residue numbers that are mutated to pAzF. The +/- correspond to presence or absence of U.V. light respectively. Each assay was carried out in 50 mM Tris pH 8.0, with 20 μ M peptide, 60 μ M SAM and 2 μ M enzyme followed by crosslinking for 30 mins at 4°C. Samples were loaded on a 12% SDS-PAGE gel followed by transfer on PVDF membrane (0.45 μ m) and incubation with anti-biotin antibody.

To pull-down our mutant along with its binding partners we incorporated the strep-tag (WSHPQFEK) at the c-terminal of enzyme. The insertion of strep-tag at the c-terminal assures pull-down of full-length protein only. Y1067pAzF, F1152pAzF and Y1165pAzF were regenerated and expressed in *E.coli*. The proteins were purified by Ni-affinity and confirmed by LCMS.



Figure 66. Methyltransferase Activity of Mutants Generated with C-strep. The unmodified peptide has a mass of 2339 Da, monomethylated peptide is 2353Da, demethylated peptide is 2367Da and trimethylated peptide is 2381Da. The name of the mutant is given on the top right corner of each spectra. Each assay was conducted on 20 μ M of peptide with 60 μ M SAM and 2 μ M enzyme in 50mM Tris pH8.0.

To confirm that the insertion did not affect the structure or function of the protein, methyltransferase activity was conducted, figure 66. The data confirmed that the activity was unaffected by insertion of strep tag at the c-terminal of the mutants. We then wanted to confirm the crosslinking efficiency of the mutants. We carried out the in-gel fluorescence assay to assess the crosslinking efficiency of mutants. The ability of the mutants to crosslink was also unaffected, figure 67 a. We then asked whether the presence of SAM was affecting the crosslinking efficiency



Figure 67. Crosslinking Efficiency of Mutants with Strep-Tag Insertion. In figure **a.** The mutants are crosslinked with TAMRA-containing trimethylated peptide in the presence or absence of UV light followed by in-gel analysis. In figure **b.** the mutants are crosslinked with TAMRA- H3K9me3 under UV light with or without SAM- cofactor, followed by in-gel analysis. In figure **c.** Enzymes are crosslinked with unmethylated (U) or trimethylated (T) TAMRA-peptide followed by in-gel analysis and in **d.** the enzymes are crosslinked with biotin-H3K9 (U) / biotin-H3K9me3 (T) followed by western blot analysis with anti-biotin antibody.

of these mutants. To address this, we carried out the in-gel assay with and without SAM, figure 67

b. Surprisingly, it did not affect the crosslinking of the mutants to the peptide.

Several groups have confirmed that G9a has a higher affinity for H3K9-peptide than H3K9me3 peptide. Considering that H3K9 is the substrate and H3K9me3 is the product, this affirms the general rule that the affinity for substrate is usually higher than the products that are released from the pocket upon formation. We carried out the in-gel fluorescence crosslinking assay with H3K9 or H3K9me3, without SAM, figure 67 c. Our data confirmed the reports, by crosslinking more strongly to H3K9 than H3K9me3 peptide.

To validate our data, we also conducted the crosslinking with tri-methylated and unmethylated biotin labeled peptide, figure 67 d. We saw similar results to our in-gel fluorescence data.

We then tested the efficiency of our system in more biologically relevant environment of histone extracts. Histone extracts were generated from HEK293T cells and upon quantification they were crosslinked followed by western blot with H3 antibody, figure 68. Surprisingly, the crosslinking was not as strong as expected. The data might be affected by the efficiency of the antibody to recognize crosslinked proteins.

To assess whether the antibody is not recognizing the crosslinked protein or the mutants do not crosslink efficiently to the full-length protein we crosslinked all the mutants without strep-tag with H3-full-length followed by western blot with the H3 antibody, figure 68. The H3 antibody worked to detect the crosslinked bands however the crosslinking efficiency of the mutants to the full-length



nm).

H3 protein was different than to the peptide as a substrate. 1085pAzF, 1154pAzF and 1158pAzF are crosslinking more efficiently than 1152pAzF or 1165pAzF.

157



This set of data suggests that some more mutants with strep-tag be generated before conducting the crosslinking and pull-down with cellular extracts for identification of the substrates of G9a by IBPP approach.

