

Structural and Biochemical Dissection of the KMT2 Core Complex

Pamela Peng Zhang

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the Doctorate in Philosophy degree in Biochemistry

Department of Biochemistry, Microbiology and Immunology
Faculty of Medicine
University of Ottawa

© Pamela Peng Zhang, Ottawa, Canada, 2015

ABSTRACT

Histone H3 lysine 4 (H3K4) methylation is an evolutionarily conserved mark commonly associated with transcription activation in eukaryotes. In mammals, this post-translational modification is deposited by the KMT2 family of H3K4 methyltransferases. Biochemical studies have shown that the enzymatic activity of the KMT2 enzymes is regulated by a core complex of four evolutionarily conserved proteins: WDR5, RbBP5, ASH2L and DPY30, collectively known as WRAD, which are all important for global H3K4 methylation. However, how these proteins interact and regulate the activity of the KMT2 enzymes is not well investigated. During my PhD, I have used structural and biochemical approaches to determine the interactions underlying formation of the core complex and regulation of KMT2 enzymatic activity. My research have shown that 1) WDR5 uses two peptide-binding clefts on opposite sides of its β -propeller domain to bridge the KMT2 enzymes to the regulatory subunit RbBP5, 2) the WDR5 peptidyl-arginine-binding cleft exhibits plasticity to accommodate the binding of all KMT2 enzymes and 3) RbBP5 S350 phosphorylation stimulates formation of the RbBP5-ASH2L complex and H3K4 methylation by the mammalian KMT2 enzymes. Collectively, these studies have provided the structural basis for understanding the important interactions governing KMT2 complex assembly and activity.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my PhD thesis supervisor **Dr. Jean-François Couture** for giving me the opportunity to begin in his lab first as undergraduate CO-OP student, a fourth year Honours student, a Master's candidate and finally a PhD candidate. I would like to thank Dr. Couture for introducing me to the world of structural biology and for initially inspiring me to pursue graduate studies. No words can describe Dr. Couture's constant guidance and support, without which I would not have thrived during my PhD. I would like to thank Dr. Couture for always pushing me to my highest potential, for instilling in me a passion and confidence for research and for ultimately nurturing me into the best researcher I can be!

Next, I would like to thank the past and present **Couture lab members** for their friendship and encouragements throughout the years and for creating a positive environment in the lab. I would especially like to thank **Sabina Sarvan** for her friendship every step of the way since our days as Honours students. I would also like to thank **members of the Blais, Figeys, Baetz and Rudner labs** as well as **Megan Tu** and **Sandy Szeto** for their friendship and support.

I would like to thank my thesis advisory committee members **Dr. Natalie Goto** and **Dr. Mary Alice Hefford** for their invaluable feedback in the past five years. I would also like to thank my PhD comprehensive exam members **Dr. Christopher N. Boddy, Dr. F. Jeffrey Dilworth, Dr. Mary Alice Hefford and Dr. Xiaohui Zha.**

I would also like to thank the various funding agencies and especially **CIHR** for granting me the Doctoral Research Award to support my PhD research.

Last but not least, I would like to thank **my family** for their unwavering support as well as my partner **Steven** for his friendship while I was conducting research in Kansas City and for his unyielding support during my last year of PhD. I would also like to thank him for showing me that California is the place to be!

TABLE OF CONTENTS

LIST OF ABBREVIATIONS	viii
LIST OF FIGURES AND TABLES	xii
CHAPTER 1. INTRODUCTION	1
1.1 Chromatin.....	1
1.2 Lysine Methylation.....	2
1.3 Histone Lysine Methylation.....	3
1.4 SET Domain.....	3
1.5 H3K4 Methylation in Transcription Activation.....	5
1.6 Evolutionary Conservation of the KMT2 Family of H3K4 Methyltransferases from Yeast to Mammals.....	6
1.7 Unique Functions of the Mammalian KMT2 Complexes.....	11
1.8 MLL1 in Leukemia.....	15
1.9 Regulatory Mechanisms Controlling H3K4 Methylation by the KMT2 Complexes... 16	
1.9.1 Site-Specific Recruitment of KMT2 Complexes.....	16
1.9.2 Regulation of H3K4 Methylation by Other Histone PTMs.....	18
1.10 Regulation of KMT2 Enzymatic Activity by the WRAD Complex.....	20
1.10.1 WDR5.....	25
1.10.2 RbBP5.....	25
1.10.3 ASH2L.....	27
1.10.4 DPY30.....	28
1.11 Rationale and Hypothesis.....	32
CHAPTER 2. STRUCTURAL AND BIOCHEMICAL INSIGHTS INTO MLL1 CORE COMPLEX ASSEMBLY	33
ABSTRACT.....	35
INTRODUCTION.....	36
RESULTS AND DISCUSSION.....	37
The C-terminal Domain of ASH2L Binds the Hinge Region of RbBP5.....	37
RbBP5 Binds WDR5 Using a Segment Neighboring its Hinge Region.....	44
Structure of the WDR5/RbBP5/MLL1 Ternary Complex.....	45
WDR5 V-Shaped Cleft and RbBP5 VDV Motif are Both Important for the Allosteric Regulation of MLL1 Methyltransferase Activity.....	49
Insights into the Organization of a Catalytically Competent WDR5-RbBP5-ASH2L-MLL1 Complex.....	53
Biological Functions of WDR5's V-Shaped Cleft.....	53
EXPERIMENTAL PROCEDURES.....	54
REFERENCES.....	59
ACKNOWLEDGEMENTS.....	62
CHAPTER 3. THE PLASTICITY OF WDR5 PEPTIDE-BINDING CLEFT ENABLES THE BINDING OF THE SET1 FAMILY OF HISTONE METHYLTRANSFERASES	63
ABSTRACT.....	65

INTRODUCTION.....	66
MATERIALS AND METHODS.....	68
RESULTS.....	71
Equilibrium Dissociation Constants of the Win Motifs for WDR5.....	71
Crystal Structures of the WDR5-MLL _{win} Complexes.....	76
Role of WDR5 in Stimulating the KMT Activity of the SET1 Family of Methyltransferases.....	87
DISCUSSION.....	93
MLL3, a Unique Member of the SET1 Family of HKMTs.....	94
The Expanding Role of WDR5.....	94
FUNDING.....	95
SUPPLEMENTARY DATA.....	95
ACKNOWLEDGEMENTS.....	96
FOOTNOTES.....	96
REFERENCES.....	96

CHAPTER 4. A PHOSPHORYLATION SWITCH ON RBBP5 REGULATES

HISTONE H3 LYS4 METHYLATION.....	100
ABSTRACT.....	102
INTRODUCTION.....	103
RESULTS AND DISCUSSION.....	105
Crystal Structure of ASH2L in Complex with RbBP5.....	105
Disruption of ASH2L/RbBP5 Interaction Impairs MLL1 Enzymatic Stimulation and Delays Erythroid Cell Terminal Differentiation.....	114
Phosphorylation of RbBP5 on S350 Potentiates WRAD Assembly.....	120
RbBP5 Phosphorylation: a Novel Regulatory Switch Controlling WRAD Assembly.....	124
RbBP5 Phosphorylation Controls Histone H3K4 Methylation by the KMT2 Enzymes.....	125
MATERIALS AND METHODS.....	134
ACKNOWLEDGEMENTS.....	135
REFERENCES.....	135
SUPPLEMENTARY RESULTS.....	138
Deconstruction of the WRAD Complex and its Allosteric Effect on KMT2 Enzymes.....	138
Deconvolution of ASH2L/RbBP5 Heterodimer Formation in MLL3 Interaction.....	139
An Evolutionarily Conserved D/E Box on RbBP5 is Important for Binding to ASH2L.....	140
SUPPLEMENTARY MATERIALS AND METHODS.....	141
SUPPLEMENTARY REFERENCES.....	144

CHAPTER 5. DISCUSSION.....	146
5.1 WDR5, a Scaffolding Protein of Chromatin Complexes.....	146
5.2 WDR5, a Scaffolding Protein of KMT2 Complexes.....	149
5.3 RbBP5 S350 Phosphorylation, an Additional Regulatory Mechanism Controlling H3K4 Methylation.....	150

5.4	ASH2L SPRY, a New Unique Phospho-reader.....	151
5.5	Limitations.....	155
5.6	Targeting MLL1 Complex Assembly for Leukemia Drug Development.....	155
5.7	Concluding Remarks.....	157
REFERENCES (FOR CHAPTERS 1 AND 5).....		158
CONTRIBUTIONS OF COLLABORATORS.....		169
Curriculum Vitae.....		171
COPYRIGHT PERMISSIONS.....		175

LIST OF ABBREVIATIONS

A

ABM	ASH2L-binding motif (of RbBP5)
Ac	<i>Anolis carolinensis</i>
AF10	ALL1-fused gene from chromosome 10 protein
Ap2delta	Activating enhancer-binding protein 2-delta
ASC-2	Activating signal cointegrator-2
ASCOM	ASC-2 complex
Ash2	Absent, small, or homeotic discs protein 2
Ash2l/ASH2L	Ash2-like protein
At	<i>Arabidopsis thaliana</i>
ATAC	Ada Two-A containing complex

B

BME	β -mercaptoethanol
BOD1	Biorientation of chromosomes in cell division protein 1
BPTF	Bromodomain and PHD finger-containing transcription factor

C

CCNA2	<i>Cyclin-A2</i>
Ce	<i>Caenorhabditis elegans</i>
CFP1	CXXC-type zinc finger protein 1
CHD1	Chromodomain helicase DNA binding protein 1
CHD8	Chromodomain helicase DNA binding protein 8
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation followed by high-throughput DNA sequencing
COMPASS	COMplex Proteins ASSociated with Set1
Cryo-EM	Cryo-electron microscopy
CPY33	Cyclophilin-33

D

DBM	DPY30-binding motif (of ASH2L)
DD	Dimerization domain (of DPY30)
D/E	Aspartate/glutamate box (of RbBP5)
Dm	<i>Drosophila melanogaster</i>
Dot1	Disrupter of telomere silencing protein 1
Dox	Doxycycline
Dpy30/DPY30	Protein dumpy-30
Dr	<i>Dario rerio</i>

E

ESCs	Embryonic stem cells
ESI-MS	Electrospray ionization mass spectrometry
EZH2	Enhancer of zeste homolog 2

F	
FL	Full-length
FOXO4	Forkhead box protein O4
FYRN/FYRC	FY-rich N-/C-terminus (of MLL1)
G	
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
Gg	<i>Gallus gallus</i>
H	
H	Hinge (of RbBP5)
HC	Hinge + C-terminus (of RbBP5)
H2Bub	H2B ubiquitylation
H3K4	Histone H3 lysine 4
H3K4me1	H3K4 mono-methylation
H3K4me2	H3K4 di-methylation
H3K4me3	H3K4 tri-methylation
H3K9	Histone H3 lysine 9
H3K27me3	Histone H3 lysine 27 tri-methylation
H3R2me2a	H3R2 asymmetric di-methylation
H4K16ac	Histone H4 K16 acetylation
HCF1	Host cell factor 1
HDAC3	Histone deacetylase 3
HKMT	Histone lysine methyltransferase
HMG	High mobility group box (of MLL4)
<i>HOTTIP</i>	<i>HOXA</i> transcript at the distal tip
<i>Hox</i>	Homeobox
Hs	<i>Homo sapiens</i>
HS2	Hypersensitive site 2 (of <i>β-globin</i> locus)
HSCs	Hematopoietic stem cells
I	
<i>IFN-γ</i>	Interferon gamma gene
IRF3	Interferon regulatory factor 3
ITC	Isothermal titration calorimetry
K	
K	Lysine
K _D /K _d	Equilibrium dissociation constant
KMTs/KMTases	Lysine methyltransferases
KMT2	Lysine methyltransferase 2
L	
LCR	Locus control region (of <i>β-globin</i> locus)
lncRNAs	Long noncoding RNAs
LPS	Lipopolysaccharide

M

MBM	Menin-binding motif (of MLL1 and MLL2)
MBT	Malignant brain tumor repeat
Mef2d	Myocyte-specific enhancer factor 2D
MEL	Murine erythroid leukemia cell
MLL1-4	Myeloid/lymphoid or mixed-lineage leukemia proteins 1-4
MSL	Male-specific lethal complex
Myc/MYC	Myc proto-oncogene protein
<i>Myf5</i>	<i>Myogenic factor 5</i>
m/z	mass/charge

N

<i>NeST</i>	Nettoie Salmonella pas Theiler's; cleanup Salmonella not Theiler's
NRC/NCoA6	Nuclear receptor coactivator 6
NSL	Non-specific lethal complex
NURF	Nucleosome remodeling factor

P

PA1	PTIP-associated protein 1
Pax2	Paired box protein Pax-2
Pax7	Paired box protein Pax-7
PBS	Phosphate buffer saline
PBST	PBS supplemented with 0.1% Triton-X 100
PEG	Polyethylene glycol
PEV	Position effect variegation
PH	β -Propeller + Hinge (of RbBP5)
PHD	Plant homeodomain finger
PHD3	Third PHD finger (of MLL1)
Phe	Phenylalanine
Pitx2	Pituitary homeobox 2
PRMT6	Protein arginine N-methyltransferase 6
PTIP	PAX-interacting protein 1
PTMs	Post-translational modifications

R

RAR	Retinoic acid receptor
Rbbp5/RbBP5	Retinoblastoma-binding protein 5
r.m.s.	Root-mean-square
r.m.s.d.	Root-mean-square deviation
RNA Pol II	RNA polymerase II
RRM	RNA recognition motif (of SET1A and SET1B)

S

SAM or AdoMet	<i>S</i> -adenosyl methionine
SAR	Scaffold attachment region
Sc	<i>Saccharomyces cerevisiae</i>

SDI	Sdc1/DPY30 interaction motif (of ASH2L)
SET	Suppressor of variegation 3-9, Enhancer of Zeste and Trithorax
Sh	<i>Sarcophilus harrisii</i>
Sp	<i>Schizosaccharomyces pombe</i>
SPRY	SPLa and RYanodine Receptor domain (of ASH2L)
T	
TAF3	Transcription initiation factor TFIID subunit 3
Tbx1	T-box transcription factor TBX1
TFIID	TATA-box-binding protein
Trr	Trithorax-related protein
Trx	Trithorax
TrxG	Trithorax group
Tyr	Tyrosine
V	
VDV	Valine-aspartate-valine motif (of RbBP5)
VISA	Mitochondrial antiviral-signaling protein
W	
WBM	WDR5-binding motif (of RbBP5)
WD40	Tryptophan-aspartate repeat
WDR5	WD repeat-containing protein 5
WDR82	WD repeat-containing protein 82
Wds	Protein will die slowly
WH	Winged-helix motif (of ASH2L)
Win	WDR5-interacting motif (of mammalian KMT2 enzymes)
WRAD	WDR5, RbBP5, ASH2L and DPY30
WT	Wild-type
X	
Xl	<i>Xenopus laevis</i>
Xt	<i>Xenopus tropicalis</i>

LIST OF FIGURES AND TABLES

CHAPTER 1. INTRODUCTION

Figure 1.1. Subunit composition of COMPASS and COMPASS-like complexes from budding yeast to mammals. (Page 8)

Figure 1.2. Domain architecture of the mammalian KMT2 enzymes. (Page 13)

Figure 1.3. Cryo-EM structure of the Y-shaped MLL1 core complex. (Page 23)

Figure 1.4. Domain architecture of the core complex subunits WRAD. (Page 30)

Table 1.1. Budding yeast COMPASS subunits and their mammalian homologs. (Page 10)

CHAPTER 2. STRUCTURAL AND BIOCHEMICAL INSIGHTS INTO MLL1 CORE COMPLEX ASSEMBLY

Figure 2.1. ASH2L and RbBP5 harbor several evolutionarily conserved domains. (Page 40)

Figure 2.2. Dissection of the core complex. (Page 42)

Figure 2.3. WDR5 binds RbBP5 and MLL1 using two independent binding clefts. (Page 47)

Figure 2.4. WDR5 V-shaped cleft and RbBP5 VDV motif are important for the stimulation of MLL methyltransferase activity. (Page 51)

Table 2.1. Data Collection and Refinement Statistics for WDR5 Structure in Complex with RbBP5 and MLL1 (Page 58)

CHAPTER 3. THE PLASTICITY OF WDR5 PEPTIDE-BINDING CLEFT ENABLES THE BINDING OF THE SET1 FAMILY OF HISTONE METHYLTRANSFERASES

Figure 3.1. WDR5 binds MLL1-4, SET1A and SET1B Win motifs. (Page 74)

Figure S3.1. A clear electron density map is observed for all the Win motifs. (Supplementary Data) (Page 81)

Figure 3.2. Crystal structure of WDR5 in complex with MLL2-4, SET1A and SET1B Win motifs. (Page 83)

Figure 3.3. Comparative analysis of Win motif binding modes. (Page 85)

Figure 3.4. Stimulation of the SET1 family of KMT activity on histone H3 by the core complex. (Page 89)

Figure 3.5. MLL3 methylates histone H3 in the absence of the core complex subunits. (Page 91)

Table 3.1. Alignment of SET1 Win Motifs (Page 73)

CHAPTER 4. A PHOSPHORYLATION SWITCH ON RBBP5 REGULATES HISTONE H3 LYS4 METHYLATION

Figure S4.1. Identification of the ASH2L-binding motif (ABM). (Supplementary Data) (Page 107)

Figure 4.1. The ASH2L SPRY domain binds a D/E box on RbBP5. (Page 110)

Figure S4.2. Simulated annealing map of RbBP5 peptides. (Supplementary Data) (Page 112)

Figure 4.2. Interaction between ASH2L and RbBP5 is essential for terminal differentiation of erythroid cells. (Page 116)

Figure S4.3. RbBP5 D/E box region is important for the stimulation of the SET1 family of lysine methyltransferases. (Supplementary Data) (Page 118)

Figure 4.3. Phosphorylation of RbBP5 stimulates WRAD complex formation. (Page 122)

Figure 4.4. RbBP5 S350 phosphorylation increases the catalytic activity of MLL3. (Page 128)

Figure S4.4. RbBP5 S350 phosphorylation increases the rate of MLL1 methylation. (Supplementary Data) (Page 130)

Figure S4.5. RbBP5/ASH2L/MLL3 and RbBP5^{phos}/ASH2L/MLL3 complexes display the same kinetics for H3K4 di- and tri-methylation. (Supplementary Data) (Page 132)

Table S4.1. Data Collection and Refinement Statistics for ASH2L SPRY in Complex with RbBP5 Peptides (Supplementary Data) (Page 109)

(Supplementary Data) These figures and tables are included accordingly in the thesis but are found in the Supplementary Data of each paper.

CHAPTER 5. DISCUSSION

Figure 5.1. Structural comparison of ASH2L SPRY and the known phospho-binding domains. (Page 153)

CHAPTER 1. INTRODUCTION

1.1 Chromatin

Chromatin is the fundamental DNA packaging apparatus in the eukaryotic cell nucleus and is composed of basic building blocks called nucleosomes, which are DNA-protein complexes composed of about 150 base pairs of DNA wrapped around a histone protein octamer of two copies each of histones H2A, H2B, H3 and H4 (Kornberg, 1974). This array of nucleosomes is tightly condensed in the nucleus by undergoing additional levels of folding (Finch & Klug, 1976). While this sophisticated packaging is essential for storing the vast amount of genetic information, the equally intricate unfolding and unwinding of chromatin is crucial for various cellular processes. Chromatin structure is regulated by ATP-dependent chromatin remodeling complexes (Peterson & Herskowitz, 1992), the incorporation of histone variants (Franklin & Zweidler, 1977), DNA modifications (Razin & Cedar, 1977) and histone post-translational modifications (PTMs) (Murray, 1964). While these chromatin modification events are defined as “epigenetic” changes (Waddington, 1942), which are changes that do not change the underlying DNA sequence (Goldberg, Allis, & Bernstein, 2007), these changes ultimately control the accessibility of DNA to influence DNA-related processes, including replication, repair and transcription. One type of epigenetic modification involves the deposition of histone PTMs, including acetylation, phosphorylation, sumoylation, ubiquitylation and methylation (Bannister & Kouzarides, 2011). These histone PTMs may act alone or in a coordinated fashion to activate or repress chromatin-mediated events.

1.2 Lysine Methylation

Lysine (K) methylation is a dynamic modification involving the addition of a maximum of three methyl groups to the epsilon amino group of a target lysine (Lanouette, Mongeon, Figeys, & Couture, 2014). Unlike other modifications, such as acetylation and phosphorylation, methylation of a lysine residue does not affect the charge of its positive side chain. In addition, the mass difference caused by the addition of a methyl group to a lysine side chain is relatively small when compared to other PTMs, such as sumoylation and ubiquitylation. Nevertheless, lysine methylation prevents the deposition of other PTMs on the same residue or influences the deposition of PTMs on neighboring residues (Lanouette et al., 2014). For example, large-scale mass spectrometry analyses on budding yeast proteins have shown that lysine methylation may prevent ubiquitylation on 43% of lysine methylation sites and thus may increase the stability of these proteins (Pang, Gasteiger, & Wilkins, 2010). However, in most cases, the function of lysine methylation is manifested by effector proteins that specifically recognize methylated lysine residues. These effector proteins harbor specific domains such as double chromodomains, double/tandem tudor domains, malignant brain tumor (MBT) repeats, plant homeodomain (PHD) fingers and tryptophan-aspartate (WD40) repeats (Taverna, Li, Ruthenburg, Allis, & Patel, 2007). In addition, the ability of methyl-lysine readers to discriminate between the different degrees of methylation, mono-, di- or tri-methylation, serves as an additional level of regulation in methyl-lysine signaling.

Since its initial discovery in 1959 (Ambler & Rees, 1959), over a thousand lysine methylation events have been documented on both histone and non-histone substrates (Lanouette et al., 2014). In particular, lysine methylation on histones H3, H4 and the linker histone H1B has important roles in regulating chromatin dynamics and transcription (Dillon, Zhang, Trievel, & Cheng, 2005; Martin & Zhang, 2005).

1.3 Histone Lysine Methylation

The first histone lysine methylation event was observed by Murray in 1964 (Murray, 1964). While it was unknown at the time, the importance of histone lysine methyltransferases in gene expression was observed even earlier in a phenomenon known as position effect variegation (PEV) in *Drosophila melanogaster* that was initially described by Mueller in 1930 (Muller, 1930), whereby euchromatic genes brought near heterochromatic regions were silenced. Later studies of PEV led to the identification of the first histone lysine methyltransferase that is specific for histone H3 lysine 9 (H3K9), suppressor of variegation 3-9 (Tschiersch et al., 1994). Since the discovery of H3K9 methylation, several methylated lysine residues on histones H3, H4 and the linker histone H1B have been identified. Depending on the site of the residue being methylated, the histone lysine methylation event leads to various biological processes including heterochromatin formation, X-chromosome inactivation and transcription regulation (Martin & Zhang, 2005). While histones H4 and H1B each have one lysine methylation site, methylation can occur on five histone H3 lysine residues. Genome-wide profiling of histone H3 lysine methylation patterns have shown that while methylation of K4, K36 and K79 are associated with transcriptionally active genes, methylation of K9 and K27 are associated with transcriptionally silenced genes (Barski et al., 2007).

1.4 SET Domain

Histone lysine methylation is predominantly catalyzed by a group of lysine methyltransferases (KMTs) that harbors a catalytic SET domain, named after the first three identified proteins harboring this domain in *Drosophila*: Suppressor of variegation 3-9 (Tschiersch et al., 1994), Enhancer of Zeste and Tritrithorax (Jones & Gelbart, 1993). Despite

differences in substrate selectivity, biological roles and tissue distribution, all SET domain enzymes use the cofactor *S*-adenosyl methionine (SAM or AdoMet) as the methyl donor.

Structural studies of SET domain proteins have illustrated the precise coordination of the cofactor and the substrate at the active site. The cofactor and the substrate bind on opposite sides of the SET domain and are connected by a narrow hydrophobic channel in which the lysine side chain is inserted and is closely aligned with the leaving methyl group of SAM at the end of the channel (Dillon et al., 2005; Qian & Zhou, 2006). Sequence alignment of SET domains shows that several conserved motifs and residues participate in cofactor and substrate binding as well as catalysis.

While the substrate specificity is determined by consensus motifs harboring the target lysine (Qian & Zhou, 2006), the intrinsic product specificity of SET domain enzymes depends on the presence of either a phenylalanine (Phe) or a tyrosine (Tyr) residue, commonly referred to as the Phe/Tyr switch, in the SET domain active site whereby the presence of a Tyr residue at the switch site restricts a given SET domain enzyme to a mono-methyltransferase (Collins et al., 2005; Couture, Collazo, Brunzelle, & Trievel, 2005; X. Zhang et al., 2003). While this rule can generally be applied to most SET domain methyltransferases, sequence alignment of SET domain enzymes reveals that the tri-methyltransferases enhancer of zeste homolog 2 (EZH2) and mixed-lineage leukemia 1 (MLL1) harbor a Tyr residue at their switch sites. Further biochemical studies have shown that the product specificity of these enzymes is controlled by regulatory proteins associating with the EZH2 (Kuzmichev, Nishioka, Erdjument-Bromage, Tempst, & Reinberg, 2002) or the MLL1 (Dou et al., 2006) SET domain.

1.5 H3K4 Methylation in Transcription Activation

Methylation of lysine 4 on histone H3 (H3K4) is an evolutionarily conserved mark that is commonly associated with transcription activation in eukaryotes. Genome-wide studies have shown that H3K4 mono- (H3K4me1), di- (H3K4me2) and tri-methylation (H3K4me3) marks are distinctly distributed across the eukaryotic genome. Despite minor differences in the H3K4 methylation patterns between yeast and metazoans, H3K4me1 and H3K4me2 are located at enhancers (Heintzman et al., 2007) and 5' ends of transcribed genes (T. Kim & Buratowski, 2009; R. Schneider et al., 2004), respectively, while H3K4me3 is located at the promoters of both “poised” and actively transcribed genes (Santos-Rosa et al., 2002; R. Schneider et al., 2004). In most instances, H3K4me3 serves as a binding platform for the recruitment of other complexes that influence chromatin structure and accessibility to the transcription machinery. For example, H3K4me3 is directly recognized by several proteins associated with chromatin remodeling complexes, such as chromodomain helicase DNA binding protein 1 (CHD1) (Sims et al., 2005) and the bromodomain and PHD finger-containing transcription factor (BPTF) subunit of the nucleosome remodeling factor (NURF) complex (Wysocka et al., 2006). In addition, H3K4me3 recruits the basal transcription factor TFIID of the transcription machinery via the PHD finger-containing TAF3 subunit (Vermeulen et al., 2007), supporting recent genome-wide studies showing co-occupancy of H3K4me3 with RNA polymerase II (RNA Pol II) (P. Wang et al., 2009).

When deposited in the promoter regions of “poised” genes, H3K4me3 co-localizes with the repressive histone H3 lysine 27 tri-methylation (H3K27me3) mark. Also referred to as bivalent domains in embryonic stem cells (ESCs) (Bernstein et al., 2006), these discrete regions contain histone H3 proteins that are tri-methylated on K4 and K27, two marks that are typically found in mutually exclusive locations in cells. Bivalent domains contain

developmentally important transcription factor genes poised for activation depending on cellular cues for self-renewal or differentiation. The expression of these poised genes are silenced by the presence of H3K27me3 but can be activated or further repressed upon demethylation of H3K27me3 or H3K4me3, respectively. Chromatin immunoprecipitation (ChIP) experiments in mouse ESCs have shown that maximal level of H3K4me3 is not required for the expression of most genes, including self-renewal genes, but is required for the expression of developmental genes during differentiation (Jiang et al., 2011). Similarly, H3K4me3 profiling during zebrafish embryonic development have shown that even though over 80% of genes are marked by H3K4me3 after genome activation, only a portion of these genes are expressed or bound by RNA Pol II (Vastenhouw et al., 2010). These results are consistent with initial studies showing that loss of H3K4 methylation by destabilizing the sole H3K4 methyltransferase complex has minimal effect on global gene transcription in the budding yeast (Miller et al., 2001). Altogether, these examples highlight the punctate functions of H3K4 methylation in defined biological processes and the expanded repertoire of biological functions performed by the different histone H3K4 methyltransferases (see below).

1.6 Evolutionary Conservation of the KMT2 Family of H3K4 Methyltransferases from Yeast to Mammals

In the budding yeast *Saccharomyces cerevisiae*, H3K4 methylation is catalyzed by a unique multi-protein complex referred to as COMplex Proteins ASSociated with Set1 (COMPASS) (Miller et al., 2001). Within COMPASS, Set1 associates with its regulatory proteins (Csp15, Cps25, Cps30, Cps35, Cps40, Cps50 and Cps60) (Figure 1.1) and is responsible for catalyzing mono-, di- and tri-methylation of H3K4 (Krogan et al., 2002;

Roguev et al., 2001). In *Drosophila melanogaster*, there are three homologs of yeast Set1: *Drosophila* Set1, Trithorax (Trx) and Trithorax-related (Trr) (Mohan et al., 2011). These enzymes are found in three distinct COMPASS-like complexes (Figure 1.1) and knockdown of Set1, Trx or Trr results in lethality, suggesting that they all play important and non-redundant functions in *Drosophila* development (Mohan et al., 2011). Despite a higher degree of diversification in the fruit fly, a common core complex of four proteins is retained in each of the three *Drosophila* COMPASS-like complexes, namely will die slowly (Wds), retinoblastoma-binding protein 5 (Rbbp5), absent, small, or homeotic discs protein 2 (Ash2) and dumpy-30 (Dpy30) (Figure 1.1). In mammals, the KMT2 family is further diversified and includes two homologs of Trx (MLL1 and MLL2), Trr (MLL3 and MLL4) and Set1 (SET1A and SET1B) (Shilatifard, 2012). These enzymes are found in six distinct COMPASS-like complexes that also contain the evolutionarily conserved core complex of four proteins, namely WD repeat-containing protein 5 (WDR5), RbBP5, Ash2-like protein (ASH2L) and DPY30, collectively known as WRAD (Figure 1.1 and Table 1.1). However, despite sharing the ability to site-specifically methylate H3K4, the mammalian COMPASS-like complexes perform distinct functions.

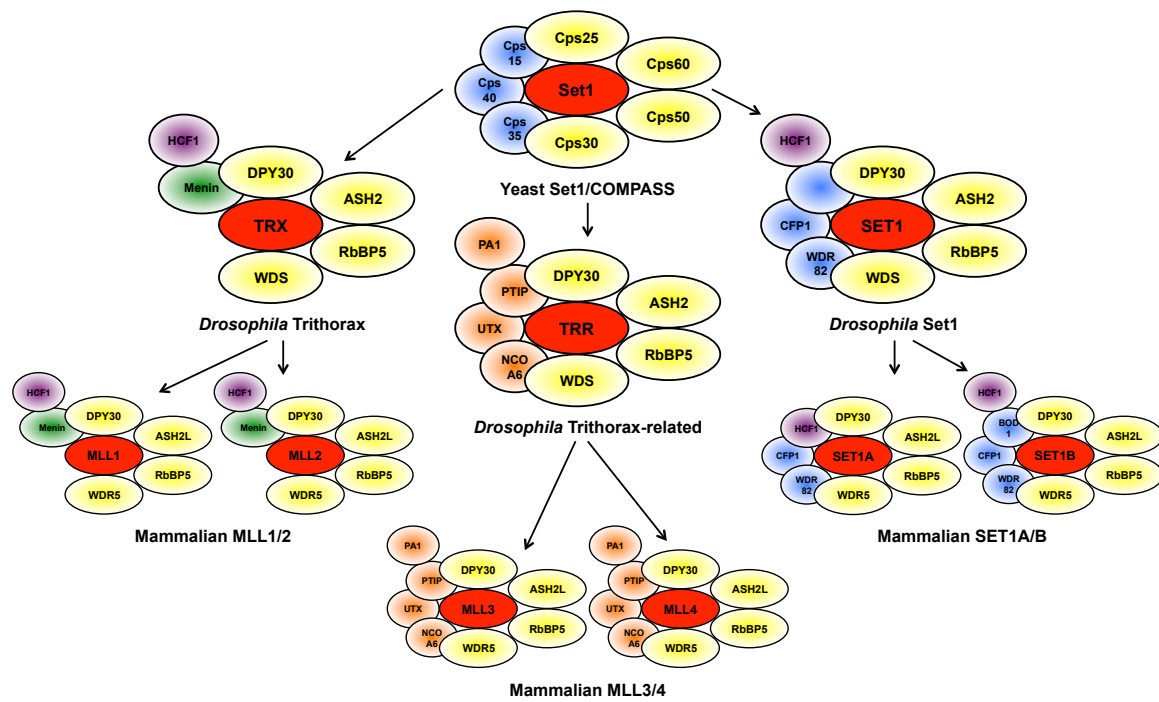


Figure 1.1. Subunit composition of COMPASS and COMPASS-like complexes from budding yeast to mammals.

The KMT2 enzymes are shown in red and the conserved core complex subunits are shown in yellow. The unique subunits are shown in purple, green, orange and blue. Figure is adapted from Shilatifard, 2012.

Table 1.1. Budding yeast COMPASS subunits and their mammalian homologs.		
Yeast COMPASS	Mammalian KMT2 complexes	Function in mammalian complexes
Set1	MLL1-4, SET1A/B	Catalytic subunits
Cps30	WDR5	Core subunits for complex stability
Cps50	RbBP5	
Cps60	ASH2L	Core subunits for H3K4me3
Cps25	DPY30	
Cps40	CFP1	SET1A/B subunits for H3K4me3
Cps35	WDR82	
Cps15	BOD1	SET1B subunit

1.7 Unique Functions of the Mammalian KMT2 Complexes

All six mammalian KMT2 family members participate redundantly in H3K4 methylation, thus deletion of any member has minimal effect on global H3K4 methylation (P. Wang et al., 2009). However, purification of the different KMT2 enzymes reveals that each COMPASS-like complex contains unique subunits (van Nuland et al., 2013). This diversity, in addition to the divergent domain architecture outside of the SET domain (Figure 1.2), reflects their gene-specific and cell-specific functions.

MLL1 and MLL2 are important in hematopoiesis and embryonic development, respectively (Glaser et al., 2006; Yu, Hess, Horning, Brown, & Korsmeyer, 1995). Correspondingly, genetic deletion of either *Mll1* or *Mll2* results in embryonic lethality in mice, suggesting that they both have important and non-redundant functions during mice embryonic development (Glaser et al., 2006; Yu et al., 1995). In stem cells, MLL1 performs gene-specific functions by maintaining the expression of homeobox (*Hox*) and neuronal genes during development (Lim et al., 2009; P. Wang et al., 2009). Consistent with its role in embryonic development, MLL2 tri-methylates H3K4 at bivalent domains in mouse ESCs, including the *Hox* gene cluster (Hu et al., 2013b). The unique roles of MLL1 and MLL2 are in part determined by their ability to interact with the protein menin, which is important for recruiting these enzymes to specific *Hox* genes (Hughes et al., 2004; P. Wang et al., 2009; Yokoyama et al., 2004).

MLL3 and MLL4 are important in nuclear signaling. This activity is controlled by their ability to associate with several unique subunits, including the PAX-interacting protein 1 (PTIP), PTIP-associated protein 1 (PA1), the H3K27 demethylase UTX (Y. W. Cho et al., 2007), and activating signal cointegrator-2 (ASC-2) (Goo et al., 2003). As subunits of ASC-2 complex (ASCOM), MLL3 or MLL4 associate with several nuclear receptors, including bile

acid receptor (Ananthanarayanan et al., 2011), retinoic acid receptor (RAR) (S. Lee et al., 2006) and liver X receptor (S. Lee, Lee, Lee, & Lee, 2008), and regulate the expression of their responsive genes. In addition to their gene-specific roles, MLL3 and MLL4 mono-methylate H3K4 at enhancer regions in several cell types (Hu et al., 2013a).

In contrast to the gene-specific regulatory roles of MLL1-4, SET1A and SET1B, along with their unique interacting partner WD repeat-containing protein 82 (WDR82), maintain global levels of H3K4me3 in mammalian cells (M. Wu et al., 2008). This activity is mediated by the ability of SET1A and SET1B to associate with CXXC-type zinc finger protein 1 (CFP1), which specifically binds unmethylated CpG islands (Voo, Carlone, Jacobsen, Flodin, & Skalnik, 2000) and recruits these enzymes to CpG islands for H3K4me3 (Thomson et al., 2010). The Timmers laboratory has recently identified a new subunit of SET1A and SET1B complexes, biorientation of chromosomes in cell division protein 1 (BOD1) (van Nuland et al., 2013), a kinetochore protein important for proper chromosomal segregation during mitosis (Porter et al., 2007). While the role of BOD1 within COMPASS is unknown, WDR82 binds the C-terminal domain of RNA Pol II and recruits SET1A and SET1B complexes to the transcription start site of transcribed genes for H3K4me3 (Lee & Skalnik, 2008). Intriguingly, WDR82, CFP1 and BOD1 are the mammalian homologs of yeast COMPASS subunits Cps35, Cps40 and Cps15, respectively (Table 1.1). The identification of yeast COMPASS homologs in SET1A and SET1B complexes reflects their evolutionary conservation from yeast Set1/COMPASS but also highlights the segregated subunit composition of the different mammalian COMPASS-like complexes compared to yeast COMPASS (Figure 1.1 and Table 1.1).

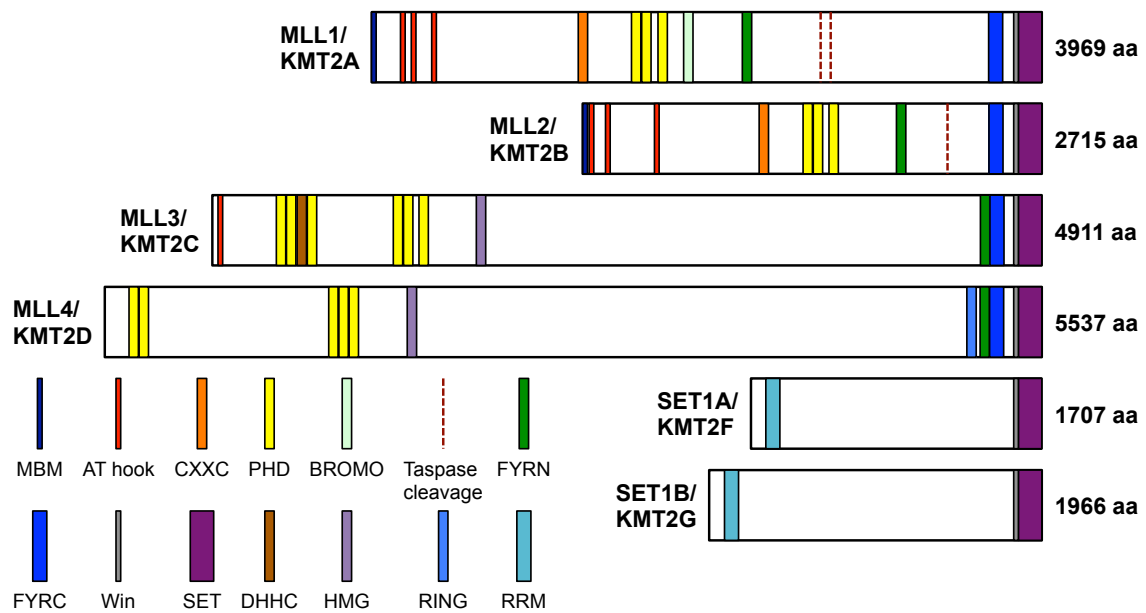


Figure 1.2. Domain architecture of the mammalian KMT2 enzymes.

The Win (WDR5-interacting) motif and the catalytic SET domain are located at the extreme N-terminus of each protein. Figure is adapted from Ernst & Vakoc, 2012.

MBM: menin-binding motif; CXXC domain; PHD finger; BROMOdomain; FYRN/FYRC: FY-rich N-/C-terminus; DHHC domain; HMG (high mobility group) box, RING finger; RRM: RNA recognition motif

1.8 MLL1 in Leukemia

Mutations of several KMT2 enzymes are associated with several types of cancer (Huntsman et al., 1999; Ruault, Brun, Ventura, Roizes, & De Sario, 2002). Due to its involvement in over 70% of acute childhood leukemia cases (Krivtsov & Armstrong, 2007), the mechanisms by which MLL1 causes leukemia has been heavily investigated. The *MLL1* gene is often targeted during oncogenic chromosomal translocations, resulting in fusion of its N-terminus with over 50 partner genes of diverse functions (Krivtsov & Armstrong, 2007). While one allele of MLL1 undergoes translocation and fusion, the other copy remains unmodified. Owing to the involvement of MLL1 in hematopoiesis, MLL1 fusion proteins alter transcription patterns of key developmental genes during stem cell maintenance and differentiation (Armstrong et al., 2002). In hematopoietic stem cells (HSCs), wild-type (WT) MLL1 regulates the expression of essential HSC self-renewal genes, such as *HOXA9* and the *Hox* cofactor *MEIS1*, which are down-regulated after differentiation (Zeisig et al., 2004). However, in leukemia patients, MLL1 fusion proteins, such as the well-characterized AF10, constitutively activate these genes through recruitment of the H3K79-specific methyltransferase Dot1 (Okada et al., 2005), resulting in the abnormal accumulation of HSCs known as leukemic transformation.

Since the SET domain of MLL1 is consistently absent in MLL1 fusion proteins, the role of WT MLL1 enzymatic activity in leukemia was elusive until a seminal study by Thiel et al. showed that upon recruitment of WT MLL1 by menin to its target genes, including *HOXA9* and the pro-proliferating cell cycle gene *CCNA2*, the enzymatic activity of WT MLL1 is required for the maximal expression of these genes, resulting in proliferation of leukemia cells (Thiel et al., 2010). This study clarified the role of WT MLL1 enzymatic activity in MLL1-fusion leukemia and provided the basis for targeting WT MLL1 activity in

treating leukemia.

1.9 Regulatory Mechanisms Controlling H3K4 Methylation by the KMT2

Complexes

The link between MLL1 and cancer suggests that several mechanisms converge to regulate MLL1 activity, including the site-specific recruitment of the enzyme to specific genes, the impact of other PTMs deposited on histone proteins and the allosteric regulation of MLL1 enzymatic activity by its regulatory proteins.

1.9.1 Site-Specific Recruitment of KMT2 Complexes

Recruitment of KMT2 complexes to their target genes is regulated by several mechanisms that involve direct interaction between specific subunits of KMT2 complexes with histone H3, transcription factors and long noncoding RNAs (lncRNAs). Several domains on the N-terminus of MLL1 have been characterized for their importance in recruiting MLL1, as well as MLL1 fusion proteins, to specific target genes (Figure 1.2). First, menin binds the menin-binding motif (MBM) of MLL1 to target the enzyme to specific *Hox* genes (Caslini et al., 2007; Yokoyama et al., 2005; Yokoyama et al., 2004). Adjacent to the MBM, the AT hook domain specifically binds the cruciform structure of DNA (Zelevnik-Le, Harden, & Rowley, 1994) and the AT-rich scaffold attachment region (SAR) of DNA (Broeker, Harden, Rowley, & Zelevnik-Le, 1996) near the attachment site of the nuclear matrix. Correspondingly, Caslini et al. showed that the three AT hooks of MLL1 localize several MLL1 fusion proteins to the nuclear matrix (Caslini et al., 2000). While the PHD fingers are consistently absent in MLL1 fusion proteins, cooperation between the CXXC domain and the PHD fingers of WT MLL1 is important for targeting both WT MLL1 and the MLL1 fusion protein to the *HoxA9* locus (Milne et al., 2010). In addition, a proline residue

within the linker region between the third PHD finger (PHD3) and the adjacent bromodomain is subjected to cis-trans isomerization by the peptidyl prolyl isomerase CYP33, resulting in down-regulation of *HOXA9* and *HOXC8* genes (Z. Wang et al., 2010). The lack of PHD3, and thus lack of CYP33 binding, in MLL1 fusion proteins may explain their constitutive activation of *Hox* genes and correspondingly, reinsertion of PHD3 in several MLL1 fusion proteins results in down-regulation of *HOXC8* (J. Chen et al., 2008). Finally, a region located at the C-terminus of MLL1 and MLL2 binds the arginine methylated form of the transcription factor Pax7 for controlling the transcription of the myogenic determination factor *Myf5* during skeletal muscle development (Kawabe, Wang, McKinnell, Bedford, & Rudnicki, 2012).

Analogous to the binding of menin to MLL1 and MLL2, the PTIP subunit of MLL3 and MLL4 complexes interacts with the developmental regulator Pax2 (Lechner, Levitan, & Dressler, 2000) and recruits these complexes for H3K4 methylation at a Pax2 response element (S. R. Patel, Kim, Levitan, & Dressler, 2007). The CFP1 and WDR82 subunits of SET1A and SET1B complexes bind DNA (Voo et al., 2000) and RNA Pol II (Lee & Skalnik, 2008), respectively, and recruit these complexes for H3K4me3 at gene promoter regions. Lastly, host cell factor 1 (HCF1), a subunit of MLL1/MLL2 and SET1A/SET1B complexes, selectively binds to the E2F transcription activator E2F1 and recruits KMT2 complexes for H3K4me3 and transcription activation at E2F responsive promoters during the G1-to-S phase transition (Tyagi, Chabes, Wysocka, & Herr, 2007).

The shared subunits WDR5, RbBP5 and ASH2L have also been implicated in the recruitment of KMT2 complexes through direct interactions with lncRNAs, histone H3 or transcription factors. While RbBP5 directly interacts with histone H3 residues 10-57 (D. H. Kim et al., 2013), ASH2L binds to several transcription factors to recruit KMT2 complexes

to maximize the expression of their responsive genes. These transcription factors include Pax7 (McKinnell et al., 2008), the phosphorylated form of Mef2d (Rampalli et al., 2007), the hematopoietic activator NF-E2p45 (Demers et al., 2007), Ap2delta (Tan et al., 2008), Tbx1 (Stoller et al., 2010), NF-Y (Fossati, Dolfini, Donati, & Mantovani, 2011) and MYC (Ullius et al., 2014). In addition to its ability to bind histone H3 (Couture, Collazo, & Trievel, 2006), WDR5 recruits KMT2 complexes through direct interactions with several lncRNAs. The direct interaction between WDR5 and the lncRNA *HOTTIP*, which is transcribed from the 5' end of the *HOXA* locus, recruits the MLL1 complex to *HOXA* genes for their expression in human fibroblasts (K. C. Wang et al., 2011). Similarly, the direct interaction between WDR5 and the lncRNA *NeST*, which is expressed adjacent to the interferon gamma (*IFN-γ*) gene, results in an increase in H3K4 methylation and production of the IFN-γ protein at the mouse interferon gamma locus (Gomez et al., 2013). Finally, the lncRNA *Mistral* (*Mira*), which is transcribed from an intergenic region located between the *Hoxa6* and *Hoxa7* genes in mouse ESCs, recruits MLL1 to this locus and stimulates gene expression (Bertani, Sauer, Bolotin, & Sauer, 2011).

1.9.2 Regulation of H3K4 Methylation by Other Histone PTMs

Histone modifications usually act in a coordinated manner to control DNA-related events. Correspondingly, other PTMs deposited on histone proteins can positively regulate H3K4 methylation by the KMT2 enzymes. In yeast, H2B ubiquitylation (H2Bub) on residue K123 by the yeast E2/E3 Rad6/Bre1 complex positively regulates H3K4 tri-methylation (Dover et al., 2002). Similarly, deposition of the same modification on the mammalian H2BK34 and H2BK120 residues by the E3 ubiquitin ligases MSL1/2 and RNF20/40 stimulates MLL1 activity and the expression of MLL1 target genes (L. Wu et al., 2013; L. Wu, Zee, Wang, Garcia, & Dou, 2011). Three models have been proposed for this interplay.

H2Bub may 1) induce conformational changes on nucleosomes thus altering chromatin structure, 2) stabilize KMT2 interaction with chromatin or 3) induce conformational changes within the KMT2 complexes or the catalytic SET domains. Another PTM positively regulating the activity of the KMT2 enzymes is histone H4 K16 acetylation (H4K16ac). Deposition of the H4K16ac mark by the MOF acetyltransferase complex is essential for transcription activation of the MLL1 target *HOXA9* (Dou et al., 2005) and promotes H3K4 methylation by other KMT2 complexes (Zhao et al., 2013).

PTMs deposited on histone proteins can also negatively regulate the activity of the KMT2 enzymes. In contrast to H2Bub, ubiquitylation of the histone H2A K119 residue by the E3 ubiquitin ligase RING1b/BMI1 inhibits H3K4 methylation by the MLL1 and MLL3 complexes (Nakagawa et al., 2008; L. Wu et al., 2013). In addition, H3R2 asymmetric dimethylation (H3R2me2a) by the arginine methyltransferase PRMT6 prevents WDR5 recruitment and H3K4me3 by the KMT2 complexes, resulting in down-regulation of *Hox* genes and Myc target genes (Hyllus et al., 2007). These results are further supported by chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-seq) studies showing that H3R2me2a and H3K4me3 occupy mutually exclusive regions in the human genome (Guccione et al., 2007). Lastly, H3K27me3 prevents H3K4 methylation by inhibiting the binding of WDR5 and RbBP5 to histone H3 (D. H. Kim et al., 2013). Correspondingly, the same study further showed that the H3K27 demethylase UTX, a unique subunit of MLL3 and MLL4 complexes, is important for the stable recruitment of ASCOM to chromatin and the expression of several RAR target genes.

Altogether, these examples illustrate the roles of the KMT2 complex subunits in sensing other PTMs deposited on chromatin to control H3K4 methylation.

1.10 Regulation of KMT2 Enzymatic Activity by the WRAD Complex

Aside from their unique subunits, all KMT2 complexes contain a common core complex of four evolutionarily conserved proteins: WDR5, RbBP5, ASH2L and DPY30, collectively known as WRAD. According to Dou et al., WRAD forms a stable structural sub-complex independent of the catalytic subunit and interacts interchangeably with each member of the mammalian KMT2 family (Dou et al., 2006).

The predominant role of WRAD is to associate to the SET domain of each KMT2 enzyme to stimulate their methyltransferase activity. Association of the WRAD complex is essential to maintain high levels of H3K4 methylation, but with different impact on the product specificity of the KMT2 enzymes *in vivo*. While knockdown of WDR5 and RbBP5 in HEK293 cells results in global loss of H3K4 methylation, knockdown of ASH2L and DPY30 results in global loss of H3K4me3 and to a lesser extent H3K4me2 (Dou et al., 2006; Jiang et al., 2011; Steward et al., 2006). These observations are supported by *in vitro* enzymatic assays showing that in the absence of WRAD, MLL1 displays weak mono-methyltransferase activity, but upon incubation with WRAD, its activity is stimulated by 600-fold (A. Patel, Dharmarajan, Vought, & Cosgrove, 2009). Recently, a study comparing the product specificity and regulation of all KMT2 enzymes by WRAD under *in vitro* conditions has shown that while WRAD significantly stimulates MLL1 and MLL4 mono- and di-methyltransferase activities, it only stimulates the mono-methyltransferase activity of MLL2 and MLL3 (Shinsky, Monteith, Viggiano, & Cosgrove, 2015). Under the same conditions, SET1A and SET1B are stimulated by WRAD to catalyze all three degrees of methylation.

Currently, there are two models for the regulation of KMT2 activity by WRAD: 1) allosteric regulation or the one active-site model and 2) the presence of a second active site

within WRAD or the two active-site model. First, structural comparison of the MLL1 SET domain with other SET domains reveals a more open active site that allows mobility of the target lysine side chain and thus cannot facilitate optimal alignment of the cofactor AdoMet and the lysine side chain for methyl transfer, leading the Wilson laboratory to propose that the core complex stimulates MLL1 methyltransferase activity by inducing closure of the active site for better alignment of AdoMet and the H3 substrate to support mono-, di- and trimethylation all within the same active site (Southall, Wong, Odho, Roe, & Wilson, 2009). In line with this model, UV-dependent AdoMet crosslinking experiments show that while all KMT2 enzymes can bind AdoMet in the absence of WRAD, WRAD significantly increases AdoMet binding by the KMT2 enzymes, presumably through a conformational change in their SET domains (Shinsky et al., 2015). In regard to the two active-site model, the Dou laboratory has demonstrated that the ASH2L-RbBP5 heterodimer displays weak methyltransferase activity on histone H3 (Cao et al., 2010). The same study also showed using UV crosslinking experiments that ASH2L directly binds AdoMet, but only in the presence of RbBP5 and MLL1. These findings were further supported by the Cosgrove laboratory showing that WRAD, independent of the MLL1 SET domain, mono-methylates histone H3 (A. Patel, Vought, Dharmarajan, & Cosgrove, 2011). These studies suggest the existence of an additional active site on WRAD that may facilitate enzyme processivity, which is to sequentially perform multiple methylation reactions on the same lysine residue. As structural evidence for the WRAD active site is lacking, the two active-site model is purely speculative at the moment.