7. REFERENCES

- Islam, K. Allele-specific chemical genetics: concept, strategies, and applications. ACS Chem Biol 2015, 10 (2), 343.
- Wu, H.; Min, J.; Lunin, V. V.; Antoshenko, T.; Dombrovski, L.; Zeng, H.; Allali-Hassani,
 A.; Campagna-Slater, V.; Vedadi, M.; Arrowsmith, C. H.et al. Structural biology of human
 H3K9 methyltransferases. *PLoS One* 2010, *5* (1), e8570.
- (3) Olby, R.; Gautrey, P. Eleven references to Mendel before 1900. *Annals of Science* 1968, 24 (1), 7.
- (4) Watson, J. D.; Crick, F. H. The structure of DNA. *Cold Spring Harb Symp Quant Biol* 1953, 18, 123.
- (5) Khorana, H. G. Total synthesis of a gene. *Science* **1979**, *203* (4381), 614.
- (6) Bedinger, P.; Hochstrasser, M.; Jongeneel, C. V.; Alberts, B. M. Properties of the T4 bacteriophage DNA replication apparatus: the T4 dda DNA helicase is required to pass a bound RNA polymerase molecule. *Cell* **1983**, *34* (1), 115.

- (7) Skinner, M. K. Role of epigenetics in developmental biology and transgenerational inheritance. *Birth Defects Res C Embryo Today* **2011**, *93* (1), 51.
- (8) Corella, D.; Ordovas, J. M. Basic Concepts in Molecular Biology Related to Genetics and Epigenetics. *Rev Esp Cardiol (Engl Ed)* 2017, 70 (9), 744.
- Morange, M. The relations between genetics and epigenetics: a historical point of view.
 Ann N Y Acad Sci 2002, 981, 50.
- (10) Mohn, F.; Schubeler, D. Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends Genet* **2009**, *25* (3), 129.
- (11) Waddington, C. H. Canalization of development and genetic assimilation of acquired characters. *Nature* **1959**, *183* (4676), 1654.
- Wei, Y.; Zhang, H.; Wang, Q.; Zhang, C. Single Cell Genetics and Epigenetics in Early Embryo: From Oocyte to Blastocyst. *Adv Exp Med Biol* 2018, *1068*, 103.
- (13) Breiling, A.; Lyko, F. Epigenetic regulatory functions of DNA modifications: 5methylcytosine and beyond. *Epigenetics Chromatin* 2015, 8, 24.

- Wang, X.; Zhao, B. S.; Roundtree, I. A.; Lu, Z.; Han, D.; Ma, H.; Weng, X.; Chen, K.; Shi, H.; He, C. N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* 2015, *161* (6), 1388.
- (15) Dominissini, D.; Nachtergaele, S.; Moshitch-Moshkovitz, S.; Peer, E.; Kol, N.; Ben-Haim,
 M. S.; Dai, Q.; Di Segni, A.; Salmon-Divon, M.; Clark, W. C.et al. The dynamic N(1) methyladenosine methylome in eukaryotic messenger RNA. *Nature* 2016, *530* (7591), 441.
- Roundtree, I. A.; Evans, M. E.; Pan, T.; He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* 2017, *169* (7), 1187.
- Bannister, A. J.; Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* 2011, 21 (3), 381.
- (18) Gardner, K. E.; Allis, C. D.; Strahl, B. D. Operating on chromatin, a colorful language where context matters. *J Mol Biol* **2011**, *409* (1), 36.
- (19) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 2009, *324* (5929), 930.

- (20) Wu, H.; Zhang, Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **2014**, *156* (1-2), 45.
- Bhutani, N.; Brady, J. J.; Damian, M.; Sacco, A.; Corbel, S. Y.; Blau, H. M.
 Reprogramming towards pluripotency requires AID-dependent DNA demethylation.
 Nature 2010, 463 (7284), 1042.
- (22) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 2011, 333 (6047), 1300.
- (23) Ko, M.; An, J.; Bandukwala, H. S.; Chavez, L.; Aijo, T.; Pastor, W. A.; Segal, M. F.; Li, H.; Koh, K. P.; Lahdesmaki, H.et al. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* 2013, 497 (7447), 122.
- (24) Bakayev, V. V.; Bakayeva, T. G.; Varshavsky, A. J. Nucleosomes and subnucleosomes: heterogeneity and composition. *Cell* **1977**, *11* (3), 619.
- (25) Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **1997**, *389* (6648), 251.
- (26) Huisinga, K. L.; Brower-Toland, B.; Elgin, S. C. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma* 2006, *115* (2), 110.