Analytical ultracentrifugation studies of the pairwise interactions between purified MLL1 and the individual WRAD subunits have demonstrated that the MLL1 complex assembles in a linear fashion: MLL1-WDR5-RbBP5-ASH2L-DPY30 (MLL1-WRAD) (A.

Patel et al., 2009). This analysis has highlighted the important scaffolding role of WDR5 and RbBP5 in the assembly of the MLL1 complex. Although a high-resolution structure is still lacking, the cryo-electron microscopy (cryo-EM) structure of the MLL1 core complex reveals a Y-shaped architecture with WDR5 and RbBP5 forming the tip of the Y-shaped structure and ASH2L/DPY30 forming the base at the bottom (Takahashi et al., 2011) (Figure 1.3). Structural analysis points to a model wherein the MLL1 SET domain is positioned at the junction of all core subunits. At this position, MLL1 directly interacts with all core subunits to methylate the histone H3 tail using a central channel.

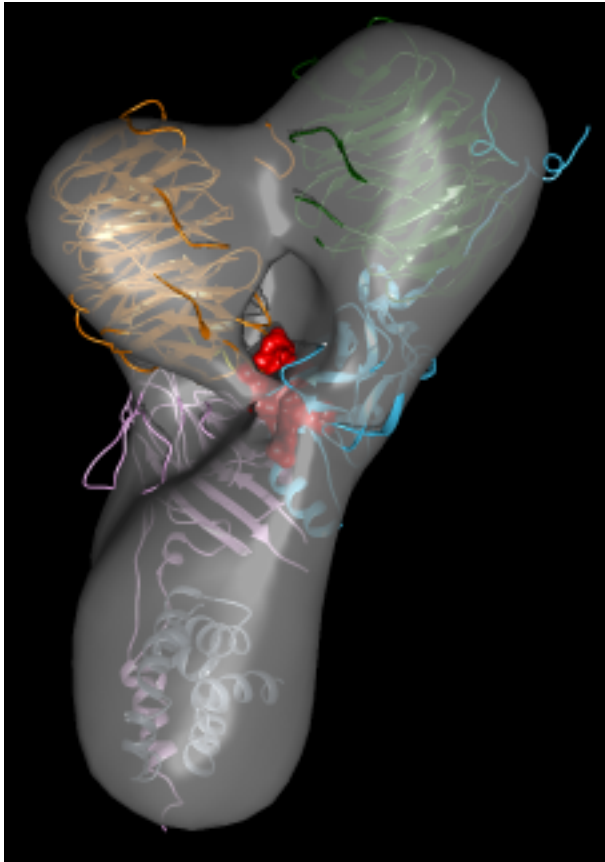


Figure 1.3. Cryo-EM structure of the Y-shaped MLL1 core complex.

MLL1 (cyan) and the H3 substrate (red) are located between the top two lobes of the Y-shape, corresponding to WDR5 (green) and RbBP5 (orange), and the bottom of the Y-shape, corresponding to ASH2L (pink) and DPY30 (grey). This figure was generated by Dr. Georgios Skiniotis using the University of California, San Francisco (UCSF) Chimera program. The modeled crystal structures of the core complex proteins (WDR5/RbBP5 WD40 domain, ASH2L SPRY domain, DPY30 DD domain and MLL1 SET domain) were mapped onto the Y-shaped structure. At the beginning of my PhD, only the crystal structure of WDR5 (RCSB 2H13.pdb) was known.

1.10.1 WDR5

WDR5 is important for vertebrate development of *Xenopus laevis* (Wysocka et al., 2005) and self-renewal of ESCs (Ang et al., 2011). Similar to the role of WDR5 in HEK293 cells (Dou et al., 2006), deletion of *Cps30* in yeast severely compromises the stability of the Set1 protein and consequently global levels of H3K4 methylation (Schneider et al., 2005).

WDR5 folds into a β -propeller domain of seven WD40 repeat domains with each domain composed of four anti-parallel β -sheets (Figure 1.4). Within the KMT2 complexes, WDR5 acts as a scaffold bridging the catalytic subunit to the other regulatory subunits, namely RbBP5 and ASH2L. Two independent studies have shown that WDR5 binds a small region, referred to as the WDR5-interacting (Win) motif, of MLL1 that is N-terminally adjacent to the SET domain (A. Patel, Dharmarajan, & Cosgrove, 2008a; Song & Kingston, 2008); Figures 1.2 and 1.4). The crystal structure of WDR5 in complex with an MLL1 Win peptide reveals that an arginine residue within the Win motif is essential for WDR5 binding whereby its side chain is anchored by two key phenylalanine residues within the WDR5 peptidyl-arginine-binding cleft. Sequence alignment shows that this arginine residue is conserved within the Win motifs of other KMT2 enzymes, suggesting that WDR5-Win interaction is conserved for all KMT2 complexes (A. Patel, Vought, Dharmarajan, & Cosgrove, 2008b). Aside from the Win motifs, WDR5 also uses the same cleft to bind histone H3 via arginine 2 (Couture et al., 2006) and residues 10-31 (D. H. Kim et al., 2013). However, how the WDR5 peptidyl-arginine-binding cleft discriminates between the Win motifs and the two histone H3 regions in mammalian cells is still a matter of debate.

1.10.2 RbBP5

In mouse ESCs, RbBP5, along with DPY30, is important for the induction of early developmental genes during neuronal differentiation and the maintenance of maximal levels

of H3K4me3 (Jiang et al., 2011). Similar to the role of RbBP5 in HEK293 cells (Dou et al., 2006), deletion of *Cps50* in yeast results in global loss of H3K4 methylation (Schneider et al., 2005).

RbBP5 is organized into three distinct domains: a predicted N-terminal β -propeller domain, a short acidic middle hinge region and an unstructured C-terminus (Figure 1.4). In line with an interaction between Cps50 and Cps30 in yeast (Dehe et al., 2006), the role of RbBP5 within the KMT2 complexes is to mediate the interaction between MLL1 and WDR5 (Dou et al., 2006). The N-terminal β -propeller domain contributes to the stimulation of MLL1 methyltransferase activity, indicating a direct interaction with the MLL1 SET domain (Odho, Southall, & Wilson, 2010), and a 50-residue region spanning the hinge region and a part of the C-terminus mediates the interaction between ASH2L and WDR5 (Avdic et al., 2011a) (Chapter 2). A yeast two-hybrid assay has mapped the ASH2L-binding motif (ABM) of RbBP5 to a region of eight residues within the hinge region (Dou et al., 2006) and enzymatic assays have further shown that these residues are important for stimulating MLL1 methyltransferase activity (Cao et al., 2010).

A recent study has demonstrated the role of RbBP5 in the recruitment of KMT2 complexes to specific genes. GST pull down experiments have shown that RbBP5, in addition to WDR5, binds histone H3, via residues 10-57, and that this interaction is strengthened by H3K4me3 (D. H. Kim et al., 2013). In turn, the additional interaction with RbBP5 further stabilizes the KMT2 complexes for efficient H3K4 methylation. This finding may reconcile the model proposed by Ruthenburg et al. (Ruthenburg et al., 2006), whereby RbBP5, instead of WDR5, propagates the H3K4me3 mark to downstream regions of genes as well as to the cognate histone H3 locus on the opposite DNA strand during replication. In

addition, ChIP experiments in HEK293 cells have shown that the occupancy of RbBP5 in ASCOM on several RAR target genes is significantly decreased upon deletion of the H3K27 demethylase UTX, further suggesting that RbBP5 can directly bind chromatin and recruit KMT2 complexes to specific genes.

1.10.3 ASH2L

ASH2L is the mammalian homolog of the *Drosophila* trithorax protein Ash2 initially characterized as an important regulator of wing development (Amoros, Corominas, Deak, & Serras, 2002). In mice, Ash2L is critical for early embryonic development as *Ash2l*-null embryos die during gestation (Stoller et al., 2010). Further studies have shown that Ash2L in mouse ESCs is important for maintaining pluripotency and open chromatin (Wan et al., 2013). Similar to the role of ASH2L in mammalian cells (Dou et al., 2006; Steward et al., 2006), deletion of *Cps60* in yeast results in global loss of H3K4me3 (Schneider et al., 2005).

ASH2L is organized into four domains: a C4 zinc finger domain, a winged-helix (WH) motif, a C-terminal SPLA and RYanodine Receptor (SPRY) domain and a DPY30-binding motif (DBM) (Figure 1.4). While the C4 zinc finger domain and the WH motif (Y. Chen et al., 2011; Sarvan et al., 2011) are important for the recruitment of KMT2 complexes, the SPRY domain is important for stimulating MLL1 methyltransferase activity (Cao et al., 2010). The crystal structure of the ASH2L SPRY domain reveals that several lysine and arginine residues on a positively charged surface are required for RbBP5 interaction and stimulation of MLL1 methyltransferase activity (Y. Chen, Cao, Wan, Dou, & Lei, 2012).

In addition to the recruitment of KMT2 complexes through ASH2L interaction with transcription factors, the ASH2L WH domain binds DNA (Y. Chen et al., 2011; Sarvan et al., 2011). Structural studies have shown that the ASH2L WH domain harbors key structural elements resembling the DNA binding domain of forkhead transcription factors such as

FOXO4. Mutational analysis combined with electrophoretic mobility shift assays and ChIP experiments have confirmed that the ASH2L WH domain is important for ASH2L interaction with DNA and the recruitment of KMT2 enzymes for H3K4me3 and gene expression at the *β -globin* and *Hox* loci (Y. Chen et al., 2011; Sarvan et al., 2011).

1.10.4 DPY30

DPY30 was initially identified in *Caenorhabditis elegans* as a critical component of the dosage compensation machinery and loss of DPY30 results in gender-specific lethality (Hsu & Meyer, 1994). The role of mammalian DPY30 was recently elucidated in ESCs showing that it is important for global H3K4me3, expression of lineage-specific developmental genes and neuronal differentiation of ESCs (Jiang et al., 2011). Similar to the role of DPY30 in mammalian cells, deletion of *Cps25* in yeast results in global loss of H3K4me3 (Schneider et al., 2005).

DPY30 is organized into two domains: an N-terminal domain and a C-terminal dimerization domain (DD) (Figure 1.4). Consistent with an interaction between yeast Cps25 and Cps60 (Dehe et al., 2006), the C-terminus of DPY30, which folds into a four-helix-bundle, interacts with the C-terminus of the ASH2L SPRY domain (Y. Chen et al., 2012). Although the DPY30-ASH2L interaction is not required for MLL1 complex stability (Dou et al., 2006), DPY30 is required for the optimal stimulation of KMT2 methyltransferase activity. Addition of DPY30 to the WRA subunits increases the stimulation of MLL1 activity by two-fold (A. Patel et al., 2009) and increases H3K4me3 by the MLL2 complex (Jiang et al., 2011). The crystal structure of DPY30 in complex with ASH2L was recently elucidated, showing that a key leucine residue within DPY30 DD is essential for ASH2L binding (Tremblay et al., 2014). Importantly, mutation of this conserved hydrophobic residue in the budding yeast results in global loss of H3K4me3, suggesting that the interaction between

ASH2L and DPY30 is critical for the allosteric regulation of the KMT2 enzymes. However, the structural basis underlying the stimulatory role of DPY30 on the kinetics of the KMT2 family of methyltransferases is still unresolved.

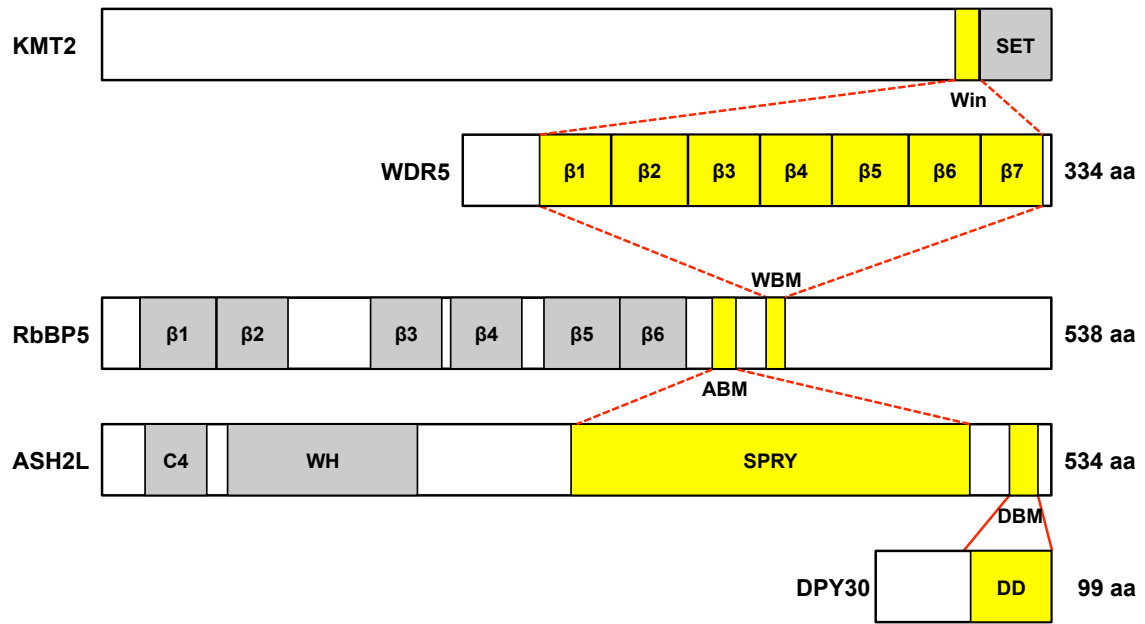


Figure 1.4. Domain architecture of the core complex subunits WRAD.

Domains that are involved in KMT2 complex assembly are shown in yellow and connected by red lines. Dashed lines represent the interactions that were characterized during my PhD (KMT2 Win-WDR5, WDR5-RbBP5 WBM, RbBP5 ABM-ASH2L SPRY).

WBM: WDR5-binding motif; ABM: ASH2L-binding motif; DBM: DPY30-binding motif

1.11 Rationale and Hypothesis

Several studies have shown that the core complex subunits WRAD stimulate the H3K4 methyltransferase activity of all KMT2 enzymes and are important for global H3K4 methylation. In addition, several studies have begun to dissect the role of the individual domains of each core complex subunit in regulating the methyltransferase activity of the KMT2 enzymes. However, a detailed structural and functional characterization of the interactions underlying KMT2-WRAD assembly and how these interactions regulate KMT2 enzymatic activity is not well investigated.

I hypothesize that the interactions between the subunits of the KMT2 core complex are important for complex assembly and subsequent stimulation of KMT2 enzymatic activity. To address this hypothesis, I have used structural biology combined with biochemical studies to determine the interactions underlying KMT2-WRAD complex assembly and activity.

The following three manuscripts focus on the structural and functional characterization of the pairwise interactions between 1) WDR5-RbBP5 2) KMT2-WDR5 and 3) RbBP5-ASH2L (Figure 1.4).

CHAPTER 2. STRUCTURAL AND BIOCHEMICAL INSIGHTS INTO MLL1 CORE COMPLEX ASSEMBLY

The following manuscript is published in *Structure*.

Avdic, V., Zhang, P., Lanouette, S., Groulx, A., Tremblay, V., Brunzelle, J., & Couture, J. F. (2011a). Structural and biochemical insights into MLL1 core complex assembly. *Structure*, 19(1), 101-108. doi: 10.1016/j.str.2010.09.022

I purified WDR5 and a library of RbBP5 peptides, used isothermal titration calorimetry (ITC) to determine their binding affinities and mapped the minimal domain of the RbBP5 C-terminus required for WDR5 binding (RbBP5 371-380) (Figure 2.2D and E). Another student (Adam) then used the RbBP5₃₇₁₋₃₈₀ peptide to co-crystallize with WDR5 and an MLL1_{WIN} peptide. Based on the MLL1-WDR5-RbBP5 structure, I purified several WDR5 and RbBP5 mutants and used ITC and *in vitro* methyltransferase assays to probe their role in RbBP5 binding (Figure 2.4B) and the role of the WDR5-RbBP5 interaction in stimulating MLL1 activity (Figure 2.4C), respectively. Lastly, I contributed to the editing of the manuscript.

Structural and biochemical insights into MLL1 core complex assembly and regulation.

Vanja Avdic^{1*}, Pamela Zhang^{1*}, Sylvain Lanouette¹, Adam Groulx¹, Veronique Tremblay¹, Joseph Brunzelle³, and Jean-Francois Couture¹

1) Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, Ontario, K1H8M5

2) Feinberg School of Medicine, Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, Chicago, Illinois 60611, USA

Running head: Dissection of the MLL core complex assembly

* These authors equally contributed to this article.

ABSTRACT

Methylation of histone H3 on Lys-4 is predominantly catalyzed by a family of methyltransferases whose enzymatic activity heavily depends on their interaction with a three-subunit complex composed of WDR5, RbBP5 and ASH2L. Here we report that a segment of 50 residues of RbBP5 bridges the ASH2L C-terminal domain to WDR5. The crystal structure of WDR5 in ternary complex with RbBP5 and MLL1 reveals that both proteins bind peptide-binding clefts located on opposite sides of WDR5's β -propeller domain. RbBP5 engages in several hydrogen bonds and van der Waals contacts within a V-shaped cleft formed by the junction of two blades on WDR5. Mutational analyses of both the WDR5 V-shaped cleft and RbBP5 residues reveal that the interactions between RbBP5 and WDR5 are important for the allosteric regulation of MLL1 methyltransferase activity. Overall, this study provides the structural basis underlying the formation of the WDR5-RbBP5 sub-complex and further highlights the important role of WDR5 in scaffolding the MLL1 core complex.

INTRODUCTION

Histone lysine methylation is a prominent post-translational modification linked to diverse biological pathways including gene transcription (Martin & Zhang, 2005), maintenance of heterochromatin (Lachner, O'Carroll, Rea, Mechtler, & Jenuwein, 2001; Schotta et al., 2002) and double stranded DNA break repair (Greeson, Sengupta, Arida, Jenuwein, & Sanders, 2008). Notably, due to its association with actively transcribed genes, methylation of Lys-4 on histone H3 (H3K4) has been heavily documented in recent years. Genome wide localization studies reveal that H3K4 tri-methylation marks the promoter regions of actively transcribed genes while H3K4 di-methylation has been recently linked to genes poised for transcription (Bernstein et al., 2005; Schneider, Dover, Johnston, & Shilatifard, 2004).

The SET1 family of enzymes, which includes MLL1 to 5 and SET1A/B, carries out the body of H3K4 di/tri-methylation *in vivo* (P. Wang et al., 2009). Studies on two archetypal members of this family, ScSET1 and HsMLL1, have provided several insights on the mechanistic determinants underlying the regulation of these enzymes (Cosgrove & Patel, 2010; Tenney & Shilatifard, 2005). Of interest, immunoprecipitation experiments have shown that, in contrast with other SET domain enzymes, members of the SET1 family are always found in multi-subunit protein complexes (Dou et al., 2005; Nakamura et al., 2002; Schneider et al., 2005; Yokoyama et al., 2004). In line with these findings, all but one member of the MLL proteins, MLL5, have been shown to interact with a common three-subunit complex composed of ASH2L, WDR5 and RbBP5. The essential role of each subunit in regulating MLL1 methyltransferase activity has been demonstrated in HELA and HEK293 cells as RNA-mediated interference targeting either ASH2L, RbBP5 or WDR5 leads to a global loss of H3K4 tri-methylation (Dou et al., 2006; Steward et al., 2006;

Wysocka et al., 2005). These results have been further confirmed by recent functional and biochemical studies showing that WDR5, RbBP5 and ASH2L are all essential in the assembly of the core complex and the allosteric regulation of MLL1 methyltransferase activity (A. Patel et al., 2008a; A. Patel et al., 2009; A. Patel et al., 2008b; Southall et al., 2009). Nevertheless, the mechanisms underlying the formation of the WDR5-RbBP5-ASH2L complex and its binding with the MLL1 SET domain have remained elusive. As both ASH2L and RbBP5 harbor several domains, it is yet unknown if one specific element or a cooperation between multiple domains is necessary for the assembly of the complex.

Here we set out to identify the minimal structural determinants underlying the formation of the WDR5-RbBP5-ASH2L complex. Dissection of the MLL1 core complex subunits reveals that a segment of 50 amino acids of RbBP5 supports the formation of the core complex in bridging WDR5 to the ASH2L C-terminal domain. The crystal structure of WDR5 in ternary complex with MLL1 and RbBP5 reveals that both proteins bind WDR5 using peptide-binding clefts located on opposite sides of WDR5's propeller domain. In addition, the tertiary complex and the accompanying mutational analyses illustrate the important role of the WDR5 V-shaped cleft in binding RbBP5 and allosterically regulating MLL1 methyltransferase activity.

RESULTS AND DISCUSSION

The C-terminal Domain of ASH2L Binds the Hinge Region of RbBP5

The core complex controls the allosteric regulation of MLL1, a representative member of the SET1 family of methyltransferases (Dou et al., 2006). Initial work performed on the core complex revealed that ASH2L directly interacts with RbBP5 (A. Patel et al., 2009). The latter is composed of an N-terminal seven-bladed β -propeller domain, a short

hinge region and a C-terminal domain of unknown function (Figure 2.1A). To define the role of the RbBP5 domains in the assembly of the core complex, RbBP5 fragments corresponding to residues 1-538 (Full-Length: FL), 1-366 (β -Propeller + Hinge: PH), 330-538 (Hinge + C-terminus: HC) and 330-366 (Hinge: H) (Figure 2.1A) were expressed as GST-fusion proteins and used as baits for *in vitro* pull-down experiments. Bound ASH2L_{FL} was eluted, resolved by SDS-PAGE and visualized by coomassie staining and western blot using an ASH2L specific antibody (Rampalli et al., 2007). As shown in Figure 2.2A, no ASH2L enrichment was detected for GST alone but consistent binding was detected for the assays performed with all the GST-RbBP5 fragments, including the peptide corresponding to the 40-residue hinge region of RbBP5.

ASH2L comprises a C4 zinc-finger motif and a SP1a and RYanodine Receptor (SPRY) domain located on its N-(ASH2L_N) and C-terminus (ASH2L_C), respectively (Figure 2.1B). To identify the region of ASH2L that interacts with RbBP5, ASH2L_N and ASH2L_C were homogeneously purified and added to pre-bound GST-RbBP5 fragments. As shown in Figure 2.2B, consistent binding was observed for ASH2L_C with all the GST-RbBP5 fragments while no binding was detected for ASH2L_N. These results reveal that a scaffolding domain, located between Ser-317 and the ASH2L C-terminus, is involved in the binding of the RbBP5 hinge region and suggest that the ASH2L SPRY domain contribute to the formation of the core complex. SPRY domains are scaffolding modules that employ a set of solvent-exposed loops to mediate protein-protein interactions (Filippakopoulos et al., 2010; Woo, Suh, Park, & Oh, 2006). Amino acid alignment of ASH2L and GUSTAVUS SPRY domains reveals that the ASH2L SPRY domain harbors a unique and evolutionarily conserved insertion of 19 amino acids (data not shown). In addition, homology modeling of the ASH2L SPRY domain shows that it harbors a similar set of loops with, however, a

noticeable increase in surface basic residues. More work is required to determine whether this insertion, the punctate charge differences on the loops or both are required for binding RbBP5 and/or stimulating MLL1 methyltransferase activity. Collectively, our results suggest that the SPRY-containing ASH2L_C domain binds the hinge region of RbBP5. In addition, as ASH2L does not bind WDR5 directly (Dou et al., 2006; A. Patel et al., 2009), our data also suggest that ASH2L_N is accessory to the assembly of the core complex and likely plays a role in linking WDR5-RbBP5-ASH2L_C-MLL1 to other regulatory subunits of MLL1.

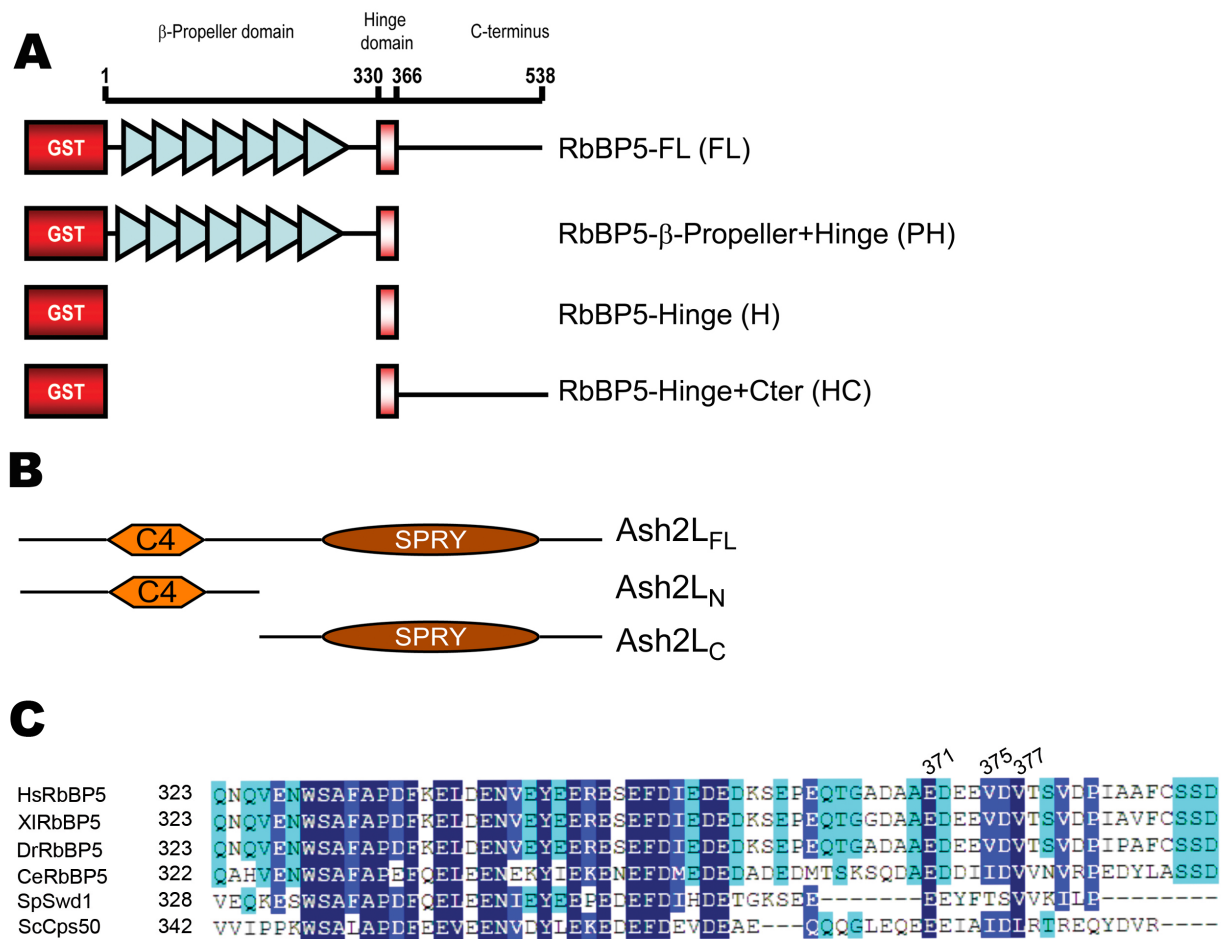


Figure 2.1. ASH2L and RbBP5 harbor several evolutionarily conserved domains.