- (27) Akinrimisi, E. O.; Bonner, J.; Tso, P. O. Binding of Basic Proteins to DNA. *J Mol Biol* 1965, *11*, 128.
- (28) Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* 2007, *14* (11), 1025.
- (29) Jenuwein, T.; Allis, C. D. Translating the histone code. *Science* **2001**, *293* (5532), 1074.
- (30) Arrowsmith, C. H.; Bountra, C.; Fish, P. V.; Lee, K.; Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat Rev Drug Discov* **2012**, *11* (5), 384.
- (31) Gillette, T. G.; Hill, J. A. Readers, writers, and erasers: chromatin as the whiteboard of heart disease. *Circ Res* **2015**, *116* (7), 1245.
- Rea, S.; Eisenhaber, F.; O'Carroll, D.; Strahl, B. D.; Sun, Z.-W.; Schmid, M.; Opravil, S.;
 Mechtler, K.; Ponting, C. P.; Allis, C. D.et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 2000, 406 (6796), 593.
- (33) Aagaard, L.; Laible, G.; Selenko, P.; Schmid, M.; Dorn, R.; Schotta, G.; Kuhfittig, S.;Wolf, A.; Lebersorger, A.; Singh, P. B.et al. Functional mammalian homologues of the
Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. *The EMBO journal* **1999**, *18*(7), 1923.

- (34) Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi,
 Y. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell*2004, *119* (7), 941.
- Becker, J. S.; Nicetto, D.; Zaret, K. S. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet* 2016, *32* (1), 29.
- (36) Saksouk, N.; Simboeck, E.; Déjardin, J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & Chromatin* **2015**, *8* (1), 3.
- (37) Trojer, P.; Reinberg, D. Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell* 2007, 28 (1), 1.
- (38) Nakayama, J.-i.; Rice, J. C.; Strahl, B. D.; Allis, C. D.; Grewal, S. I. S. Role of Histone H3
 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. *Science* 2001, 292 (5514), 110.
- Martens, J. H. A.; O'Sullivan, R. J.; Braunschweig, U.; Opravil, S.; Radolf, M.; Steinlein,
 P.; Jenuwein, T. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *The EMBO Journal* 2005, *24* (4), 800.

- (40) Margueron, R.; Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* 2011, 469, 343.
- Li, B.; Zheng, Y.; Yang, L. The Oncogenic Potential of SUV39H2: A Comprehensive and Perspective View. *J Cancer* 2019, *10* (3), 721.
- (42) Nielsen, P. R.; Nietlispach, D.; Mott, H. R.; Callaghan, J.; Bannister, A.; Kouzarides, T.;
 Murzin, A. G.; Murzina, N. V.; Laue, E. D. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* 2002, *416* (6876), 103.
- (43) Hou, H.; Yu, H. Structural insights into histone lysine demethylation. *Curr Opin Struct Biol* 2010, 20 (6), 739.
- (44) Labbe, R. M.; Holowatyj, A.; Yang, Z. Q. Histone lysine demethylase (KDM) subfamily
 4: structures, functions and therapeutic potential. *Am J Transl Res* 2013, 6 (1), 1.
- (45) Elgin, S. C. R.; Reuter, G. Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harb Perspect Biol* **2013**, *5* (8), a017780.
- (46) Simon, J. A.; Lange, C. A. Roles of the EZH2 histone methyltransferase in cancer epigenetics. *Mutat Res* **2008**, *647* (1-2), 21.

- (47) Zhao, Y.; Jensen, O. N. Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques. *Proteomics* 2009, 9 (20), 4632.
- (48) Hayashi-Takanaka, Y.; Yamagata, K.; Wakayama, T.; Stasevich, T. J.; Kainuma, T.; Tsurimoto, T.; Tachibana, M.; Shinkai, Y.; Kurumizaka, H.; Nozaki, N.et al. Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. *Nucleic Acids Res* 2011, 39 (15), 6475.
- (49) Phair, R. D.; Scaffidi, P.; Elbi, C.; Vecerova, J.; Dey, A.; Ozato, K.; Brown, D. T.; Hager, G.; Bustin, M.; Misteli, T. Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Mol Cell Biol* 2004, *24* (14), 6393.
- (50) Darzacq, X.; Singer, R. H.; Shav-Tal, Y. Dynamics of transcription and mRNA export. *Curr Opin Cell Biol* 2005, *17* (3), 332.
- (51) Metivier, R.; Penot, G.; Hubner, M. R.; Reid, G.; Brand, H.; Kos, M.; Gannon, F. Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **2003**, *115* (6), 751.
- (52) Surade, S.; Blundell, T. L. Structural biology and drug discovery of difficult targets: the limits of ligandability. *Chem Biol* **2012**, *19* (1), 42.

- (53) Wells, J. A.; McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007**, *450* (7172), 1001.
- (54) Hwang, Y. W.; Miller, D. L. A mutation that alters the nucleotide specificity of elongation factor Tu, a GTP regulatory protein. *J Biol Chem* 1987, 262 (27), 13081.
- (55) Bishop, A. C.; Buzko, O.; Shokat, K. M. Magic bullets for protein kinases. *Trends Cell Biol* 2001, *11* (4), 167.
- (56) Strader, C. D.; Gaffney, T.; Sugg, E. E.; Candelore, M. R.; Keys, R.; Patchett, A. A.; Dixon,
 R. A. Allele-specific activation of genetically engineered receptors. *J Biol Chem* 1991, 266
 (1), 5.
- (57) Holt, J. R.; Gillespie, S. K.; Provance, D. W.; Shah, K.; Shokat, K. M.; Corey, D. P.; Mercer, J. A.; Gillespie, P. G. A chemical-genetic strategy implicates myosin-1c in adaptation by hair cells. *Cell* **2002**, *108* (3), 371.
- (58) Kapoor, T. M.; Mitchison, T. J. Allele-specific activators and inhibitors for kinesin. *Proc Natl Acad Sci U S A* 1999, 96 (16), 9106.

- (59) Yang, C.; Mi, J.; Feng, Y.; Ngo, L.; Gao, T.; Yan, L.; Zheng, Y. G. Labeling lysine acetyltransferase substrates with engineered enzymes and functionalized cofactor surrogates. *J Am Chem Soc* **2013**, *135* (21), 7791.
- (60) Carter-O'Connell, I.; Jin, H.; Morgan, R. K.; David, L. L.; Cohen, M. S. Engineering the substrate specificity of ADP-ribosyltransferases for identifying direct protein targets. *J Am Chem Soc* 2014, *136* (14), 5201.
- (61) Breski, M.; Dey, D.; Obringer, S.; Sudhamalla, B.; Islam, K. Engineering Biological C-H Functionalization Leads to Allele-Specific Regulation of Histone Demethylases. J Am Chem Soc 2016, DOI:10.1021/jacs.6b08653 10.1021/jacs.6b08653.
- (62) Liu, C. C.; Schultz, P. G. Adding new chemistries to the genetic code. *Annu Rev Biochem* **2010**, 79, 413.
- (63) O'Carroll, D.; Scherthan, H.; Peters, A. H.; Opravil, S.; Haynes, A. R.; Laible, G.; Rea, S.;
 Schmid, M.; Lebersorger, A.; Jerratsch, M.et al. Isolation and characterization of Suv39h2,
 a second histone H3 methyltransferase gene that displays testis-specific expression. *Mol Cell Biol* 2000, *20* (24), 9423.
- (64) Ait-Si-Ali, S.; Guasconi, V.; Fritsch, L.; Yahi, H.; Sekhri, R.; Naguibneva, I.; Robin, P.;
 Cabon, F.; Polesskaya, A.; Harel-Bellan, A. A Suv39h-dependent mechanism for silencing
 S-phase genes in differentiating but not in cycling cells. *EMBO J* 2004, *23* (3), 605.