Schematic representation of RbBP5 (A) and ASH2L (B) constructs used in the GST pull-down assays.

C) A protein sequence alignment of *Homo sapiens* (Hs), *Xenopus laevis* (Xl), *Danio rerio* (Dr), *Drosophila melanogaster* (Dm), *Caenorhabditis elegans* (Ce), *Schizosaccharomyces pombe* (Sp) and *Saccharomyces cerevisiae* (Sc) RbBP5 zoomed on a region neighboring the hinge region. Positions with 100-80%, 80-60% and less than 60% of amino acid conservation are represented in dark, medium and pale blue, respectively.

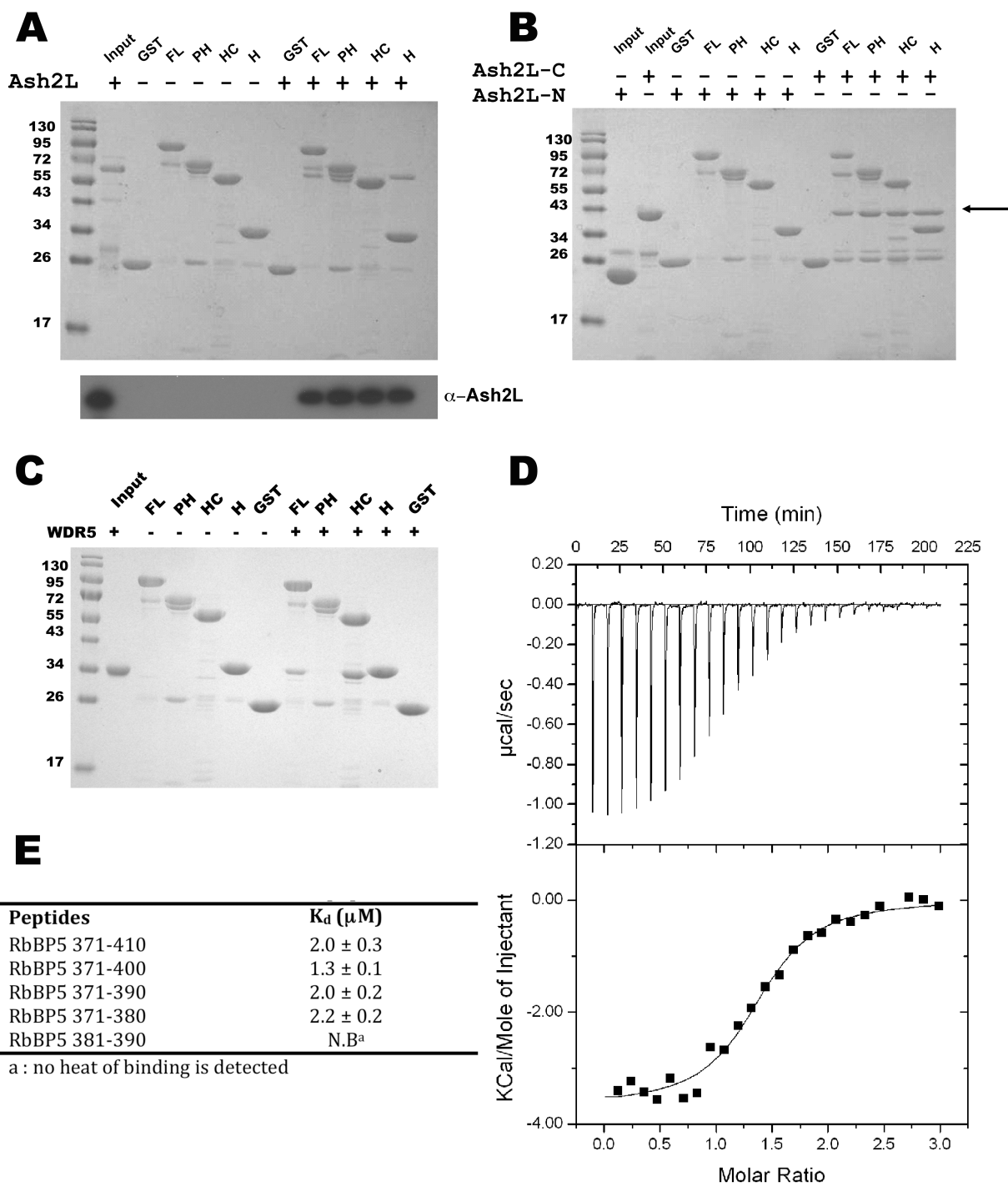


Figure 2.2. Dissection of the core complex.

A) ASH2L interacts with the GST-RbBP5-hinge fragment. GST-tagged RbBP5 domains were incubated with and without equal amounts of ASH2L_{FL}. The lower panel shows a western blot of the ASH2L protein used as input or recovered after pull-down by the RbBP5 fragments using a specific ASH2L antibody.

B) ASH2L_C domain binds the hinge region of RbBP5. Equal amount of ASH2L N- and C-terminal domains were co-incubated with the GST-RbBP5 fragments. The black arrow highlights the domain of ASH2L that has been retained by the RbBP5 fragments.

C) WDR5 interacts with the GST-RbBP5-HC fragment. GST-tagged RbBP5 domains were incubated with and without equal amounts of WDR5. Bound proteins were resolved on 15% SDS-PAGE and coomassie stained. Loading controls for GST-RbBP5 fragments are shown on the left hand side of the gel. Abbreviations are as follow: FL: GST-RbBP5-Full length, PH: GST-RbBP5- β -propeller+hinge, HC: GST-RbBP5-Hinge+C-terminus, H: GST-RbBP5-Hinge.

D) Representative isothermal titration calorimetry data for the binding of RbBP5₃₇₁₋₃₈₀ to WDR5. The upper and lower panels show the heat of binding for peptide binding and integrated data, respectively.

E) Equilibrium dissociation constants for the titration of WDR5 with RbBP5 peptides. Indicated are the RbBP5 peptides with their corresponding K_D values for WDR5.

RbBP5 Binds WDR5 Using a Segment Neighboring its Hinge Region

After identifying the minimal domain of ASH2L sufficient to interact with RbBP5, we sought to map the minimal structural domains mediating the interaction between RbBP5 and WDR5. We performed GST pull-down experiments using the RbBP5 fragments listed in Figure 2.1A with purified WDR5 and subsequently determined that WDR5 binds to full-length RbBP5 and the RbBP5-HC fragment (Figure 2.2C), indicating that WDR5 binds a region located between the RbBP5 hinge region and its C-terminal domain. Our observations also suggest that ASH2L and WDR5 bind distinct regions of RbBP5. To identify the residues responsible for the interaction, a multiple sequence alignment was performed using several eukaryotic homologs of RbBP5 (Figure 2.1C). Given the functional overlap between WDR5-RbBP5 and Cps35-Cps50 and the analogous requirement for their interaction within the MLL complex and COMPASS (Trievel & Shilatifard, 2009), respectively, it is expected that the region of interaction would show a significant degree of conservation. Close inspection of the alignment reveals that the only conserved domain outside of the β -propeller domain and the hinge region of RbBP5 is located between residues 371-410 (Figure 2.1C).

To confirm our observations and to further define the elements of RbBP5 that contribute to the binding of WDR5, we determined the equilibrium dissociation constants (K_D) of a small library of peptides spanning this region of RbBP5 (Figure 2.2E). Using isothermal titration calorimetry (ITC) (Figure 2.2D), we found that WDR5 binds with similar affinities ($\sim 2 \mu\text{M}$), the RbBP5 peptides corresponding to residues 371-410, 371-400, 371-390 and 371-380 while no binding could be detected with a 10mer peptide corresponding to residues 381-390 of RbBP5 (Figure 2.2E). Given that initial binding studies established that full-length RbBP5 binds to WDR5 with a K_D of $2.4 \mu\text{M}$ (A. Patel et al., 2009), our data

suggest that RbBP5₃₇₁₋₃₈₀ comprises the minimal structural determinants underlying the interaction between WDR5 and RbBP5.

Structure of the WDR5/RbBP5/MLL1 Ternary Complex

To gain insights into the mechanisms underlying the recognition of RbBP5 by WDR5, we have determined the 2.35Å-resolution crystal structure of WDR5 bound to RbBP5₃₇₁₋₃₈₀ and the MLL1 WIN motif (MLL1_{WIN}). Overall, the structure of WDR5/RbBP5₃₇₁₋₃₈₀/MLL1_{WIN} is similar to apo-WDR5 with an overall root-mean-square (r.m.s.) of 0.58Å for all protein atoms. Close inspection of the ternary complex reveals that MLL1_{WIN} is maintained in a 3_{10} helix conformation with its C-terminus exiting the WDR5 peptidyl-arginine-binding cleft toward blade 6 (Figure 2.3A). As previously described, MLL1_{WIN} is pinioned in the peptidyl-arginine-binding cavity by several hydrogen bonds and hydrophobic contacts (A. Patel et al., 2008a; Song & Kingston, 2008). Upon closer inspection of the WDR5 surface, we found a continuous electron density corresponding to residues 374-380 of RbBP5. The first three residues of the RbBP5 peptide were not modeled, as no discernable electronic density was observed. Analysis of the RbBP5₃₇₁₋₃₈₀ binding mode reveals that the peptide adopts an elongated conformation with a small kink centered on V377 (the single letter refers therein to RbBP5 residues). The RbBP5 peptide is found in proximity of a V-shaped cleft formed by the junction of blades 5 and 6 of WDR5 (Figure 2.3A) and related to the WDR5 peptidyl-arginine-binding cleft, is maintained on the opposite side of the β -propeller.

Close inspection of the V-shaped cleft reveals that RbBP5₃₇₁₋₃₈₀ is maintained by several hydrogen bonds and van der Waals contacts. Notably, E374 engages in several van der Waals contacts with the Gln289 side chain while the V375 side chain participates in

hydrophobic interactions within a small hydrophobic pocket composed of Leu249, Tyr228, Leu240 and Lys250. In addition, the carbonyl oxygen and backbone amide of V375 engage in two hydrogen bonds with the Gln289 amide side chain (Figure 2.3C). Following V375 in RbBP5, D376 interacts with WDR5 through a hydrogen bond between its backbone carbonyl and the Asn225 amide side chain. In addition, D376 also engages in several intra-molecular hydrogen bonds between its carboxylate group and the T378 backbone amide, S379 hydroxyl group and backbone amide (Figure 2.3C). This network of hydrogen bonds likely stabilizes RbBP5's kink conformation and orients the V377 side chain in a hydrophobic pocket composed of Phe266, Tyr228, Ser223 and Leu288 of WDR5. Moreover, the residues T378 and S379 participate in water-mediated and direct hydrogen bonds with WDR5's Pro224 carbonyl group and Asn225 amide side chain, respectively. Overall, the crystal structure of the WDR5/RbBP5₃₇₁₋₃₈₀/MLL1_{WIN} ternary complex and sequence homology of RbBP5 homologs (Figure 2.1C) suggest that an aspartate residue bordered by a pair of hydrophobic moieties are key in conferring binding activity of RbBP5 within WDR5's V-shaped cleft.

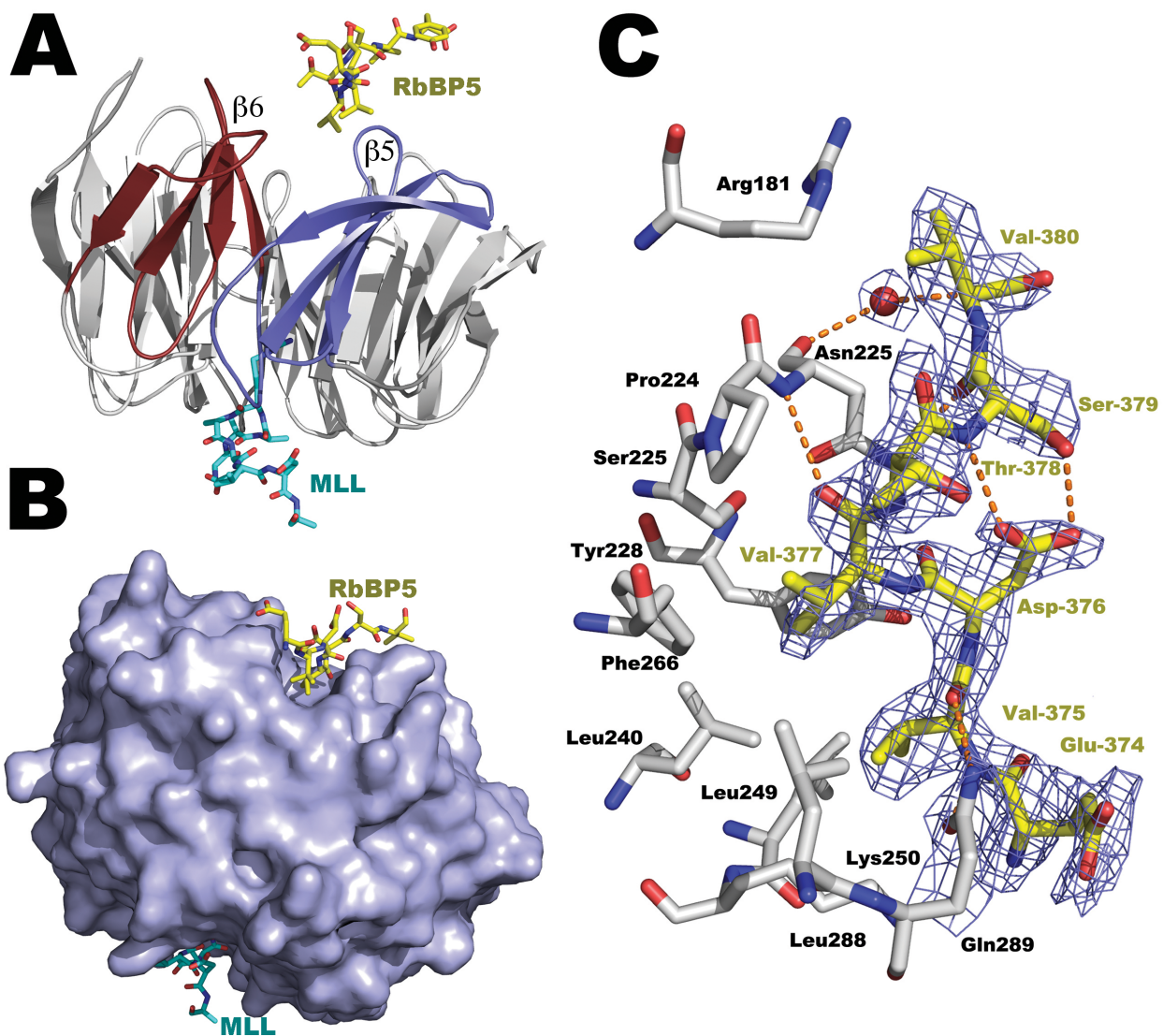


Figure 2.3. WDR5 binds RbBP5 and MLL1 using two independent binding clefts.

A) Overall structure of WDR5 in complex with MLL1_{WIN} and RbBP5₃₇₁₋₃₈₀ peptides. Blades 5 and 6 are highlighted in blue and red, respectively. MLL1_{WIN} and RbBP5₃₇₁₋₃₈₀ carbon atoms are colored in cyan and yellow, respectively.

B) Surface representation of WDR5's peptide-binding clefts.

C) Zoomed view of the simulated annealing $F_o - F_c$ omit map (blue) contoured at 2σ in which WDR5 and RbBP5 carbon atoms are rendered in grey and yellow, respectively. Hydrogen bonds are displayed as dashed orange lines.

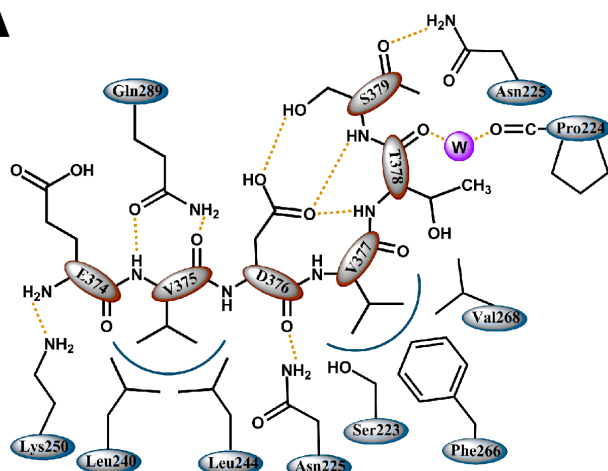
WDR5 V-Shaped Cleft and RbBP5 VDV Motif are Both Important for the Allosteric Regulation of MLL1 Methyltransferase Activity

To further probe the interactions between RbBP5 and WDR5, we mutated Asn225, Leu240 and Gln289 and measured their effects on the binding of RbBP5 and the subsequent stimulation of MLL1 methyltransferase activity (Figure 2.4B and C). Mutation of Asn225 to alanine abolished RbBP5 binding and resulted in a 2-fold decrease in the stimulation of MLL1 methyltransferase activity (Figure 2.4B and C). Similar to Asn225, substitution of Leu240 to a lysine residue, which likely disrupts the formation of the hydrophobic cleft, impaired RbBP5 binding and weakened the allosteric regulation of MLL1 (Figure 2.4B and C). Finally, mutation of Gln289 to alanine decreased the affinity of WDR5 for RbBP5 by 20-fold approximately and correlatively led to a 3-fold decrease in the methylation of histone H3.

RbBP5's amino acid side chains are recognized by WDR5 through several hydrophobic contacts and hydrogen bonds. To identify the RbBP5 residues that confer its binding to WDR5, we mutated E374, V375, D376 and V377 and determined their effects on the methyltransferase activity of MLL1 (Figure 2.4C). Consistent with the loss of methyltransferase activity for the WDR5 mutants, all of the RbBP5 mutants failed to stimulate, to the same extent of the complex reconstituted with the wild-type proteins, MLL1 methyltransferase activity. Notably, mutation of E374 to an alanine severely impaired the stimulation of MLL1 methyltransferase activity by the core complex subunits. Similarly, replacement of V375 and V377 by glutamic acid residues weakened the allosteric regulation of MLL1 by 9- and 4-fold, respectively while substitution of D376 to an alanine residue led to a 4-fold decrease in the methylation of histone H3. Given that D376 engages in several intra-molecular hydrogen bonds, our results support a model in which the D376 carboxylate

group is required to maintain RbBP5's kinked conformation in WDR5's V-shaped cleft. Collectively, our results suggest that both the RbBP5 VDV motif and the WDR5 V-shaped cleft are important in stimulating MLL1 methyltransferase activity.

Similar to HsASH2L and Cps60 SPRY domains, the central region of RbBP5₃₇₁₋₃₈₀, including the triad V375-D376-V377, is conserved in the RbBP5 yeast homolog Cps50. Inspection of Cps50 reveals a similar triad composed of I291-D292-L293 (Figure 2.1C), suggesting that the interaction of this pair of hydrophobic residues onto Cps35's β -propeller domain is important for the binding of Cps50 and complements previous findings showing that the heterodimer Cps35-Cps50 is crucial for COMPASS assembly (Dehe et al., 2006; Halbach et al., 2009). Taken together, our structural studies and binding assays indicate that the RbBP5₃₃₀₋₃₈₀ region is of importance in the formation of the core complex, and point to a shared mechanism between the human SET1 family of methyltransferases and its yeast homolog COMPASS.

A**B**

Equilibrium dissociation constants between WDR5 mutants and RbBP5

Mutants	K _d (μM)
WDR5 WT	2.2 ± 0.2
WDR5 N225A	N.B. ^a
WDR5 L240K	N.B. ^a
WDR5 Q289A	44.64 ± 8.45

a : no heat of binding is detected

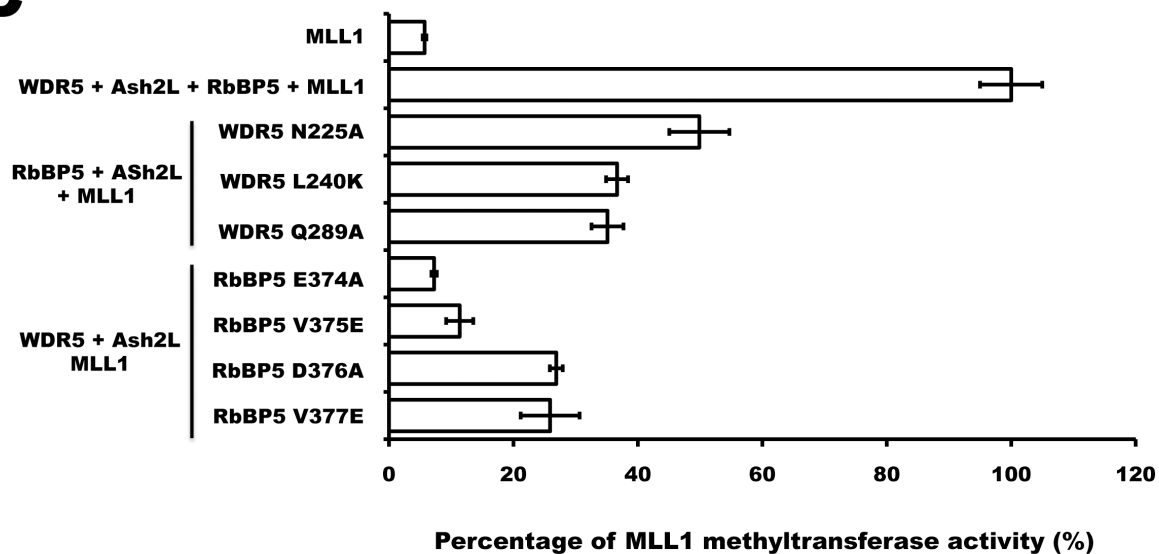
C

Figure 2.4. WDR5 V-shaped cleft and RbBP5 VDV motif are important for the stimulation of MLL1 methyltransferase activity.

A) Schematic representation of WDR5-RbBP5₃₇₁₋₃₈₀ interactions. WDR5 and RbBP5₃₇₁₋₃₈₀ residues are represented with blue and red ovals, respectively. Hydrogen bonds and hydrophobic contacts are highlighted as orange dashed lines and blue half-ovals, respectively. The last two residues of the peptide were not modeled as they engage in hydrogen bond and hydrophobic contacts with a symmetry related molecule.

B) WDR5 V-shaped cleft is important in the binding of RbBP5. Table showing the equilibrium dissociation constants between WDR5 mutants and RbBP5.

C) The interactions between RbBP5 and WDR5 are important for stimulating MLL1 methyltransferase activity. Radiometric methyltransferase assays performed with the MLL1 SET domain alone or reconstituted with the indicated wild-type or mutant proteins. Average values of methyltransferase assays performed in triplicate are shown; error bars represent SD. Each experiment was performed twice with two different batches of recombinant proteins.

Insights into the Organization of a Catalytically Competent WDR5-RbBP5-ASH2L-MLL1 Complex

Protein lysine methyltransferases are often found in multi-protein complexes *in vivo*. However, the SET1 family of methyltransferases is, along with Ezh1/2, one of the few that requires such associations to achieve di- and/or tri-methyltransferase activity. In COMPASS, binding of Cps60 to SET1 is required for H3K4 tri-methylation in yeast (Steward et al., 2006; Takahashi et al., 2009) while ASH2L is essential for stimulating MLL1 methyltransferase activity in mammals (Dou et al., 2006; A. Patel et al., 2008b; Southall et al., 2009). To explain the role of these subunits, it has been proposed that the association of the WDR5-RbBP5-ASH2L complex with the MLL1 SET domain reorganizes its iSET region and contributes to the formation of the histone H3-binding cleft (Southall et al., 2009). While mutations of residues found in the iSET region of SET8 (Couture et al., 2005) and DIM-5 (X. Zhang et al., 2003) are deleterious for their enzymatic activity, similar mutations in the MLL1 iSET region do not impair, to the same extent, its methyltransferase activity when bound to its core complex subunits (Southall et al., 2009). The WDR5/RbBP5₃₇₁₋₃₈₀/MLL1_{WIN} ternary complex and pull-down studies presented herein may suggest a mechanistic basis for these observations. The anti-parallel orientation of RbBP5₃₇₁₋₃₈₀ and MLL1_{WIN} peptides would orient the members of the core complex in a position favorable to interact with the MLL1 SET domain and thus play a role analogous to SET8 or DIM-5 iSET regions.

Biological Functions of WDR5's V-Shaped Cleft

The family of β -propeller proteins plays major functions in several biological pathways through the assembly of multi-subunit protein complexes. Interestingly, recent

studies have shown that WDR5 plays a key role in the assembly of complexes unrelated to MLL1. WDR5 associates with the virus-induced VISA-associated complex (Y. Y. Wang et al., 2010). In human, WDR5 is also found in the ATAC histone acetyltransferase complex (Y. L. Wang, Faiola, Xu, Pan, & Martinez, 2008) and the ATP-dependent chromatin remodeling complex comprising CHD8 (Dou et al., 2005; Thompson, Tremblay, Lin, & Bochar, 2008). Additional studies will be required to establish a possible role of WDR5's V-shaped cleft in the assembly of these complexes but the WDR5/RbBP5₃₇₁₋₃₈₀/MLL1_{WIN} ternary complex illustrates WDR5's ability in binding proteins with two distinct binding clefts.

Overall, our results indicate that only specific elements of a limited region of RbBP5 and the SPRY-containing ASH2L_C domain are necessary for the formation of the MLL1 core complex. Given the functional overlap between WDR5-RbBP5 and Cps35-Cps50 complexes and the role of WDR5's V-shaped cleft in the regulation of histone H3 Lys-4 methylation, our study also points to a common mechanism of allosteric regulation between COMPASS and MLL1.