- (65) Sone, K.; Piao, L.; Nakakido, M.; Ueda, K.; Jenuwein, T.; Nakamura, Y.; Hamamoto, R. Critical role of lysine 134 methylation on histone H2AX for gamma-H2AX production and DNA repair. *Nat Commun* **2014**, *5*, 5691.
- (66) Piao, L.; Suzuki, T.; Dohmae, N.; Nakamura, Y.; Hamamoto, R. SUV39H2 methylates and stabilizes LSD1 by inhibiting polyubiquitination in human cancer cells. *Oncotarget* 2015, 6 (19), 16939.
- (67) Copeland, R. A. Protein methyltransferase inhibitors as precision cancer therapeutics: a decade of discovery. *Philos Trans R Soc Lond B Biol Sci* **2018**, *373* (1748).
- (68) Wang, R.; Luo, M. A journey toward Bioorthogonal Profiling of Protein Methylation inside living cells. *Curr Opin Chem Biol* 2013, *17* (5), 729.
- (69) Islam, K.; Chen, Y.; Wu, H.; Bothwell, I. R.; Blum, G. J.; Zeng, H.; Dong, A.; Zheng, W.;
 Min, J.; Deng, H.et al. Defining efficient enzyme-cofactor pairs for bioorthogonal profiling of protein methylation. *Proc Natl Acad Sci U S A* 2013, *110* (42), 16778.
- Wang, R.; Islam, K.; Liu, Y.; Zheng, W.; Tang, H.; Lailler, N.; Blum, G.; Deng, H.; Luo,
 M. Profiling genome-wide chromatin methylation with engineered posttranslation apparatus within living cells. *J Am Chem Soc* 2013, *135* (3), 1048.

- (71) Rea, S.; Eisenhaber, F.; O'Carroll, D.; Strahl, B. D.; Sun, Z. W.; Schmid, M.; Opravil, S.;
 Mechtler, K.; Ponting, C. P.; Allis, C. D.et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 2000, 406 (6796), 593.
- Islam, K.; Zheng, W.; Yu, H.; Deng, H.; Luo, M. Expanding cofactor repertoire of protein lysine methyltransferase for substrate labeling. *ACS Chem Biol* 2011, 6 (7), 679.
- (73) Fussner, E.; Djuric, U.; Strauss, M.; Hotta, A.; Perez-Iratxeta, C.; Lanner, F.; Dilworth, F.
 J.; Ellis, J.; Bazett-Jones, D. P. Constitutive heterochromatin reorganization during somatic cell reprogramming. 2011, *30* (9), 1778.
- (74) Chao, C.; You, J.; Li, H.; Xue, H.; Tan, X. Elevated SUV39H2 attributes to the progression of nasopharyngeal carcinoma via regulation of NRIP1. *Biochem Biophys Res Commun* 2019, *510* (2), 290.
- (75) Muttach, F.; Rentmeister, A. A Biocatalytic Cascade for Versatile One-Pot Modification of mRNA Starting from Methionine Analogues. *Angew Chem Int Ed Engl* 2016, 55 (5), 1917.
- James, T. C.; Elgin, S. C. Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. *Mol Cell Biol* 1986, *6* (11), 3862.

- (77) Muller, H. J.; Altenburg, E. The Frequency of Translocations Produced by X-Rays in Drosophila. *Genetics* 1930, 15 (4), 283.
- (78) Eissenberg, J. C.; James, T. C.; Foster-Hartnett, D. M.; Hartnett, T.; Ngan, V.; Elgin, S. C. Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. *Proc Natl Acad Sci* USA 1990, 87 (24), 9923.
- Hard, R.; Li, N.; He, W.; Ross, B.; Mo, G. C. H.; Peng, Q.; Stein, R. S. L.; Komives, E.;
 Wang, Y.; Zhang, J.et al. Deciphering and engineering chromodomain-methyllysine peptide recognition. *Sci Adv* 2018, *4* (11), eaau1447.
- (80) Singh, P. B.; Miller, J. R.; Pearce, J.; Kothary, R.; Burton, R. D.; Paro, R.; James, T. C.; Gaunt, S. J. A sequence motif found in a Drosophila heterochromatin protein is conserved in animals and plants. *Nucleic Acids Res* **1991**, *19* (4), 789.
- (81) Bannister, A. J.; Zegerman, P.; Partridge, J. F.; Miska, E. A.; Thomas, J. O.; Allshire, R. C.; Kouzarides, T. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 2001, *410* (6824), 120.
- (82) Sampath, S. C.; Marazzi, I.; Yap, K. L.; Sampath, S. C.; Krutchinsky, A. N.; Mecklenbrauker, I.; Viale, A.; Rudensky, E.; Zhou, M. M.; Chait, B. T.et al. Methylation

of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. *Mol Cell* **2007**, *27* (4), 596.