EXPERIMENTAL PROCEDURES

Peptides

Peptides were purchased either from New England Peptides or Peptide 2.0. Peptides were ordered with 95-99% purity with an additional tyrosine on their C-terminus for UV quantification. Peptides were suspended in water at 100 mM concentration and stored at -20°C.

Protein Expression and Purification

Fragments corresponding to the full-length (ASH2L_{FL}), N-(residues 96-281 (referred therein as ASH2L_N)) and C-terminal (residues 317-628 (referred therein as ASH2L_C)) domains of ASH2L were PCR amplified using clone #3921999 (OpenBiosystem). PCR fragments of ASH2L_{FL} and ASH2L_N were cloned in a modified version of pET3d (Couture, Collazo, Ortiz-Tello, Brunzelle, & Trievel, 2007) while ASH2L_C was cloned in frame with a His-SUMO tag. A DNA construct equivalent to full-length RbBP5 was PCR amplified and cloned in pHis2 (Sheffield, Garrard, & Derewenda, 1999). Proteins were overexpressed with 0.1 mM IPTG in Rosetta cells (Novagen) for 16 hours at 18°C. Cells were harvested in 50 mM sodium phosphate (NaPi) pH 7.0, 500 mM sodium chloride (NaCl) and 5 mM β-mercaptoethanol (BME), lysed by sonication, clarified by centrifugation and purified by Talon Co²⁺ affinity chromatography. TEV-cleaved ASH2L_{FL} and ASH2L_N were further purified by metal affinity and size exclusion chromatographies (Superdex 200). ULP1-cleaved ASH2L_C was purified by Q-sepharose ion exchange and size exclusion chromatographies (Superdex 200). GST-RbBP5 constructs were expressed in BL-21 using the same experimental conditions as for ASH2L_N. Following GST-RbBP5 overexpression, cells were centrifuged and harvested in phosphate buffer saline (PBS). The catalytic domain of MLL1 was produced as previously described (A. Patel et al., 2008b). WDR5 and RbBP5 mutants were prepared as for the wild-type proteins.

GST Pull-Down Assays

GST and GST-fusion RbBP5 proteins were applied onto glutathione-sepharose beads for one hour and washed extensively with PBS. Binding assays were performed in PBS supplemented with 0.1% Triton-X 100 (PBST) for two hours with 10 mg of recombinant

ASH2L or WDR5 proteins. Binding reactions were extensively washed with PBST and bound proteins were eluted with 50 mM Tris-HCl pH 8.0 and 20 mM reduced glutathione for one hour at 4°C. Each sample was loaded on a 12% polyacrylamide gel and coomassie stained. ASH2L binding was also immunodetected using an ASH2L antibody (kindly provided by Dr. Jeffrey Dilworth) (Rampalli et al., 2007).

***In Vitro* Methyltransferase Assays**

Methyltransferase assays were conducted using 5 μ M of recombinantly purified MLL1 SET domain (3745-3969) with equimolar amounts of ASH2L, RbBP5 and WDR5. Reactions were performed in 50 mM Tris-HCl pH 8.5, 200 mM NaCl, 3 mM dithiothreitol, 5 mM magnesium chloride, 5% glycerol and 250 μ M of peptide (residue 1-20 of histone H3). Each assay was initiated by the addition of 1 μ Ci of radiolabeled AdoMet and incubated for two hours at 22°C. Reactions were stopped by spotting the reactions onto Whatman P-81 filter papers and free AdoMet was removed by washing the filter papers in 250 mL of 50 mM sodium bicarbonate pH 9.0. Activity was quantified by liquid scintillation counts.

Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) experiments were performed using a VP-ITC calorimeter (MicroCal). RbBP5 peptides (0.65 mM) were injected into a solution of WDR5 (0.06 mM) as previously described (Couture et al., 2006). Experiments were performed at 19°C in 20 mM NaPi pH 7.0, 100 mM NaCl, and 5 mM BME. The titration data were analyzed using Origin 7.0 (OriginLab Corp.) and showed that the RbBP5 peptides bind WDR5 with binding stoichiometries (N-values) between 0.8-1.1. The K_D , the error and the

N-value were determined by fitting the integrated data to a standard single-site binding model.

Crystallization, Data Collection and Structure Determination

Purified WDR5 was mixed in equal molar ratio with MLL1_{WIN} (A. Patel et al., 2008a) and RbBP5₃₇₁₋₃₈₀ and incubated on ice for 30 minutes. Diffraction quality crystals were grown in 0.2 M ammonium citrate (buffered with hydrochloric acid/sodium hydroxide to pH 7.0) and 18% PEG3350. Crystals were sequentially soaked in the mother liquor supplemented with 10% and 20% glycerol, harvested and flash frozen in liquid nitrogen. A full data set was collected on a HomeSource MicroMax-007 (Rigaku) equipped with a Raxis IV image plate detector. Diffraction data were indexed and scaled using d*TREK[®] (Pflugrath, 1999). A molecular replacement solution was found using MOLREP (A. Vagin & Teplyakov, 2000) and apo-WDR5 as a starting model (Couture et al., 2006). The structure was further refined using iterative cycles of energy minimization and modeling using Refmac (A. A. Vagin et al., 2004) and Coot (Emsley & Cowtan, 2004), respectively. Quality of the structure was assessed using Molprobit (V. B. Chen et al., 2010) (Table 2.1).

Table 2.1: Data Collection and Refinement Statistics for WDR5 Structure in Complex with RbBP5 and MLL1

Data collection	
Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	53.9, 81.4, 81.8
Resolution	32.6 – 2.35 (2.43 – 2.35)!
R _{meas}	8.9 (44.2)
<i>I</i> / σI	8.3 (2.1)
Completeness (%)	99.2 (98.9)
Redundancy	3.3 (3.3)
Refinement	
Resolution (Å)	32.6 – 2.35
No. reflections	15413
R _{work} / R _{free}	19.3 / 24.9
No. Atoms	
Protein	2346
Peptides	134
Water	244
B-factors (Å ²)	
Protein	41.1
Ligands	51.1
Water	50.4
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.3
Molprobability scores	1.68
Ramachandran favored (%)	96.2
Ramachandran allowed (%)	3.8

! Highest resolution shell is shown in parentheses.

REFERENCES

- Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., *et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* *120*, 169-181.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* *66*, 12-21.
- Cosgrove, M.S., and Patel, A. (2010). Mixed lineage leukemia: a structure-function perspective of the MLL1 protein. *FEBS J* *277*, 1832-1842.
- Couture, J.F., Collazo, E., Brunzelle, J.S., and Trievel, R.C. (2005). Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes Dev* *19*, 1455-1465.
- Couture, J.F., Collazo, E., Ortiz-Tello, P.A., Brunzelle, J.S., and Trievel, R.C. (2007). Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat Struct Mol Biol* *14*, 689-695.
- Couture, J.F., Collazo, E., and Trievel, R.C. (2006). Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat Struct Mol Biol* *13*, 698-703.
- Dehe, P.M., Dichtl, B., Schaft, D., Roguev, A., Pamblanco, M., Lebrun, R., Rodriguez-Gil, A., Mkandawire, M., Landsberg, K., Shevchenko, A., *et al.* (2006). Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J Biol Chem* *281*, 35404-35412.
- Dou, Y., Milne, T.A., Ruthenburg, A.J., Lee, S., Lee, J.W., Verdine, G.L., Allis, C.D., and Roeder, R.G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol* *13*, 713-719.
- Dou, Y., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Allis, C.D., Chait, B.T., Hess, J.L., and Roeder, R.G. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* *121*, 873-885.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* *60*, 2126-2132.
- Filippakopoulos, P., Low, A., Sharpe, T.D., Uppenberg, J., Yao, S., Kuang, Z., Savitsky, P., Lewis, R.S., Nicholson, S.E., Norton, R.S., *et al.* (2010). Structural Basis for Par-4 Recognition by the SPRY Domain- and SOCS Box-Containing Proteins SPSB1, SPSB2, and SPSB4. *J Mol Biol*.

Greeson, N.T., Sengupta, R., Arida, A.R., Jenuwein, T., and Sanders, S.L. (2008). Di-methyl H4 Lysine 20 Targets the Checkpoint Protein Crb2 to Sites of DNA Damage. *Journal of Biological Chemistry* 283, 33168-33174.

Halbach, A., Zhang, H., Wengi, A., Jablonska, Z., Gruber, I.M., Halbeisen, R.E., Dehe, P.M., Kemmeren, P., Holstege, F., Geli, V., *et al.* (2009). Cotranslational assembly of the yeast SET1C histone methyltransferase complex. *EMBO J* 28, 2959-2970.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410, 116-120.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* 6, 838-849.

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C.M., and Canaani, E. (2002). ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol Cell* 10, 1119-1128.

Patel, A., Dharmarajan, V., and Cosgrove, M.S. (2008a). Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. *J Biol Chem* 283, 32158-32161.

Patel, A., Dharmarajan, V., Vought, V.E., and Cosgrove, M.S. (2009). On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem* 284, 24242-24256.

Patel, A., Vought, V.E., Dharmarajan, V., and Cosgrove, M.S. (2008b). A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. *J Biol Chem* 283, 32162-32175.

Pflugrath, J.W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr* 55, 1718-1725.

Rampalli, S., Li, L., Mak, E., Ge, K., Brand, M., Tapscott, S.J., and Dilworth, F.J. (2007). p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation. *Nat Struct Mol Biol* 14, 1150-1156.

Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2004). Global proteomic analysis of *S. cerevisiae* (GPS) to identify proteins required for histone modifications. *Methods Enzymol* 377, 227-234.

Schneider, J., Wood, A., Lee, J.S., Schuster, R., Dueker, J., Maguire, C., Swanson, S.K., Florens, L., Washburn, M.P., and Shilatifard, A. (2005). Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Mol Cell* 19, 849-856.

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J* 21, 1121-1131.

- Sheffield, P., Garrard, S., and Derewenda, Z. (1999). Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors. *Protein Expr Purif* 15, 34-39.
- Song, J.J., and Kingston, R.E. (2008). WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *J Biol Chem* 283, 35258-35264.
- Southall, S.M., Wong, P.S., Odho, Z., Roe, S.M., and Wilson, J.R. (2009). Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol Cell* 33, 181-191.
- Steward, M.M., Lee, J.S., O'Donovan, A., Wyatt, M., Bernstein, B.E., and Shilatifard, A. (2006). Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat Struct Mol Biol* 13, 852-854.
- Takahashi, Y.H., Lee, J.S., Swanson, S.K., Saraf, A., Florens, L., Washburn, M.P., Trievel, R.C., and Shilatifard, A. (2009). Regulation of H3K4 trimethylation via Cps40 (Spp1) of COMPASS is monoubiquitination independent: implication for a Phe/Tyr switch by the catalytic domain of Set1. *Mol Cell Biol* 29, 3478-3486.
- Tenney, K., and Shilatifard, A. (2005). A COMPASS in the voyage of defining the role of trithorax/MLL-containing complexes: linking leukemogenesis to covalent modifications of chromatin. *J Cell Biochem* 95, 429-436.
- Thompson, B.A., Tremblay, V., Lin, G., and Bochar, D.A. (2008). CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes. *Mol Cell Biol* 28, 3894-3904.
- Trievel, R.C., and Shilatifard, A. (2009). WDR5, a complexed protein. *Nat Struct Mol Biol* 16, 678-680.
- Vagin, A., and Teplyakov, A. (2000). An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* 56, 1622-1624.
- Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G.N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* 60, 2184-2195.
- Wang, P., Lin, C., Smith, E.R., Guo, H., Sanderson, B.W., Wu, M., Gogol, M., Alexander, T., Seidel, C., Wiedemann, L.M., *et al.* (2009). Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol Cell Biol* 29, 6074-6085.
- Wang, Y.L., Faiola, F., Xu, M., Pan, S., and Martinez, E. (2008). Human ATAC Is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. *J Biol Chem* 283, 33808-33815.

Wang, Y.Y., Liu, L.J., Zhong, B., Liu, T.T., Li, Y., Yang, Y., Ran, Y., Li, S., Tien, P., and Shu, H.B. (2010). WDR5 is essential for assembly of the VISA-associated signaling complex and virus-triggered IRF3 and NF-kappaB activation. *Proc Natl Acad Sci U S A* *107*, 815-820.

Woo, J.S., Suh, H.Y., Park, S.Y., and Oh, B.H. (2006). Structural basis for protein recognition by B30.2/SPRY domains. *Mol Cell* *24*, 967-976.

Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H., and Allis, C.D. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* *121*, 859-872.

Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D.J., Kitabayashi, I., Herr, W., and Cleary, M.L. (2004). Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol Cell Biol* *24*, 5639-5649.

Zhang, X., Yang, Z., Khan, S.I., Horton, J.R., Tamaru, H., Selker, E.U., and Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. *Mol Cell* *12*, 177-185.

ACKNOWLEDGEMENTS

We would like to thank Drs. Marjorie Brand and Jeffrey Dilworth for insightful discussions and for sharing their reagents. We wish to thank Dr. Alexandre Blais for reviewing this manuscript. A Canadian Institutes of Health Research grant to Jean-Francois Couture supported this work. Dr. Couture holds a Canadian Research Chair in Structural Biology and Epigenetics.

CHAPTER 3. THE PLASTICITY OF WDR5 PEPTIDE-BINDING CLEFT ENABLES THE BINDING OF THE SET1 FAMILY OF HISTONE METHYLTRANSFERASES

The following manuscript is published in *Nucleic Acids Research*.

Zhang, P., Lee, H., Brunzelle, J. S., & Couture, J. F. (2012). The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Res*, 40(9), 4237-4246. doi: 10.1093/nar/gkr1235

I used ITC with purified WDR5 and the KMT2/SET1 Win peptides to determine their binding affinities (Figure 3.1). Next, I co-crystallized WDR5 in complex with these peptides (Figures 3.2 and 3.3). Finally, I developed a protocol for purifying the Win and SET domain of each KMT2 enzyme and performed *in vitro* methyltransferase assays to investigate their activity in the absence and presence of the core complex subunits WDR5, RbBP5 and ASH2L (Figures 3.4 and 3.5). I also contributed to the writing and editing of the manuscript.

The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases.

Pamela Zhang¹, Hwabin Lee¹, Joseph S. Brunzelle² and Jean-Francois Couture¹

¹Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, 451 Smyth Road, Ottawa, Ontario K1H 8M5, Canada.

²Feinberg School of Medicine, Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, Chicago, Illinois 60611, USA.

Correspondence should be addressed to J.-F.C.

Short title: WDR5 scaffolds the SET1 family of methyltransferases.

Corresponding author: Jean-Francois Couture

ABSTRACT

In mammals, the SET1 family of lysine methyltransferases (KMTs), which includes MLL1-5, SET1A and SET1B, catalyzes the methylation of lysine-4 (Lys-4) on histone H3. Recent reports have demonstrated that a three-subunit complex composed of WD-Repeat protein-5 (WDR5), Retinoblastoma-Binding Protein-5 (RbBP5) and Absent, Small, Homeotic discs-2-Like (ASH2L) stimulates the methyltransferase activity of MLL1. On the basis of studies showing that this stimulation is in part controlled by an interaction between WDR5 and a small region located in close proximity of the MLL1 catalytic domain (referred to as the WDR5-interacting motif (Win)), it has been suggested that WDR5 might play an analogous role in scaffolding the other SET1 complexes. We herein provide biochemical and structural evidence showing that WDR5 binds the Win motifs of MLL2-4, SET1A and SET1B. Comparative analysis of WDR5-Win complexes reveals that the binding of Win motifs is achieved by the plasticity of the WDR5 peptidyl-arginine-binding cleft in allowing the C-terminal ends of the Win motifs to be maintained in structurally divergent conformations. Consistently, enzymatic assays reveal that WDR5 plays an important role in the optimal stimulation of MLL2-4, SET1A and SET1B methyltransferase activity by the RbBP5-ASH2L heterodimer. Overall, our findings illustrate the function of WDR5 in scaffolding the SET1 family of KMTs and further emphasize on the important role of WDR5 in regulating global histone H3 Lys-4 methylation.

INTRODUCTION

The landscape of each chromosome emanates from an intricate nucleoprotein complex referred to as the nucleosome. Composed of two copies each of histone H3, H4, H2A and H2B and wrapped by a DNA fragment averaging 146 bp, this protein-DNA complex constitutes the first step of DNA compaction and plays an important role in controlling the access of different DNA-binding machineries to various regions of chromosomes. This process is dynamically regulated by ATP-dependent chromatin remodelers along with histone modifying enzymes and regulates a myriad of biological processes including, but not limited to, chromosome silencing, transcription, DNA replication and DNA damage repair (1).

Several post-translational modifications (PTMs), such as acetylation, phosphorylation, ubiquitylation/sumoylation, ADP-ribosylation, PARylation and methylation, have been mapped onto histone proteins and linked to various cellular functions (2). Of interest, methylation of lysine residues by SET domain proteins (Suppressor of Variegation 3-9, Enhancer of Zeste and Trithorax), which frequently exhibit preference towards a precise residue and has the ability to transfer up to three methyl groups to a lysine ϵ -amine, has been linked to important biological processes. Notably, systematic mapping of methyl-lysine marks on histone H3 demonstrated that histone H3 Lys-4 mono-, di- and tri-methylation (H3K4me1/2/3) decorate the promoter and strong enhancer regions of actively transcribed genes. The same study also showed that H3K4me1 localizes to poised enhancers and chromatin regions undergoing transcriptional transition (3, 4). These results further support initial findings linking H3K4 methylation with the promoter regions of actively transcribed genes (5-7).

In the budding yeast *Saccharomyces cerevisiae*, the COMpLex ASSociated with SET1 (COMPASS) (8, 9) is the sole protein complex catalyzing H3K4 methylation while in mammals, this reaction is catalyzed by at least six members: MLL1-4, SET1A, and SET1B. Each mammalian methyltransferase assembles in a COMPASS-like complex and despite sharing similarities in substrate specificity, the SET1 KMTs are not functionally redundant (10-16) and correlatively, binding studies have demonstrated that members of the SET1 family interact with unique subsets of proteins. The KMTs MLL2 and MLL3 have been shown to interact with Activating Signal Cointegrator 2 (ASC2) (17) while WDR82 and menin associate with SET1A/SET1B (18-20) and MLL1/MLL2 (11, 15), respectively. Despite these differences, members of the SET1 family share several features including the complexity of their domain organization and the presence of a catalytic SET domain. In addition, all but MLL5 associate with a three-subunit complex composed of WDR5, RbBP5 and ASH2L, collectively referred to as the MLL core complex (21, 22), which is required for the stimulation of MLL1 di/tri-methyltransferase activity both *in vitro* (23) and *in vivo* (22, 24).

Biochemical and structural studies have recently assigned discrete functions to RbBP5, ASH2L and WDR5 in the assembly of the core complex and stimulation of MLL1 KMT activity. Apart from its DNA binding domain (25, 26), ASH2L binds RbBP5 via its SPIa and Ryanodine receptor (SPRY) domain and the ASH2L/RbBP5 heterodimer stimulates MLL1 methyltransferase activity (27). In addition to its ASH2L-interacting region, RbBP5 comprises residues that directly bind WDR5, an interaction important for maintaining the integrity of the core complex and indirectly, the stimulation of MLL1 KMT activity (28, 29). On the opposite side of the WDR5 β -propeller domain, WDR5 binds a small motif preceding the SET domain of MLL1 (23, 30). Also referred to as the WDR5-

interacting motif (Win), the interaction between the MLL1 Win motif (MLL1_{Win}) and WDR5 is important in the scaffolding of the MLL1-WDR5-RbBP5-ASH2L complex and stimulation of MLL1 KMT activity *in vitro* (23). Based on these findings and sequence homology between the Win motifs, the same authors surmised that WDR5 would play an analogous role in binding the other members of the SET1 family, namely MLL2-4, SET1A and SET1B (31).

We herein provide biochemical evidence demonstrating that WDR5 binds with different equilibrium dissociation constants the peptides corresponding to the Win motifs of the other SET1 family members. Correspondingly, the crystal structures of WDR5 in complex with MLL2-4, SET1A and SET1B Win motifs reveal that WDR5 accommodates the different Win motifs using divergent networks of hydrogen bonds and hydrophobic contacts. Finally, we show that WDR5, when in complex with the ASH2L-RbBP5 heterodimer, is important for the full stimulation of MLL2-4, SET1A and SET1B methyltransferase activity. Overall these findings further support the fundamental role of WDR5 in scaffolding the SET1 complexes and in indirectly stimulating H3K4 methylation.

MATERIALS AND METHODS

Protein Purification

Constructs corresponding to the WDR5-interacting (Win) and SET domains of MLL1 (3745-3969), MLL2 (5344-5536) and MLL4 (2512-2715) were cloned into a pGST based vector while those for MLL3 (4714-4911), SET1A (1487-1707) and SET1B (1701-1923) were cloned into a home-made vector (pSMT3). MLL1 and MLL3 were overexpressed as GST and His-SUMO fusion proteins, respectively, in Rosetta cells (Novagen) with 0.1 mM IPTG for 16 hours at 18°C. Due to their poor solubility, MLL2, MLL4, SET1A and SET1B were

overexpressed in Arctic cells (Stratagene) with 1 mM IPTG for 48 hours at 10°C. Cells were harvested in either 50 mM sodium phosphate (NaPi) pH 7.0, 500 mM sodium chloride (NaCl) and 5 mM β -mercaptoethanol (BME) (pSMT3) or phosphate buffered saline (pGST), lysed by sonication, clarified by centrifugation and filtration, purified by affinity chromatography and ULP1-(pSMT3) or TEV-(pGST2) cleaved on beads for 16 hours at 4°C. Following cleavage, the enzymes were dialyzed into a final buffer of 20 mM Tris-HCl pH 7.5, 300 mM NaCl, and 5 mM BME. The members of the core complex, including ASH2L, RbBP5 and WDR5 (22-334) were overexpressed and purified as previously described (33). Following gel filtration, the core complex subunits were dialyzed in 20 mM Tris-HCl pH 7.5, 300 mM NaCl, and 5 mM BME. The concentrations of the SET1 proteins were quantified either by UV or coomassie-stained SDS-PAGE gel.

Isothermal Titration Calorimetry (ITC)

ITC experiments were performed using a VP-ITC calorimeter (MicroCal) as previously described (29, 33). Briefly, a solution of peptide (0.5-1 mM) derived from the Win motif of MLL1, MLL2, MLL3, MLL4, SET1A or SET1B was injected into a solution of WDR5 (0.06 mM). Titrations were performed at 19°C in 20 mM NaPi pH 7.0, 100 mM NaCl and 5 mM BME. To decrease the rate of saturation of the WDR5 peptidyl-arginine-binding cleft, the volume of injection of the MLL4_{win} peptide was decreased by half (from 10 μ l to 5 μ l) for injections 9-19. Titration curves were analyzed using the Origin software. The K_D and the error were determined by fitting the integrated data to a standard single-site binding model.

Protein Crystallization, Data Collection and Structure Determination

Equimolar amounts of purified WDR5 and the Win peptides were incubated on ice for one hour and crystallized using the sitting- or hanging-drop vapor diffusion methods.

Crystallization conditions for each complex vary but the predominant components in the mother liquor are ammonium sulfate and polyethylene glycol (PEG) 3350. Crystals were sequentially soaked in the mother liquor supplemented with 5%, 10%, 15% and 20% of glycerol, harvested and flash frozen in liquid nitrogen. Data sets were collected at the Life Science Collaborative Access Team (LS-CAT) beamline at the Advance Photon Source of the Argonne National Laboratory and indexed with HKL2000 (34). The structure of each WDR5-Win complex was solved by molecular replacement using Phaser and histone H3 bound WDR5 as a search model (RCSB 2H13.pdb) (35). The structure of each WDR5-Win complex was further improved using iterative cycles of refinement, using either Refmac (36) or Buster (37), and model building with COOT (38). The quality of the structures was validated using Molprobity (39).

***In vitro* Methyltransferase Assay**

Methyltransferase assays were performed as previously described (29, 33). Briefly, each recombinantly purified SET1 HKMT was incubated with equimolar amounts of WDR5, RbBP5 and ASH2L in 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 3 mM dithiothreitol, 5 mM magnesium chloride, 5% glycerol and 10 μ M of histone H3. Reactions were initiated by adding 1 μ Ci of radiolabeled S-adenosyl-L-Methionine (AdoMet) (0.07 mM of tritiated AdoMet) and stopped by spotting the reactions onto Whatman P-81 filter papers. Free AdoMet was removed by washing the filter papers four times in 250 ml of 50 mM sodium

bicarbonate pH 9.0 and methyltransferase activity was quantified by liquid scintillation counting.

RESULTS

Equilibrium Dissociation Constants of the Win Motifs for WDR5

A recent biochemical analysis of the WDR5 peptidyl-arginine-binding cleft revealed that this protein binds a 10-residue peptide corresponding to MLL1_{Win} with an equilibrium dissociation constant (K_D) of 1.7 μ M (31). The crystal structure of WDR5 in complex with MLL1_{Win} identified that R3765 of MLL1 was critical in conferring binding to WDR5, scaffolding the MLL1-WDR5-RbBP5-ASH2L complex and correlatively, stimulating MLL1 methyltransferase activity (23). Accordingly, the same authors posited that a similar region found on other members of the SET1 family would play an analogous role (31) while another study proposed that only MLL1 and MLL4 could bind to WDR5 (30). In order to investigate these hypotheses, we first sought to measure the K_D values of WDR5 for MLL1-4 and SET1A/B Win motifs. Using isothermal titration calorimetry (ITC), we determined that WDR5 binds MLL1_{Win} and MLL4_{Win} peptides with a K_D of 1.13 μ M and 0.03 μ M, respectively. Similar titration experiments of WDR5 with MLL2_{Win} and MLL3_{Win} peptides resulted in K_D of 0.16 μ M and 2.03 μ M, respectively, while SET1A_{Win} and SET1B_{Win} peptides bound to WDR5 with a K_D of 0.26 μ M and 0.10 μ M, respectively. Overall, these results demonstrate that WDR5 binds the Win region of all members of the SET1 family of KMTs. However, given the divergence in the amino acid sequence succeeding the conserved arginine residue and the ~37-fold range of binding affinities of WDR5 for the Win motifs

(Table 3.1 and Figure 3.1), our binding analysis suggests that WDR5 interacts with SET1 KMTs using divergent binding modes.

Table 3.1: Alignment of SET1 Win Motifs											
	Amino acid sequence										
	P ₋₃	P ₋₂	P ₋₁	P ₀	P ₊₁	P ₊₂	P ₊₃	P ₊₄	P ₊₅	P ₊₆	P ₊₇
MLL1 _{WIN}	G ₃₇₆₂	S	A	R	A	E	V	H	L	R	K
MLL2 _{WIN}	G ₅₀₆₂	C	A	R	S	E	P	K	I	L	T
MLL3 _{WIN}	G ₄₇₀₇	S	A	R	A	E	P	K	M*	<u>S</u>	<u>A</u>
MLL4 _{WIN}	G ₂₅₀₈	A	A	R	A	E	V	Y	L	R	<u>K</u>
SET1A _{WIN}	G ₁₄₉₂	S	A	R	S	E	G	Y	<u>Y</u>	<u>P</u>	<u>I</u>
SET1B _{WIN}	<u>G</u> ₁₈₈₂	C	A	R	S	E	G	F	Y	T	<u>I</u>

* Side chain is missing

Underlined residues have not been modeled in the crystal structure.

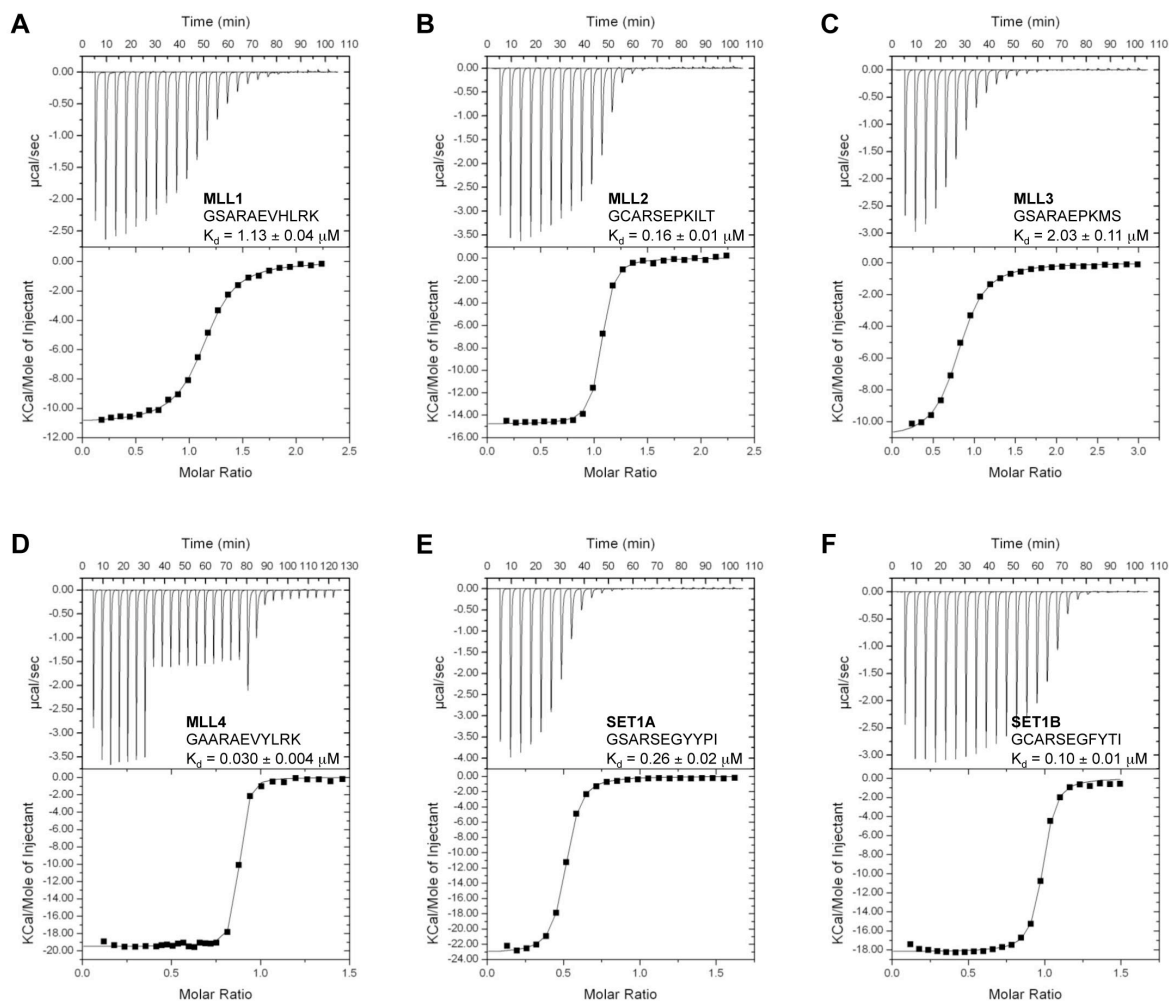


Figure 3.1. WDR5 binds MLL1-4, SET1A and SET1B Win motifs.

ITC titration experiments with MLL1_{Win} (A), MLL2_{Win} (B), MLL3_{Win} (C), MLL4_{Win} (D), SET1A_{Win} (E) and SET1B_{Win} (F) peptides and WDR5 (*upper* trace) and the fitted binding curve (*lower* trace). The sequence of the Win motifs and K_D are indicated as insets. Titration of WDR5 with the Win peptides displays binding stoichiometries (N values) between 0.9 and 1.0.

Crystal Structures of the WDR5-MLL_{Win} Complexes

To gain insights into the mechanisms underlying WDR5 ability to bind the structurally divergent C-terminal ends of the Win peptides (Table 3.1), we have determined the crystal structures of WDR5 in complex with peptides corresponding to the Win motifs of MLL2 (1.4 Å resolution), MLL3 (2.2 Å resolution), MLL4 (1.57 Å resolution), SET1A (2.1 Å resolution) and SET1B (2.2 Å resolution) (Figure S3.1). The WDR5-Win complexes superimpose with an overall root-mean-square deviation (r.m.s.d.) of less than 0.3 Å for all aligned atoms, showing that binding of the Win peptides does not result in noticeable changes in the overall structure of WDR5. Simulated annealing omit maps revealed clear electron density for several residues flanking the central arginine residue of each Win motif (Figure S3.1). However, the distal residues in the peptides, including the M4715 side chain (single letter code refers to residues of the Win motifs), S4716 and A4717 of MLL3_{Win}, K2518 of MLL4_{Win}, P1501 and I1502 of SET1A_{Win} and G1882 and I1892 as well as the side chains of F1889 and Y1890 of SET1B_{Win}, lack interpretable electron density and were not modeled. Close inspection of the WDR5-Win structures reveals that all the peptides bind the WDR5 peptidyl-arginine-binding cleft, a crevice located on one side of the WDR5 β -propeller structure (Figure 3.2), in which each peptide engages in several hydrogen bonds, van der Waals contacts and hydrophobic interactions with WDR5. To facilitate the discussion of the interactions between WDR5 and the Win motifs, we have used a modified version of the Schechter-Berger notation for assigning residues of the Win motifs in which the central arginine residue is denoted as P₀ (Table 3.1). In addition, interactions between WDR5 and the Win motifs have been separated into two categories based on interactions with residues preceding and succeeding the P₊₂ glutamate residue of the Win motifs (Table 3.1).

Consistent with the sequence homology of the residues flanking the central arginine (P_0), including those at P_{-3} , P_{-2} , P_{-1} , P_{+1} and P_{+2} position (Table 3.1), this region of the peptide adopts similar orientations and engages in several shared direct and water-mediated hydrogen bonds and hydrophobic interactions with WDR5. Similar to MLL1_{Win}, the P_{-2} to P_{+2} region folds as a 3_{10} helix in which the side chains of the P_{-3} , P_{-2} , and P_{-1} residues pack tightly within the WDR5 peptidyl-arginine-binding cleft and make hydrophobic contacts with the Ala-65, Gly-89, Ile-90 and Asp-107 side chains. The same residues in WDR5 form a wall that maintains the side chain of the P_{-2} residue in an orientation permissive to make a hydrogen bond with the backbone amide of the P_{+1} residue. The alanine residue in the P_{-1} position binds analogously in all the WDR5-Win complexes. Indeed, the P_{-1} amide group engages in a hydrogen bond with the Asp-107 carboxylate group and its side chain makes hydrophobic contacts with the side chains of Phe-133 and Phe-149. Similar to histone H3 R2 (35, 40-42) and MLL1 R3765 (30, 31), the central arginine (P_0) of the Win motifs extends inside the central toroid cavity of WDR5 in which its guanidinium group interacts with the side chains of Phe-133 and Phe-263. The P_0 residue also makes several direct hydrogen bonds with the main chains of Ser-91, Phe-133 and Cys-261 and a water-mediated hydrogen bond with the Ser-175 backbone carbonyl group. Following the central arginine residue, the P_{+1} position is occupied by a small residue such as a serine (MLL2, SET1A and SET1B) or an alanine (MLL3 and MLL4) while the P_{+2} position is occupied by a glutamate residue. Consistent with the sequence homology between the P_{+1} and P_{+2} residues, this portion of the Win motifs shares several interactions with WDR5. The P_{+1} residue makes van der Waals contacts with the side chains of Ala-47 and Tyr-260 while in MLL4 the P_{+2} carboxylate moiety makes two additional water-mediated hydrogen bonds with the backbone amides of

the P₊₃ and P₋₃ residues. In addition, the hydroxyl group of the MLL2_{Win} and SET1B_{Win} P₊₁ serine residue makes a hydrogen bond with the sulfhydryl group of the P₋₂ cysteine residue.

Despite high sequence homology and analogous binding modes of the P₋₃, P₋₂, P₋₁ and P₀ residues, structural alignment of the WDR5-Win complexes reveals notable structural differences. In comparison with MLL1, the backbone of the MLL2_{Win} and MLL4_{Win} P₋₃ residue is rotate 180°, which results in additional water-mediated and direct hydrogen bonds with WDR5 (Figure 3.3). This rotation leads to a small structural displacement of the P₋₂ backbone with negligible consequences on the positioning of the C5063 (MLL2), S4708 (MLL3), A2509 (MLL4), S1493 (SET1A) and C1883 (SET1B) side chains and their interactions with WDR5. However, in clear contrast to the P₋₂, P₋₁, P₀, P₊₁ and P₊₂ residues, the P₊₃ to P₊₅ residues adopt divergent orientations. MLL2_{Win} and MLL3_{Win} peptides, which bind analogously to WDR5, interact in such a way that their respective C-termini point towards blade 5 of WDR5 while MLL4_{Win} and MLL1_{Win} adopt a conformation placing their C-terminal ends towards blade 4 of the β-propeller. SET1A_{Win} and SET1B_{Win} peptides bind in an intermediate orientation such that their C-termini are maintained on the surface of blades 4 and 5 of WDR5 (Figure 3.2A). Given the structural differences adopted by the residues succeeding the P₊₂ residue (Figure 3.3), we purported that each peptide would engage in distinct interactions with WDR5. The backbone of P₊₃ Y2514 of the MLL4_{Win} peptide, which binds similarly to H3769 of MLL1_{Win}, forms direct and water-mediated hydrogen bonds with the Asp-172 and Lys-259 carboxylate and carbonyl groups, respectively. In addition, its aromatic ring makes hydrophobic contacts with the side chains of Tyr-191, Pro-173 and Phe-149. In contrast to the solvent exposed P₊₄ L2516, the P₊₅ R2517 side chain is nestled in a small cleft formed by the surface of blade 4 and the loop connecting blade 4 and blade 5 of WDR5 (Figure 3.3A). This crevice, which is composed of

Tyr-191, Asp-192, Gly-193, Asn-214, Pro-215 and Pro-216, likely orients the R2517 side chain in a permissive orientation to engage in direct hydrogen bonds with the backbone carbonyl groups of Tyr-191 and Asn-214 and van der Waals contacts with Pro-216.

Titration analysis revealed that MLL4_{win} binds ~37-fold more tightly to WDR5 than the highly homologous MLL1_{win} peptide (Figure 3.1 and Table 3.1). Owing to the high sequence homology between these two motifs and structural homology between the WDR5 bound MLL1_{win} and MLL4_{win} complexes, the differences in binding affinities are seemingly at odds. However, while a comparative analysis of the WDR5-MLL1_{win} and WDR5-MLL4_{win} complexes reveals that both MLL4_{win} and MLL1_{win} P₊₄ residues share extensive interactions with WDR5, it also shows that a water-mediated interaction of the MLL1_{win} P₊₄ H3769 side chain with the WDR5 Asp-172 carboxylate group is replaced by a direct hydrogen bond between the same residue in WDR5 and the side chain hydroxyl group of MLL4_{win} P₊₄ Y2514; a difference that presumably explains the dissimilarity between MLL1_{win} and MLL4_{win} binding affinities for WDR5.

Owing to the presence of a proline residue in the P₊₃ position of MLL2_{win} and MLL3_{win}, we postulated that these peptides would bind WDR5 differently than MLL1_{win} and MLL4_{win}. Consistent with this hypothesis, structural alignment of WDR5-MLL2_{win} and WDR5-MLL1_{win} complexes reveals that the C-terminal end of MLL2_{win} shifts towards the loop preceding β 18, resulting in several novel interactions between MLL2_{win} and WDR5 (Figure 3.3B). In MLL2_{win}, the P₊₃ P5068 makes close van der Waals contacts with the side chains of Lys-259 and Tyr-260. The WDR5 peptidyl-arginine-binding cleft also undergoes notable structural changes. The phenolic side chain of Tyr-191 rotates 90°, positioning its aromatic ring in a near parallel orientation with the P₊₄ lysine residue of MLL2_{win}/MLL3_{win}. In addition, the same lysine residue in MLL2_{win}/MLL3_{win} forms a novel hydrogen bond with

the Lys-259 carbonyl group and hydrophobic contacts with Leu-234. Finally, in contrast to all of the other Win motifs, the MLL2_{Win} P₊₅ residue L5071 lies within a hydrophobic cleft composed of Pro-216 and Leu-234. Overall, the crystal structure of MLL2_{Win} bound to WDR5 suggests that the P₊₃ residue acts as a structural determinant underlying the binding mode of a peptide within the WDR5 peptidyl-arginine-binding cleft.

After observing that the P₊₃ position in MLL2_{Win}/MLL3_{Win} peptides is a key point of structural divergence between the Win peptides, we postulated that the P₊₃ glycine residue of SET1A_{Win}/SET1B_{Win} motifs would also exhibit structural differences compared to the other Win motifs. Accordingly, in contrast to the MLL1-4 Win peptides, the backbone carbonyl of the SET1A_{Win} P₊₂ glutamate residue undergoes a 180° rotation that results in novel hydrogen bonds between the amide groups of the P₊₃ and P₊₄ residues and the backbone carbonyls of the P₋₁ residue and Lys-259 of WDR5, respectively. Notably, in comparison with the WDR5-MLL1_{Win} complex, the side chains of Lys-259 and Tyr-191 shift toward the Win peptides, resulting in the formation of additional hydrophobic contacts with the SET1A_{Win} P₊₃ glycine and P₊₅ tyrosine residues (Figure 3.3C). In addition, the same tyrosine residue in SET1A engages in several hydrophobic contacts with the lateral chain of Leu-234 and Pro-216.

Overall, the crystal structures of the WDR5-Win complexes highlight WDR5's ability in binding the Win motifs using alternative structural determinants. More importantly, our structural studies suggest that the plasticity of the WDR5 peptidyl-arginine-binding cleft stems from its ability to allow subtle but key structural changes that accommodate the binding of the structurally divergent residues in position P₊₃. Finally, the crystal structure of the WDR5-Win complexes suggest that the type of residue at the P₊₃ position will likely play an important role in the binding mode adopted by the peptide within the WDR5 peptidyl-arginine-binding cleft.

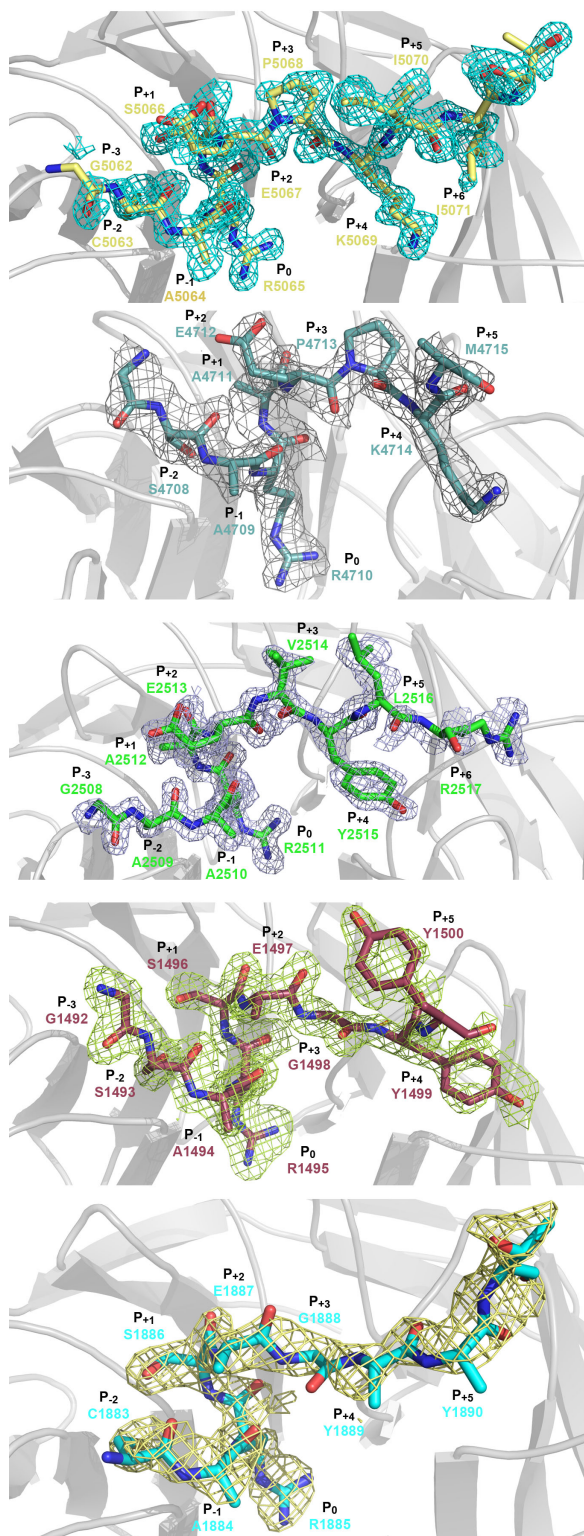


Figure S3.1. A clear electron density map is observed for all the Win motifs.

Shown is the omit map for the Win motif of the SET1 proteins. Contoured at 2σ , the maps for A) MLL2 (cyan), B) MLL3 (black), C) MLL4 (blue), D) SET1A (green) and E) SET1B (yellow) are rendered with different colors and the carbon atoms for each peptide are highlighted as in Figure 3.2.

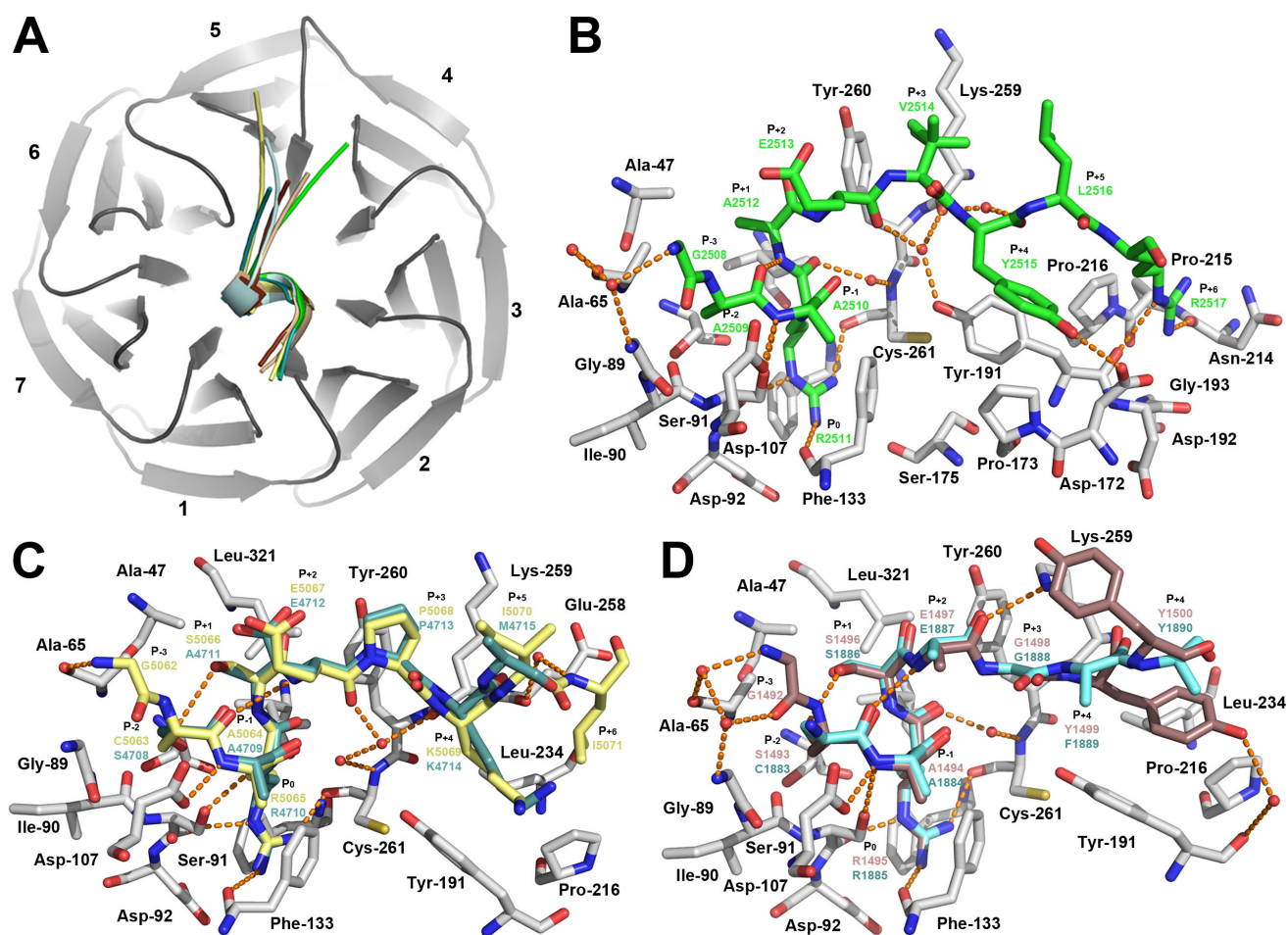


Figure 3.2. Crystal structure of WDR5 in complex with MLL2-4, SET1A and SET1B Win motifs.

A) Overall structure of WDR5 (dark grey) showing the seven blades and the relative orientations of MLL1 (beige), MLL2 (yellow), MLL3 (turquoise), MLL4 (green), SET1A (firebrick) and SET1B (cyan) Win motifs.

B) Crystal structure of the WDR5-MLL4_{Win} complex. Zoomed view of the WDR5 peptidyl-arginine-binding cleft in which WDR5 and MLL4 carbon atoms are rendered in grey and green, respectively.

C) Crystal structures of WDR5-MLL2_{Win} and WDR5-MLL3_{Win} complexes. MLL2 and MLL3 carbon atoms are highlighted as in (A).

D) Crystal structures of WDR5-SET1A_{Win} and WDR5-SET1B_{Win} complexes. SET1A and SET1B carbon atoms are colored as in (A). Water molecules and key hydrogen bonds are depicted as red spheres and orange dashed lines, respectively.

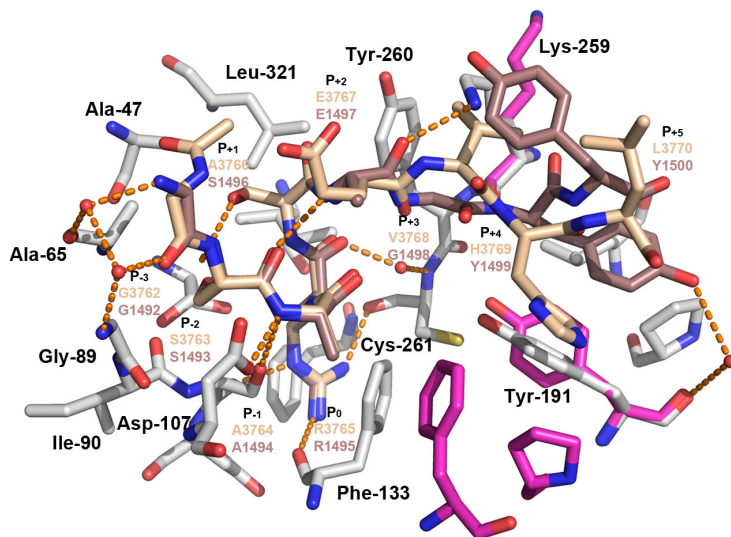
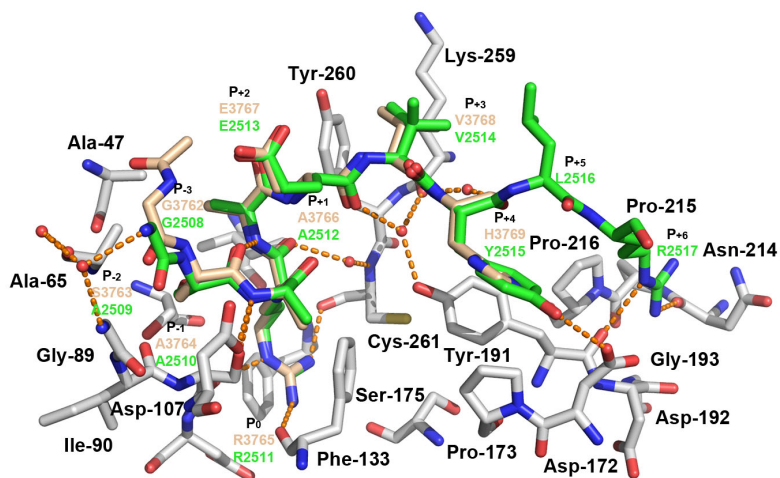
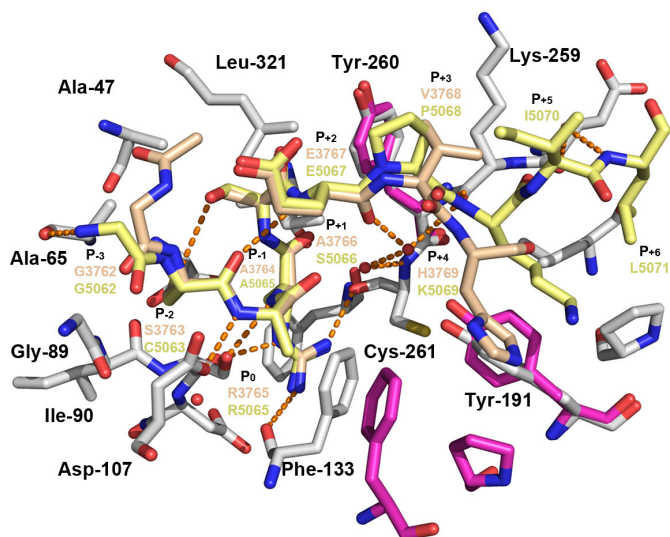


Figure 3.3. Comparative analysis of Win motif binding modes.