- (83) Vermeulen, M.; Eberl, H. C.; Matarese, F.; Marks, H.; Denissov, S.; Butter, F.; Lee, K. K.;
 Olsen, J. V.; Hyman, A. A.; Stunnenberg, H. G.et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* 2010, *142* (6), 967.
- (84) Kaustov, L.; Ouyang, H.; Amaya, M.; Lemak, A.; Nady, N.; Duan, S.; Wasney, G. A.; Li,
 Z.; Vedadi, M.; Schapira, M.et al. Recognition and specificity determinants of the human cbx chromodomains. *J Biol Chem* 2011, 286 (1), 521.
- (85) Dialynas, G. K.; Terjung, S.; Brown, J. P.; Aucott, R. L.; Baron-Luhr, B.; Singh, P. B.;
 Georgatos, S. D. Plasticity of HP1 proteins in mammalian cells. *J Cell Sci* 2007, *120* (Pt 19), 3415.
- (86) Aucott, R.; Bullwinkel, J.; Yu, Y.; Shi, W.; Billur, M.; Brown, J. P.; Menzel, U.; Kioussis, D.; Wang, G.; Reisert, I.et al. HP1-beta is required for development of the cerebral neocortex and neuromuscular junctions. *J Cell Biol* 2008, *183* (4), 597.
- (87) Mattout, A.; Aaronson, Y.; Sailaja, B. S.; Raghu Ram, E. V.; Harikumar, A.; Mallm, J. P.;Sim, K. H.; Nissim-Rafinia, M.; Supper, E.; Singh, P. B.et al. Heterochromatin Protein

1beta (HP1beta) has distinct functions and distinct nuclear distribution in pluripotent versus differentiated cells. *Genome Biol* **2015**, *16*, 213.

- (88) Milosevich, N.; Hof, F. Chemical Inhibitors of Epigenetic Methyllysine Reader Proteins.
 Biochemistry 2016, 55 (11), 1570.
- (89) Islam, K. The Bump-and-Hole Tactic: Expanding the Scope of Chemical Genetics. *Cell Chem Biol* 2018, 25 (10), 1171.
- (90) Seeliger, D.; Soeroes, S.; Klingberg, R.; Schwarzer, D.; Grubmuller, H.; Fischle, W. Quantitative assessment of protein interaction with methyl-lysine analogues by hybrid computational and experimental approaches. ACS Chem Biol 2012, 7 (1), 150.
- (91) Baril, S. A.; Koenig, A. L.; Krone, M. W.; Albanese, K. I.; He, C. Q.; Lee, G. Y.; Houk, K. N.; Waters, M. L.; Brustad, E. M. Investigation of Trimethyllysine Binding by the HP1 Chromodomain via Unnatural Amino Acid Mutagenesis. *J Am Chem Soc* 2017, *139* (48), 17253.
- (92) Ruan, J.; Ouyang, H.; Amaya, M. F.; Ravichandran, M.; Loppnau, P.; Min, J.; Zang, J.
 Structural basis of the chromodomain of Cbx3 bound to methylated peptides from histone h1 and G9a. *PLoS One* 2012, *7* (4), e35376.