A) Superimposition of WDR5 bound MLL1 (beige) and MLL2 (yellow) peptides.

B) Overlay of MLL1 (beige) and MLL4 (green) peptides in complex with WDR5 (grey).

C) Structural alignment of WDR5-MLL1_{Win} and WDR5-SET1A_{Win} complexes in which carbons of the peptides are highlighted as in Figure 3.2. Residues of the WDR5 peptidyl-arginine-binding cleft adopting divergent conformation between the complexes are rendered as pink carbon atoms. Hydrogen bonds and water molecules are rendered as in Figure 3.2.

Role of WDR5 in Stimulating the KMT Activity of the SET1 Family of

Methyltransferases

The Win motif of MLL1 is important for its interaction with the WDR5-RbBP5-ASH2L complex and thereby stimulation of its methyltransferase activity (23). After showing that WDR5 binds the Win motifs of MLL1-4 and SET1A/B, we examined the role of the WDR5-RbBP5-ASH2L complex in the stimulation of other members of the SET1 family of methyltransferases. A comparison of SET1 KMT activity in the absence or presence of the WDR5-RbBP5-ASH2L complex confirmed that the core complex enhances the methyltransferase activity of all members, with the exception of MLL5, with however differences in the fold of stimulation. Specifically, the WDR5-RbBP5-ASH2L complex stimulated MLL2/MLL3 and MLL4 activity by ~52/256- and ~15-fold, respectively, while co-incubation of SET1A and SET1B with the core complex subunits increased methylation of histone H3 by ~40 and ~1290-fold, respectively. The stimulatory roles of the core complex on MLL4 and SET1A can be rationalized by the nearly undetectable activity of these KMTases on histone H3 in the absence of the WDR5-RbBP5-ASH2L complex. Overall, these observations support initial *in vivo* data showing that the members of the core complex are critical for global level of histone H3K4 methylation.

After showing that the core complex subunits stimulate the methyltransferase activity of MLL2-4, SET1A and SET1B, we sought to measure the contribution of WDR5 in the enzymatic activity of each complex (Figure 3.4). Interestingly, incubation of the ASH2L/RbBP5 heterodimer with each KMT failed to fully stimulate, to various extents, the enzymatic activity of each member of the SET1 family. Specifically, the absence of WDR5 resulted in a 2-fold decrease of activity for SET1B, MLL3 and MLL4 while co-incubation of SET1A and MLL2 with the ASH2L-RbBP5 heterodimer decreased the methylation of

histone H3 by 5-fold and 1.5-fold, respectively (Figure 3.4). Overall, these results clearly demonstrate the importance of WDR5 in stimulating the methyltransferase activity of MLL2-4, SET1A and SET1B and support recent studies suggesting that WDR5 is key for the activity of all members of the SET1 family of KMTases.

Following the identification that the WDR5-RbBP5-ASH2L complex stimulates all members of the SET1 family, we sought to evaluate the methylation of histone H3 by MLL1-4 in the absence of the core complex subunits. To examine the catalytic activity of each MLL, we incubated each HKMTase with the histone H3 peptide and the tritiated cofactor. As shown in Figure 3.5, we failed to detect histone H3 methylation by MLL1, MLL2 and MLL4 while a similar assay performed with MLL3 resulted in a notable accumulation of the methylated peptide. These observations suggest that MLL3 enzymatic activity may play a role independent of the WDR5-RbBP5-ASH2L complex.

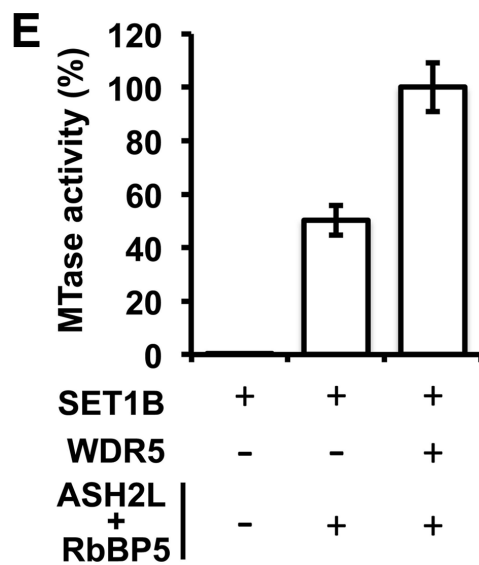
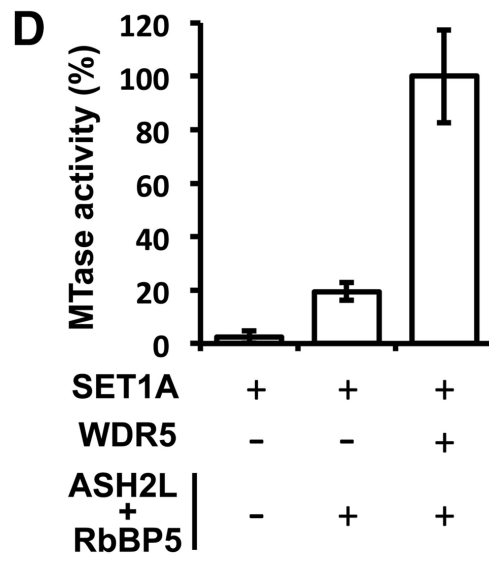
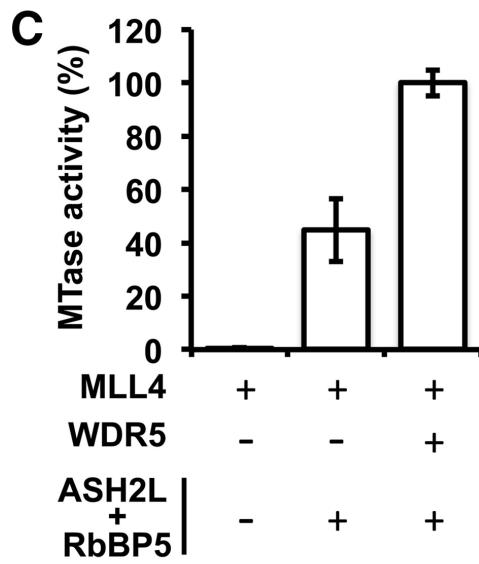
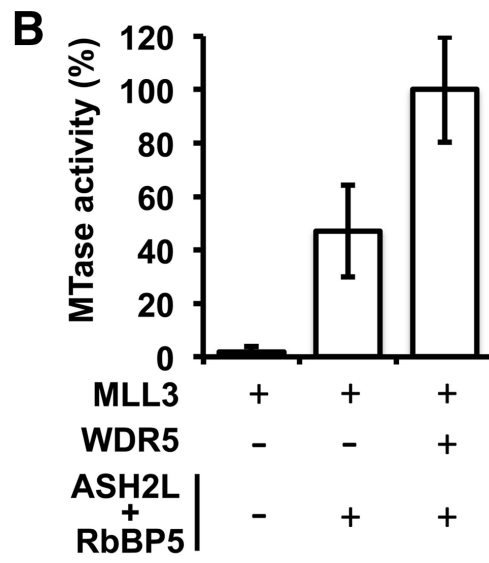
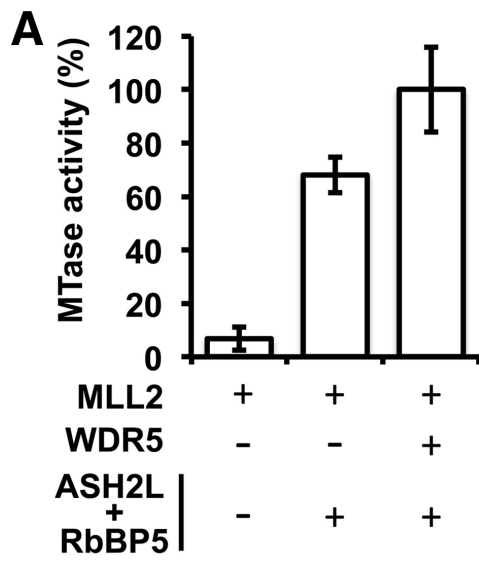


Figure 3.4. Stimulation of the SET1 family of KMT activity on histone H3 by the core complex.

Radiometric methyltransferase assays performed with MLL2 (A), MLL3 (B), MLL4 (C), SET1A (D) and SET1B (E) either in the absence or presence of WDR5 and the RbBP5-ASH2L complex. Methyltransferase assays were performed at a concentration of the enzyme in the linear range of activity of MLL2 (1 μ M), MLL3 (0.5 μ M), MLL4 (1 μ M), SET1A (1 μ M) and SET1B (0.5 μ M) complexes. Activity is represented as the average of three independent experiments performed in triplicate. Error bars indicate the standard deviation between the assays.

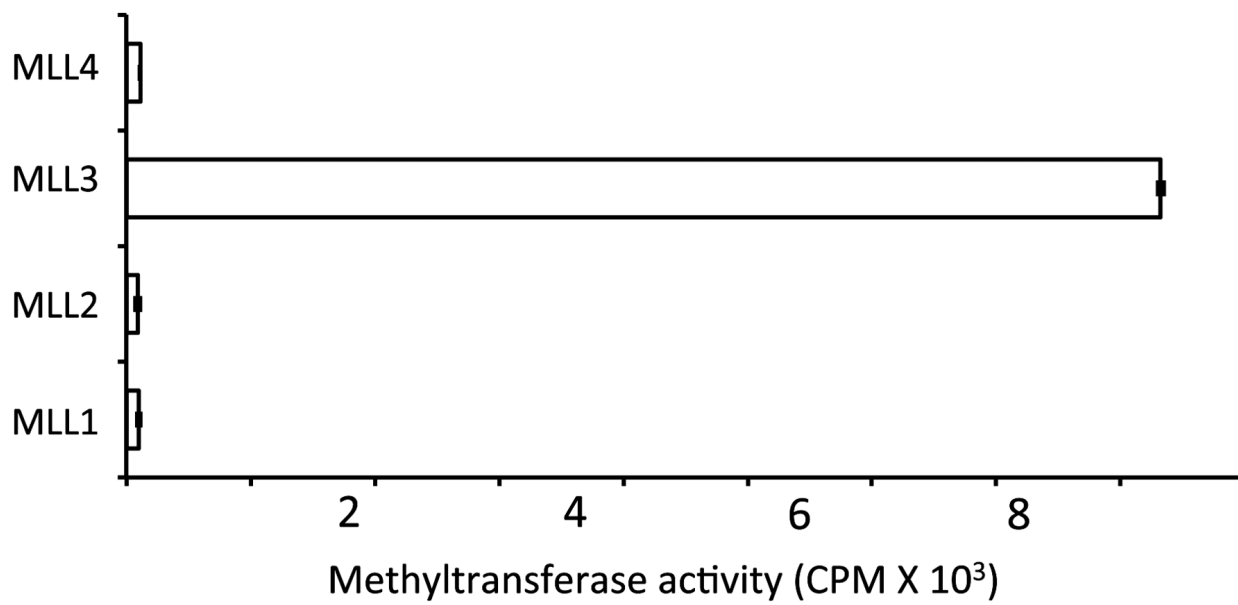


Figure 3.5. MLL3 methylates histone H3 in the absence of the core complex subunits. Comparative analysis of MLLs enzymatic activity in the absence of the core complex subunits. Assays were performed in the presence of 0.5 μ M of enzymes.

DISCUSSION

The results presented herein provides biochemical and structural evidence that the WDR5 peptidyl-arginine-binding cleft is important in the binding of all members of the SET1 family of methyltransferases stimulated by the core complex.

Of interest, we observed that MLL2 and MLL3 Win motifs bind similarly to WDR5. Owing to the presence of a proline residue in P₊₃, Song & Kingston had previously proposed that MLL2/MLL3 might interact with WDR5 using a different mode of binding (30). Comparative analyses of the WDR5-MLL2_{Win} and WDR5-MLL1_{Win} complexes support this hypothesis as the C-terminal ends of both peptides are maintained by specific sets of hydrogen bonds, hydrophobic contacts and van der Waals interactions. Similarly, we also observed noticeable differences between the WDR5-SET1A_{Win} and WDR5-MLL1_{Win} complexes that include rotation of the SET1A_{Win} P₊₂ backbone, a change that could be attributed to the presence of a glycine residue in P₊₃, which results in the reorientation of the P₊₄ and P₊₅ residues. These structural differences are accommodated by the reorganization of residues forming the WDR5 peptidyl-arginine-binding cleft including Lys-259 and Tyr-191. Although it was evidenced by several mutational studies (23, 28, 30, 33, 35, 41) that the N-terminal end of the Win motifs is important for binding to WDR5, our comparative analysis suggest that the propensity of WDR5 in binding the members of the SET1 family of HKMTs is achieved by the plasticity of its peptidyl-arginine-binding cleft and more specifically its ability to accommodate the divergent C-terminal region of the Win motifs.

In metazoans, WDR5 plays a critical role in regulating/maintaining proper levels of histone H3K4 methylation (21, 22, 24) while in *Drosophila*, deletion of the will die slowly protein (WDR5 homolog) leads to global loss of methylation and ultimately lethality (43). Similarly, morpholino knockdown of WDR5 in *Xenopus* leads to severe developmental

deficiencies including gut, hematopoietic and somatic defects (21). We propose that the plasticity of the WDR5 peptidyl-arginine-binding cleft plays an important role in scaffolding all of the SET1 complexes and underlie the fundamental roles of WDR5 in regulating various developmental programs (21, 44).

MLL3, a Unique Member of the SET1 Family of HKMTs

The characterization of SET1 enzymatic activity on histone H3 presented herein provides the first comparative analysis for the SET1 family of methyltransferases. While the SET domains of MLL2, MLL4 and SET1B displayed negligible enzymatic activity in the absence of the WDR5-RbBP5-ASH2L complex, we observed detectable histone H3 methyltransferase activity for MLL3 independent of the core complex. This observation that MLL3 efficiently methylates histone H3 in the absence of the core complex while MLL2 (MLL4 in rodent) strictly depends on the presence of the core complex subunits is interesting. Indeed, while initial studies have suggested that MLL3 and MLL4 are functionally redundant in retinoic acid receptor transactivation in mouse embryonic fibroblasts (45), targeted inactivation of MLL3 resulted in several defects including lower body mass, hypofertility and increased formation of ureter epithelial tumors (45, 46). Based on our findings, we propose that the ability of MLL3 to methylate histone H3 in the absence of the core complex subunits may underlie some of the non-overlapping functions between MLL3 and MLL4.

The Expanding Role of WDR5

The recent findings that WDR5 binds to proteins unrelated to the trithorax group proteins shed new light on the ever-expanding role of WDR5. It has been shown recently that

Oct4 forms a complex with WDR5 independent of the presence of RbBP5, ASH2L or the SET1 family members (44). By identifying an interaction between WDR5 and the embryonic stem cell core transcriptional network, the authors of this study established that WDR5 expression is necessary for the efficient formation of induced pluripotent stem cells (44). In addition, Gan et al. recently demonstrated that WDR5 regulates smooth muscle cell-selective gene activation by interacting with the homeodomain of pituitary homeobox 2 (Pitx2) (47). Similarly, WDR5 has been shown to interact with the VISA-associated signaling complex during virus infection-triggered IRF3 and NF- κ B activation (48) and with the NRC/NCoA6 interacting factor 1 (49). Finally, recent MudPit analysis identified WDR5 as a subunit of the non-specific lethal complex (NSL) (50). These examples further emphasize the important role of WDR5 in scaffolding these complexes and we propose that the plasticity derived from the WDR5 peptidyl-arginine-binding cleft, or its recently characterized V-shape cleft (29), plays an important role in these biological processes.

FUNDING

This work was supported by a Canadian Institute of Health Research grant (J-F C) and an early research award from the Ministry of Research and Innovation (province of Ontario). Dr. Couture holds a Canada Research Chair in Structural Biology and Epigenetics. P.Z. is supported by a scholarship from the Natural Sciences and Engineering Research Council of Canada (CREATE).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online: Supplementary Table S1 and Supplementary Figure S1

ACKNOWLEDGEMENTS

We would like to thank Dr. Alain Doucet, Sylvain Lanouette and Vanessa Mongeon for providing comments on the manuscript.

FOOTNOTES

Coordinates and structure factors for each WDR5-Win complex will be released after publication of this manuscript.

REFERENCES

1. Allis, C.D., Jenuwein, T. and Reinberg, D. (2007) *Epigenetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
2. Bannister, A.J. and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Res*, **21**, 381-395.
3. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43-49.
4. Zhou, V.W., Goren, A. and Bernstein, B.E. (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet*, **12**, 7-18.
5. Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P. *et al.* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **448**, 553-560.
6. Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J. and Kouzarides, T. (2002) Active genes are trimethylated at K4 of histone H3. *Nature*, **419**, 407-411.
7. Schneider, R., Bannister, A.J., Myers, F.A., Thorne, A.W., Crane-Robinson, C. and Kouzarides, T. (2004) Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol*, **6**, 73-77.
8. Tenney, K. and Shilatifard, A. (2005) A COMPASS in the voyage of defining the role of trithorax/MLL-containing complexes: linking leukemogenesis to covalent modifications of chromatin. *J Cell Biochem*, **95**, 429-436.
9. Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F. and Shilatifard, A. (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A*, **98**, 12902-12907.
10. Cho, Y.W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G.R., Copeland, T.D., Kalkum, M. *et al.* (2007) PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem*, **282**, 20395-20406.

11. Hughes, C.M., Rozenblatt-Rosen, O., Milne, T.A., Copeland, T.D., Levine, S.S., Lee, J.C., Hayes, D.N., Shanmugam, K.S., Bhattacharjee, A., Biondi, C.A. *et al.* (2004) Menin associates with a trithorax family histone methyltransferase complex and with the *hoxc8* locus. *Mol Cell*, **13**, 587-597.
12. Patel, S.R., Kim, D., Levitan, I. and Dressler, G.R. (2007) The BRCT-domain containing protein PTIP links PAX2 to a histone H3, lysine 4 methyltransferase complex. *Dev Cell*, **13**, 580-592.
13. Issaeva, I., Zonis, Y., Rozovskaia, T., Orlovsky, K., Croce, C.M., Nakamura, T., Mazo, A., Eisenbach, L. and Canaani, E. (2007) Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol Cell Biol*, **27**, 1889-1903.
14. Cho, E.A., Prindle, M.J. and Dressler, G.R. (2003) BRCT domain-containing protein PTIP is essential for progression through mitosis. *Mol Cell Biol*, **23**, 1666-1673.
15. Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D.J., Kitabayashi, I., Herr, W. and Cleary, M.L. (2004) Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol Cell Biol*, **24**, 5639-5649.
16. Lee, J.H. and Skalnik, D.G. (2005) CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J Biol Chem*, **280**, 41725-41731.
17. Goo, Y.H., Sohn, Y.C., Kim, D.H., Kim, S.W., Kang, M.J., Jung, D.J., Kwak, E., Barlev, N.A., Berger, S.L., Chow, V.T. *et al.* (2003) Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. *Mol Cell Biol*, **23**, 140-149.
18. Wu, M., Wang, P.F., Lee, J.S., Martin-Brown, S., Florens, L., Washburn, M. and Shilatifard, A. (2008) Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol Cell Biol*, **28**, 7337-7344.
19. Lee, J.H. and Skalnik, D.G. (2008) Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol Cell Biol*, **28**, 609-618.
20. Lee, J.H., Tate, C.M., You, J.S. and Skalnik, D.G. (2007) Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *J Biol Chem*, **282**, 13419-13428.
21. Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H. and Allis, C.D. (2005) WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*, **121**, 859-872.
22. Dou, Y., Milne, T.A., Ruthenburg, A.J., Lee, S., Lee, J.W., Verdine, G.L., Allis, C.D. and Roeder, R.G. (2006) Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol*, **13**, 713-719.
23. Patel, A., Vought, V.E., Dharmarajan, V. and Cosgrove, M.S. (2008b) A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. *J Biol Chem*, **283**, 32162-32175.
24. Steward, M.M., Lee, J.S., O'Donovan, A., Wyatt, M., Bernstein, B.E. and Shilatifard, A. (2006) Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat Struct Mol Biol*, **13**, 852-854.

25. Chen, Y., Wan, B., Wang, K.C., Cao, F., Yang, Y., Protacio, A., Dou, Y., Chang, H.Y. and Lei, M. (2011) Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding. *EMBO Rep*, **12**, 797-803.
26. Sarvan, S., Avdic, V., Tremblay, V., Chaturvedi, C.P., Zhang, P., Lanouette, S., Blais, A., Brunzelle, J.S., Brand, M. and Couture, J.F. (2011) Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain. *Nat Struct Mol Biol*, **18**, 857-859.
27. Cao, F., Chen, Y., Cierpicki, T., Liu, Y., Basrur, V., Lei, M. and Dou, Y. (2010) An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain. *PLoS One*, **5**, e14102.
28. Odho, Z., Southall, S.M. and Wilson, J.R. (2010) Characterization of a novel WDR5-binding site that recruits RbBP5 through a conserved motif to enhance methylation of histone H3 lysine 4 by mixed lineage leukemia protein-1. *J Biol Chem*, **285**, 32967-32976.
29. Avdic, V., Zhang, P., Lanouette, S., Groulx, A., Tremblay, V., Brunzelle, J. and Couture, J.F. (2011a) Structural and biochemical insights into MLL1 core complex assembly. *Structure*, **19**, 101-108.
30. Song, J.J. and Kingston, R.E. (2008) WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *J Biol Chem*, **283**, 35258-35264.
31. Patel, A., Dharmarajan, V. and Cosgrove, M.S. (2008a) Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. *J Biol Chem*, **283**, 32158-32161.
32. Couture, J.F., Collazo, E., Ortiz-Tello, P.A., Brunzelle, J.S. and Trievel, R.C. (2007) Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat Struct Mol Biol*, **14**, 689-695.
33. Avdic, V., Zhang, P., Lanouette, S., Voronova, A., Skerjanc, I. and Couture, J.F. (2011b) Fine-tuning the stimulation of MLL1 methyltransferase activity by a histone H3-based peptide mimetic. *FASEB J*, **25**, 960-967.
34. Minor, Z.O.a.W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Macromolecular Crystallography, part A*, **276**, 307-326.
35. Couture, J.F., Collazo, E. and Trievel, R.C. (2006) Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat Struct Mol Biol*, **13**, 698-703.
36. Potterton, E., Briggs, P., Turkenburg, M. and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr*, **59**, 1131-1137.
37. Blanc, E., Roversi, P., Vonnrhein, C., Flensburg, C., Lea, S.M. and Bricogne, G. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr*, **60**, 2210-2221.
38. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.
39. Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **66**, 12-21.
40. Han, Z., Guo, L., Wang, H., Shen, Y., Deng, X.W. and Chai, J. (2006) Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. *Mol Cell*, **22**, 137-144.

41. Ruthenburg, A.J., Wang, W., Graybosch, D.M., Li, H., Allis, C.D., Patel, D.J. and Verdine, G.L. (2006) Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nat Struct Mol Biol*, **13**, 704-712.
42. Schuetz, A., Allali-Hassani, A., Martin, F., Loppnau, P., Vedadi, M., Bochkarev, A., Plotnikov, A.N., Arrowsmith, C.H. and Min, J. (2006) Structural basis for molecular recognition and presentation of histone H3 by WDR5. *EMBO J*, **25**, 4245-4252.
43. Hollmann, M., Simmerl, E., Schafer, U. and Schafer, M.A. (2002) The essential *Drosophila melanogaster* gene *wds* (will die slowly) codes for a WD-repeat protein with seven repeats. *Mol Genet Genomics*, **268**, 425-433.
44. Ang, Y.S., Tsai, S.Y., Lee, D.F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B. *et al.* (2011) Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell*, **145**, 183-197.
45. Lee, S., Lee, D.K., Dou, Y., Lee, J., Lee, B., Kwak, E., Kong, Y.Y., Lee, S.K., Roeder, R.G. and Lee, J.W. (2006) Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. *Proc Natl Acad Sci U S A*, **103**, 15392-15397.
46. Lee, J., Kim, D.H., Lee, S., Yang, Q.H., Lee, D.K., Lee, S.K., Roeder, R.G. and Lee, J.W. (2009) A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. *Proc Natl Acad Sci U S A*, **106**, 8513-8518.
47. Gan, Q., Thiebaud, P., Theze, N., Jin, L., Xu, G., Grant, P. and Owens, G.K. (2011) WD repeat-containing protein 5, a ubiquitously expressed histone methyltransferase adaptor protein, regulates smooth muscle cell-selective gene activation through interaction with pituitary homeobox 2. *J Biol Chem*, **286**, 21853-21864.
48. Wang, Y.Y., Liu, L.J., Zhong, B., Liu, T.T., Li, Y., Yang, Y., Ran, Y., Li, S., Tien, P. and Shu, H.B. (2010) WDR5 is essential for assembly of the VISA-associated signaling complex and virus-triggered IRF3 and NF-kappaB activation. *Proc Natl Acad Sci U S A*, **107**, 815-820.
49. Garapaty, S., Xu, C.F., Trojer, P., Mahajan, M.A., Neubert, T.A. and Samuels, H.H. (2009) Identification and characterization of a novel nuclear protein complex involved in nuclear hormone receptor-mediated gene regulation. *J Biol Chem*, **284**, 7542-7552.
50. Cai, Y., Jin, J., Swanson, S.K., Cole, M.D., Choi, S.H., Florens, L., Washburn, M.P., Conaway, J.W. and Conaway, R.C. (2010) Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. *J Biol Chem*, **285**, 4268-4272.

CHAPTER 4. A PHOSPHORYLATION SWITCH ON RBBP5 REGULATES HISTONE H3 LYS4 METHYLATION

The following manuscript is published in *Genes & Development*.