- (93) Yan, H.; Xiang, X.; Chen, Q.; Pan, X.; Cheng, H.; Wang, F. HP1 cooperates with CAF-1 to compact heterochromatic transgene repeats in mammalian cells. *Sci Rep* 2018, 8 (1), 14141.
- (94) Gracheva, E.; Dus, M.; Elgin, S. C. Drosophila RISC component VIG and its homologVig2 impact heterochromatin formation. *PLoS One* 2009, *4* (7), e6182.
- (95) Fuks, F.; Hurd, P. J.; Deplus, R.; Kouzarides, T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res* 2003, *31* (9), 2305.
- (96) Piacentini, L.; Fanti, L.; Negri, R.; Del Vescovo, V.; Fatica, A.; Altieri, F.; Pimpinelli, S. Heterochromatin protein 1 (HP1a) positively regulates euchromatic gene expression through RNA transcript association and interaction with hnRNPs in Drosophila. *PLoS Genet* 2009, 5 (10), e1000670.
- (97) Netzer, C.; Rieger, L.; Brero, A.; Zhang, C. D.; Hinzke, M.; Kohlhase, J.; Bohlander, S. K. SALL1, the gene mutated in Townes-Brocks syndrome, encodes a transcriptional repressor which interacts with TRF1/PIN2 and localizes to pericentromeric heterochromatin. *Hum Mol Genet* 2001, *10* (26), 3017.
- Nielsen, A. L.; Ortiz, J. A.; You, J.; Oulad-Abdelghani, M.; Khechumian, R.; Gansmuller,
 A.; Chambon, P.; Losson, R. Interaction with members of the heterochromatin protein 1

(HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. *EMBO J* **1999**, *18* (22), 6385.

- (99) McDowell, T. L.; Gibbons, R. J.; Sutherland, H.; O'Rourke, D. M.; Bickmore, W. A.; Pombo, A.; Turley, H.; Gatter, K.; Picketts, D. J.; Buckle, V. J.et al. Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. *Proc Natl Acad Sci U S A* **1999**, *96* (24), 13983.
- (100) Ryu, H. W.; Lee, D. H.; Florens, L.; Swanson, S. K.; Washburn, M. P.; Kwon, S. H. Analysis of the heterochromatin protein 1 (HP1) interactome in Drosophila. *J Proteomics* 2014, *102*, 137.
- (101) Abdouh, M.; Hanna, R.; El Hajjar, J.; Flamier, A.; Bernier, G. The Polycomb Repressive Complex 1 Protein BMI1 Is Required for Constitutive Heterochromatin Formation and Silencing in Mammalian Somatic Cells. *J Biol Chem* **2016**, *291* (1), 182.
- (102) Prasanth, S. G.; Shen, Z.; Prasanth, K. V.; Stillman, B. Human origin recognition complex is essential for HP1 binding to chromatin and heterochromatin organization. *Proc Natl Acad Sci U S A* **2010**, *107* (34), 15093.
- (103) Shareef, M. M.; King, C.; Damaj, M.; Badagu, R.; Huang, D. W.; Kellum, R. Drosophila heterochromatin protein 1 (HP1)/origin recognition complex (ORC) protein is associated

with HP1 and ORC and functions in heterochromatin-induced silencing. *Mol Biol Cell* **2001,** *12* (6), 1671.

- (104) Rosnoblet, C.; Vandamme, J.; Volkel, P.; Angrand, P. O. Analysis of the human HP1 interactome reveals novel binding partners. *Biochem Biophys Res Commun* 2011, *413* (2), 206.
- (105) Cammas, F.; Janoshazi, A.; Lerouge, T.; Losson, R. Dynamic and selective interactions of the transcriptional corepressor TIF1 beta with the heterochromatin protein HP1 isotypes during cell differentiation. *Differentiation* **2007**, *75* (7), 627.
- (106) Le Douarin, B.; Nielsen, A. L.; Garnier, J. M.; Ichinose, H.; Jeanmougin, F.; Losson, R.; Chambon, P. A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors. *EMBO J* **1996**, *15* (23), 6701.
- (107) Knights, C. D.; Catania, J.; Di Giovanni, S.; Muratoglu, S.; Perez, R.; Swartzbeck, A.;
 Quong, A. A.; Zhang, X.; Beerman, T.; Pestell, R. G.et al. Distinct p53 acetylation cassettes differentially influence gene-expression patterns and cell fate. *J Cell Biol* 2006, *173* (4), 533.
- (108) Chau, C. M.; Deng, Z.; Kang, H.; Lieberman, P. M. Cell cycle association of the retinoblastoma protein Rb and the histone demethylase LSD1 with the Epstein-Barr virus latency promoter Cp. *J Virol* 2008, 82 (7), 3428.

- (109) Sudhamalla, B.; Dey, D.; Breski, M.; Nguyen, T.; Islam, K. Site-specific azide-acetyllysine photochemistry on epigenetic readers for interactome profiling. *Chem Sci* 2017, 8 (6), 4250.
- (110) Krishnamurthy, M.; Dugan, A.; Nwokoye, A.; Fung, Y. H.; Lancia, J. K.; Majmudar, C. Y.; Mapp, A. K. Caught in the act: covalent cross-linking captures activator-coactivator interactions in vivo. ACS Chem Biol 2011, 6 (12), 1321.
- (111) Liu, D. R.; Magliery, T. J.; Pastrnak, M.; Schultz, P. G. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc Natl Acad Sci U S A* **1997**, *94* (19), 10092.
- (112) Liu, D. R.; Magliery, T. J.; Schultz, P. G. Characterization of an 'orthogonal' suppressor tRNA derived from E. coli tRNA2(Gln). *Chem Biol* 1997, *4* (9), 685.
- (113) Eggertsson, G.; Soll, D. Transfer ribonucleic acid-mediated suppression of termination codons in Escherichia coli. *Microbiol Rev* 1988, 52 (3), 354.
- (114) Buchmueller, K. L.; Hill, B. T.; Platz, M. S.; Weeks, K. M. RNA-tethered phenyl azide photocrosslinking via a short-lived indiscriminant electrophile. *J Am Chem Soc* 2003, *125* (36), 10850.

- (115) Chen, Y.; Ebright, R. H. Phenyl-azide-mediated photocrosslinking analysis of Cro-DNA interaction. *J Mol Biol* **1993**, 230 (2), 453.
- (116) Liu, H.; Galka, M.; Mori, E.; Liu, X.; Lin, Y. F.; Wei, R.; Pittock, P.; Voss, C.; Dhami, G.;
 Li, X.et al. A method for systematic mapping of protein lysine methylation identifies functions for HP1beta in DNA damage response. *Mol Cell* 2013, *50* (5), 723.
- (117) Gupta, R.; Kapoor, N.; Raleigh, D. P.; Sakmar, T. P. Nucleobindin 1 caps human islet amyloid polypeptide protofibrils to prevent amyloid fibril formation. *J Mol Biol* 2012, *421* (2-3), 378.
- (118) Park, M.; Sivertsen, B. B.; Els-Heindl, S.; Huber, T.; Holst, B.; Beck-Sickinger, A. G.; Schwartz, T. W.; Sakmar, T. P. Bioorthogonal Labeling of Ghrelin Receptor to Facilitate Studies of Ligand-Dependent Conformational Dynamics. *Chem Biol* **2015**, *22* (11), 1431.
- (119) Ye, S.; Kohrer, C.; Huber, T.; Kazmi, M.; Sachdev, P.; Yan, E. C.; Bhagat, A.; RajBhandary, U. L.; Sakmar, T. P. Site-specific incorporation of keto amino acids into functional G protein-coupled receptors using unnatural amino acid mutagenesis. *J Biol Chem* 2008, 283 (3), 1525.
- (120) Shechter, D.; Dormann, H. L.; Allis, C. D.; Hake, S. B. Extraction, purification and analysis of histones. *Nat Protoc* **2007**, *2* (6), 1445.

- (121) Huber, T.; Sakmar, T. P. Probing Antibody Binding Sites on G Protein-Coupled Receptors
 Using Genetically Encoded Photo-Activatable Cross-Linkers. *Methods Mol Biol* 2018, 1785, 65.
- (122) Cano-Rodriguez, D.; Rots, M. G. Epigenetic Editing: On the Verge of Reprogramming Gene Expression at Will. *Curr Genet Med Rep* 2016, *4* (4), 170.
- (123) Jeffries, M. A. Epigenetic editing: How cutting-edge targeted epigenetic modification might provide novel avenues for autoimmune disease therapy. *Clin Immunol* 2018, 196, 49.
- (124) Falahi, F.; Sgro, A.; Blancafort, P. Epigenome engineering in cancer: fairytale or a realistic path to the clinic? *Front Oncol* **2015**, *5*, 22.
- (125) Stolzenburg, S.; Goubert, D.; Rots, M. G. Rewriting DNA Methylation Signatures at Will: The Curable Genome Within Reach? *Adv Exp Med Biol* 2016, *945*, 475.