Zhang, P., Chaturvedi, C. P., Tremblay, V., Cramet, M., Brunzelle, J. S., Skiniotis, G., . . . Couture, J. F. (2015). A phosphorylation switch on RbBP5 regulates histone H3 Lys4 methylation. *Genes Dev*, 29(2), 123-128. doi: 10.1101/gad.254870.114

I designed and purified an ASH2L SPRY construct that was amenable for crystallization. Using ITC, I determined the binding affinity between ASH2L SPRY and the RbBP5 344-357 peptide and then co-crystallized this complex (Figure 4.1). Based on the ASH2L-RbBP5 structure, I used ITC to probe the role of several ASH2L residues in RbBP5 binding and performed *in vitro* methyltransferase assays to probe the role of ASH2L-RbBP5 interaction in stimulating MLL1 activity (Figure 4.2AB). I prepared maxipreps of Flag-ASH2L wild-type and mutants for transfection into erythroid cells to probe the role of ASH2L-RbBP5 interaction in erythroid cell terminal differentiation. I performed preliminary immunoprecipitation experiments to probe the role of RbBP5 S350 in ASH2L binding in HEK293 cells (Figure 4.3B). Using ITC, I determined the binding affinity between ASH2L SPRY and the RbBP5 S350phos peptide and co-crystallized this complex (Figure 4.3CDE). Finally, I performed the pull down and enzymatic assays to probe the role of RbBP5 S350phos in MLL3 binding and stimulation of MLL1 and MLL3 activity, respectively (Figure 4.4BCD). I also wrote most of the manuscript.

A phosphorylation switch on RbBP5 regulates histone H3 Lys-4 methylation.

Pamela Zhang¹, Chandra-Prakash Chaturvedi^{3,4}, Veronique Tremblay¹, Myriam Cramet¹, Joseph S. Brunzelle⁶, Georgios Skiniotis⁵, Marjorie Brand^{3,4}, Ali Shilatifard² and Jean-François Couture¹

1) Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, Canada, K1H 8M5.

2) Department of Biochemistry and Molecular Genetics, Northwestern University, Searle Building, 320 E Superior, Chicago, Illinois, USA, 60611.

3) The Sprott Center for Stem Cell Research, Regenerative Medicine Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada, K1H 8L6.

4) Department of Cellular and Molecular Medicine, University of Ottawa, Ontario, Canada, K1H 8L6.

5) Life Sciences Institute and Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan, USA, 48109.

6) Feinberg School of Medicine, Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, Chicago, Illinois, USA, 60611.

Correspondence should be addressed to J.-F.C.

Short title: RbBP5 phosphorylation regulates H3K4 methylation

Corresponding author: Jean-Francois Couture

ABSTRACT

The methyltransferase activity of the trithorax group (TrxG) protein MLL1 found within its COMPASS-like complex is allosterically regulated by a four-subunit complex composed of WDR5, RbBP5, ASH2L and DPY30 (also referred to as WRAD). We report structural evidence showing that in WRAD, a concave surface of the ASH2L SPRY domain binds to a cluster of acidic residues, referred to as the D/E box, in RbBP5. Mutational analysis shows that residues forming the ASH2L/RbBP5 interface are important for heterodimer formation, stimulation of MLL1 catalytic activity and erythroid cell terminal differentiation. We also demonstrate that a phosphorylation switch on RbBP5 stimulates WRAD complex formation and significantly increases KMT2 enzymes' methylation rates. Overall, our findings provide structural insights into the assembly of the WRAD complex and point to a novel regulatory mechanism controlling the activity of the KMT2/COMPASS family of lysine methyltransferases.

INTRODUCTION

In mammals, the KMT2 (Lysine (K) MethylTransferase 2) family of enzymes (C. D. Allis et al., 2007), which includes MLL1-4 and SET1A/1B of the COMPASS family, catalyzes the site-specific methylation of the ϵ -amine of Lys-4 (K4) of histone H3. While these enzymes share the ability to methylate the same residue on histone H3, the catalytic activity of these enzymes is linked to different biological processes. MLL1/MLL2 di/tri-methylate H3K4 (H3K4me_{2/3}) and regulate *Hox* gene expression during embryonic development (Dou et al., 2006; Yu et al., 1995). MLL3/MLL4 regulate adipogenesis (J. Lee et al., 2008) and primarily mono-methylate H3K4 (H3K4me₁) at both enhancer (Herz et al., 2012; Hu et al., 2013a) and promoter regions (J. Cheng et al., 2014) while SET1A/B are the primary H3K4 tri-methyltransferases (M. Wu et al., 2008). However, despite divergence in catalytic activity and functional roles, enzymes of the KMT2/COMPASS family must assemble into multi-subunit complexes to carry out their biological functions.

Our molecular understanding of the protein complexes involved in H3K4 methylation stems from the isolation of COMpLex ASSociated with Set1 (COMPASS) from *Saccharomyces cerevisiae* (Dehe et al., 2006; Krogan et al., 2002; Miller et al., 2001; Roguev et al., 2001). These studies demonstrated that regulatory subunits found within COMPASS and mammalian COMPASS-like complexes play key roles in stabilizing the enzyme and stimulating its methyltransferase activity as well as targeting the protein complex to specific genomic loci (Couture & Skiniotis, 2013). While each of these multi-subunit protein complexes contain unique subunits, each member of the KMT2 family associates with a common set of four evolutionarily conserved regulatory proteins, namely WDR5, RbBP5, ASH2L and DPY30 (WRAD) (Couture & Skiniotis, 2013). The four-subunit complex directly binds the SET domain of KMT2 enzymes and serves as an essential

modulatory platform stimulating the enzymatic activity of each member within this family (Avdic et al., 2011a; Dou et al., 2006; A. Patel et al., 2009; Steward et al., 2006; P. Zhang, Lee, Brunzelle, & Couture, 2012).

In an attempt to understand the structural mechanisms underlying the assembly of the WRAD complex and stimulation of MLL1 methyltransferase activity, several studies have dissected the structure and function of each WRAD subunit (Couture & Skiniotis, 2013). WDR5 is crucial for the structural integrity of the complex and acts as a bridge linking each member of the KMT2 family (Dharmarajan, Lee, Patel, Skalnik, & Cosgrove, 2012; P. Zhang et al., 2012) to the regulatory subunits RbBP5, ASH2L and DPY30 (Avdic et al., 2011a; Odho et al., 2010). In RbBP5, a predicted unstructured region binds to ASH2L and WDR5 and is important for the stimulation of MLL1 methyltransferase activity (Avdic et al., 2011a; Cao et al., 2010). Furthermore, mapping analysis identified that this region of RbBP5 binds to the SP1a and Ryanodine receptor (SPRY) domain of ASH2L (Y. Chen et al., 2012); however, the structural basis underlying the interaction of RbBP5 with ASH2L is unknown.

Here, we report the crystal structure of the ASH2L SPRY domain in complex with RbBP5. We show that ASH2L/RbBP5 binding specificity is conferred by several conserved residues on both ASH2L and RbBP5. Structure-guided mutational analysis reveals that disruption of this binding interface impairs formation of the WRAD complex, stimulation of MLL1 methyltransferase activity and terminal differentiation of erythroid cells. Interestingly, the structure reveals that a phosphorylation switch on RbBP5 stimulates WRAD complex formation and increases methylation of H3K4 by the KMT2 enzymes.

RESULTS AND DISCUSSION

Crystal Structure of ASH2L in Complex with RbBP5

After determining that the ASH2L SPRY domain binds residues 344-357 of RbBP5 (Figure S4.1), we sought to gain structural insights into the interaction between ASH2L and RbBP5 and solved the crystal structure of ASH2L^{SPRYdel} in complex with a peptide corresponding to residues 344-357 of RbBP5 at a resolution of 2.20 Å (Table 4.1). The ASH2L^{SPRYdel} domain adopts a twisted β -sandwich composed of two anti-parallel β -sheets (referred to as A and B). Sheet A is composed of β 2, β 4, β 5, β 6, β 7 and β 11 while sheet B is composed of β 1, β 3, β 8, β 9, β 10 and β 12. The two sheets are linked by several interconnecting loops of varying length that extend out of the β -sandwich fold and the ASH2L^{SPRYdel} domain ends with a short α -helix (α 1) (Figure 4.1A).

Simulated annealing omit maps reveal clear electron density for the RbBP5 peptide including residues 345 to 354 (Figure S4.2A). No electron density is observed for the RbBP5 E344 side chain (single letter denotes RbBP5 residues) and residues 355-357 and therefore not modeled in the structure. The RbBP5 peptide adopts a chair-like conformation and sits on a shallow surface formed by β 4- β 5- β 6- β 7 of sheet A. The N-terminal half of the peptide (residues 344-348) adopts an elongated conformation and protrudes perpendicularly down towards the basic surface of the ASH2L SPRY domain (Figure 4.1A and B). In this region of the peptide, the RbBP5 E347 side chain makes van der Waals contacts with the backbone of ASH2L residues forming the β 1- β 2 loop while the R348 side chain is solvent-exposed. In stark contrast, the E349 side chain binds in a deep pocket formed by the side chains of Tyr313 and Arg367 (Figure 4.1A and C). The main chain carbonyl of E349 makes a hydrogen bond with the ASH2L Tyr313 hydroxyl group while its carboxylate group engages in several hydrogen bonds with the guanidinium group of Arg367. Located in the bulge of

the S-shaped conformation, the F352 phenyl side chain makes hydrophobic contacts with Tyr313, Pro356 and Tyr359 side chains. Similar to E349, the D353 carboxylate group makes two hydrogen bonds with the Arg343 guanidinium group, suggesting that the ASH2L^{SPRYdel} positively charged cleft is important for binding this region predominantly occupied by glutamic acid and aspartic acid residues (subsequently referred to as the D/E box) of RbBP5 (Figure 4.1B and C).

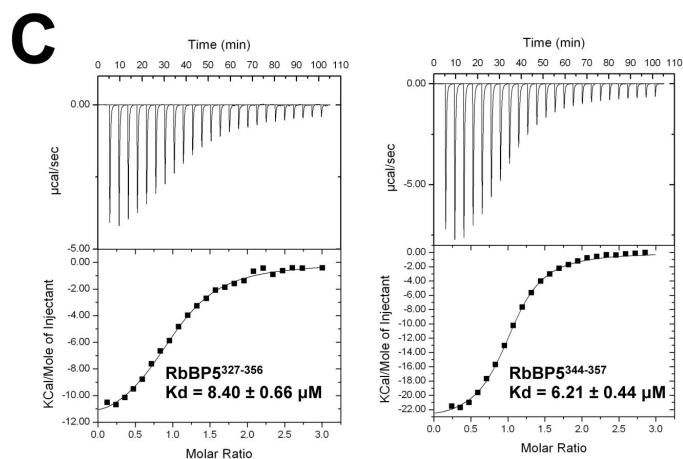
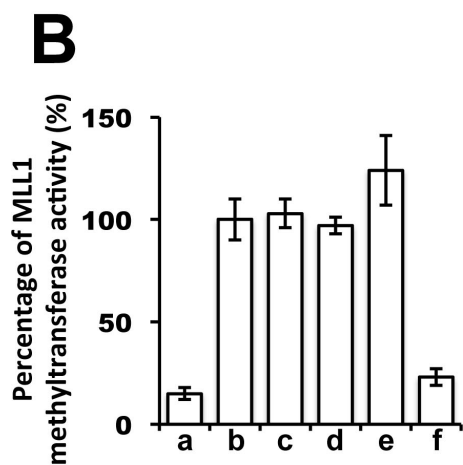
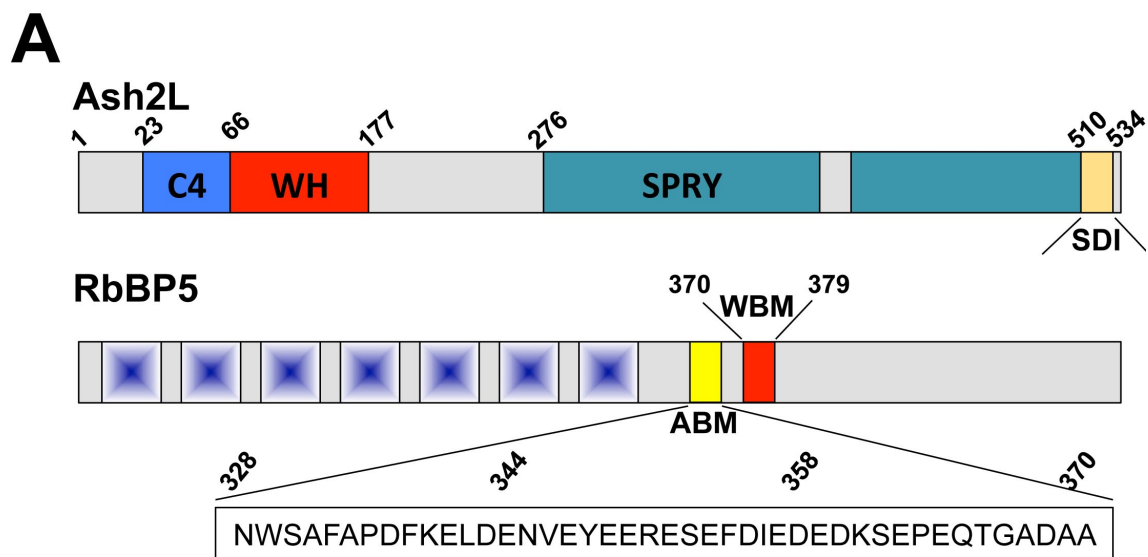


Figure S4.1. Identification of the ASH2L-binding motif (ABM).

A) Schematic representation of ASH2L and RbBP5 domains. ASH2L C4 zinc finger, winged-helix (WH) and SPIA and Ryanodine receptor (SPRY) domains as well as the Sdc DPY30 interacting region (SDI) are illustrated. RbBP5 WD repeat proteins are illustrated as blue squares while the ASH2L- and WDR5-binding motifs (ABM and WBM) are highlighted.

B) The RbBP5 ABM motif is important for binding to ASH2L and stimulation of MLL1 methyltransferase activity. Methyltransferase assays performed in the absence (a) or presence of full-length RbBP5 (b), RbBP5³³⁰⁻³⁸⁰ (c), RbBP5³³⁰⁻³⁷⁰ (d), RbBP5³³⁰⁻³⁶⁰ (e) and RbBP5³⁴⁴⁻³⁵⁷ (f). Assays were repeated twice with two independent preparations of each protein and performed in triplicates.

C) RbBP5 residues 344-357 bind to ASH2L but are insufficient to stimulate MLL1 methyltransferase activity. Equilibrium dissociation constants (K_d) determined using isothermal titration calorimetry experiments. Each K_d represents an average of two replicates performed with different batches of protein.

Table 4.1. Data Collection and Refinement Statistics for ASH2L SPRY in Complex with RbBP5 Peptides		
	RbBP5	RbBP5 S350^{Phos}
Data Collection		
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions		
a, b, c (Å)	51.5, 59.1, 70.8	38.3, 51.5, 110.2
Resolution	29.6 – 2.10 (2.2 – 2.1) ^a	29.9 – 2.1 (2.1 – 2.1)
R _{meas}	5.0 (12.5)	6.9 (16.2)
I / σI	16.5 (7.6)	11.1 (3.9)
Completeness (%)	96.6 (87.1)	98.6 (99.9)
Redundancy	3.4 (3.2)	3.5 (3.4)
Refinement		
Resolution (Å)	29.2 – 2.2	29.9 – 2.1
No. reflections	11232	13897
R _{work} / R _{free}	21.1 / 26.4	23.2 / 26.6
No. Atoms		
Protein	1377	1389
Peptides	94	75
Water	159	176
B-factors (Å²)		
Protein	16.4	16.2
Ligands	17.2	21.0
Water	24.0	24.4
Rmsds		
Bond lengths (Å)	0.008	0.008
Bond angles (°)	1.0	1.0
Molprobability scores	1.19	1.60
Ramachandran favored (%)	96.2	95.5
Ramachandran allowed (%)	3.8	4.5

^a Values in () are for the last resolution shell.

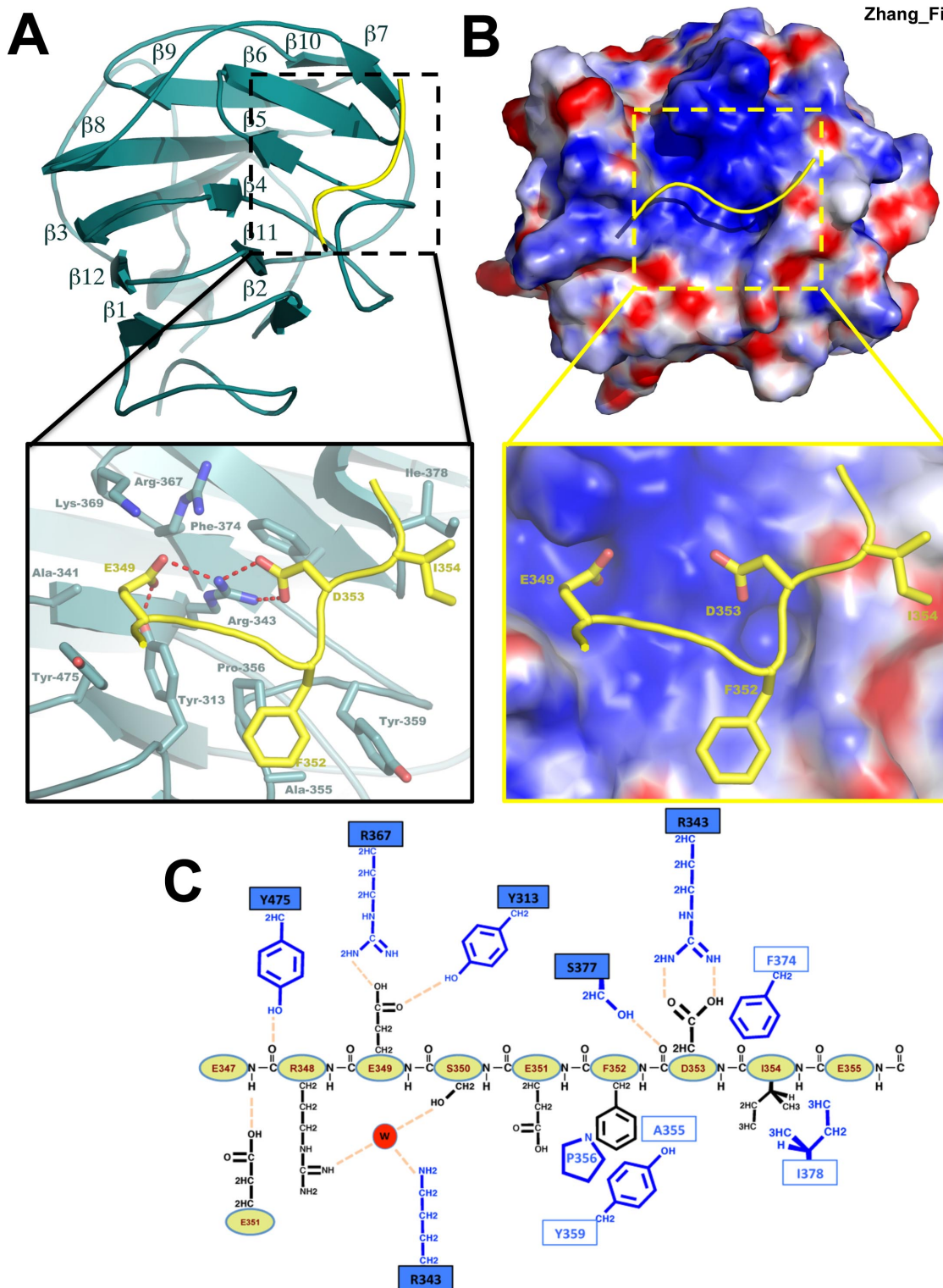


Figure 4.1. ASH2L SPRY domain binds a D/E box on RbBP5.

A) Cartoon representation of the ASH2L SPRY domain (green) in complex with RbBP5 (yellow) and a zoomed view on the interactions between the ASH2L SPRY domain and RbBP5. ASH2L and RbBP5 carbon atoms are highlighted in light green and yellow, respectively. Key hydrogen bonds are rendered as red dashed lines. For clarity, only a subset of interactions is shown.

B) Electrostatic potentials are contoured from -10 kbTe^{-1} (red) to $+10 \text{ kbTe}^{-1}$ (blue) (e = charge of an electron, kb = Boltzmann's constant and T = temperature in Kelvin) and zoomed view on the positively charged cleft of ASH2L.

C) Schematic representation of the interactions stabilizing RbBP5 into the ASH2L SPRY peptide-binding pocket. Yellow spheres represent RbBP5 residues. ASH2L residues making hydrogen bonds (filled boxes), hydrophobic or van der Waals (empty boxes) contacts with RbBP5 are rendered in blue. Hydrogen bonds are highlighted as orange dashed lines. For clarity, some interactions were omitted from the figure.

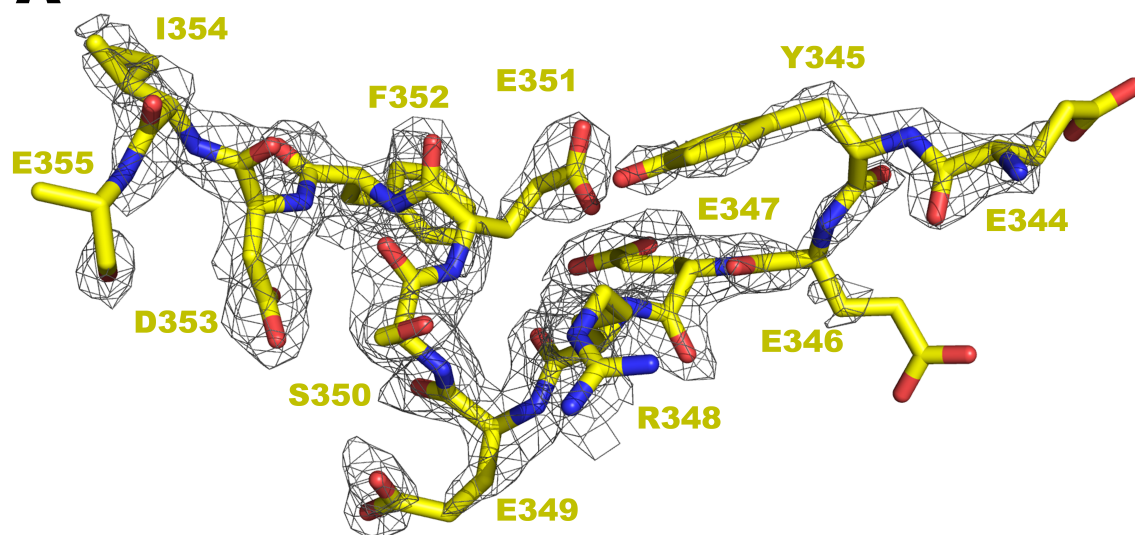
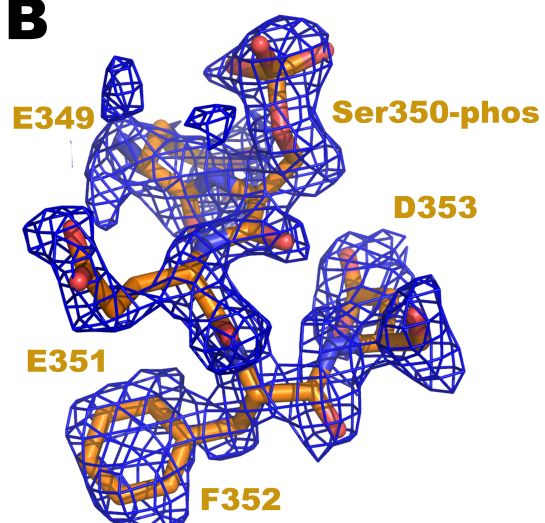
A**B**

Figure S4.2. Simulated Annealing map of RbBP5 peptides.

Omit maps were generated using Buster and ASH2L coordinates in which RbBP5 (A) or RbBP5^{phos} (B) peptides were deleted. The maps are contoured at 2σ and highlighted in black (A) or blue (B). Carbon atoms of RbBP5 (A) and RbBP5^{phos} (B) peptides are rendered in yellow and orange, respectively.

Disruption of ASH2L/RbBP5 Interaction Impairs MLL1 Enzymatic Stimulation and Delays Erythroid Cell Terminal Differentiation

Following structural analysis of the ASH2L/RbBP5 complex, we first sought to identify ASH2L residues that are key for binding to RbBP5. Using ITC (Figure 4.2A), we found that replacement of Tyr313 and Arg343, two residues lining the base of the ASH2L^{SPRYdel} D/E-binding pocket and interacting with RbBP5 E347 and D353, respectively, with alanine severely impaired binding of RbBP5. Accordingly, enzymatic assays performed with the same mutants resulted in approximately five-fold reduction of MLL1 methyltransferase activity compared to wild-type (WT) ASH2L (Figures 4.2B and S4.3B). Mutation of Pro356 and Arg367, residues interacting with the hydrophobic bulge and E349 of the RbBP5 D/E box, resulted in six- and 13-fold reduction in binding, respectively. Accordingly, reconstitution of the complex with the ASH2L Pro356Ala and Arg367Ala mutants failed to stimulate MLL1 methyltransferase activity to the same extent as WT ASH2L, demonstrating that an ASH2L positively charged pocket lined by hydrophobic residues is important for WRAD assembly and MLL1 methyltransferase activity (Figure 4.2A and B).

ASH2L is essential for maintaining high levels of histone H3K4 tri-methylation (Demers et al., 2007; Steward et al., 2006) and knockdown of ASH2L in murine erythroid leukemia (MEL) cells results in a decrease of the H3K4me3 mark at the hypersensitive site 2 (HS2) of the *β-globin* locus control region (LCR) and a concomitant loss of *β-globin* gene transcription, a marker of erythroid cell terminal differentiation (Demers et al., 2007). To test the impact of mutations impairing ASH2L/RbBP5 complex formation, we transfected FLAG-tagged constructs corresponding to ASH2L WT and single point mutant of residues forming the base of the RbBP5-binding pocket in MEL cells stably expressing a doxycycline

(Dox)-inducible small hairpin shRNA directed against ASH2L (Demers et al., 2007). Treatment of the cells with Dox resulted in a 40% decrease of H3K4me3 at the HS2 locus and a corresponding loss of 50% in *β-globin* gene expression (Figure 4.2C and D). Transfection of MEL cells with sh-resistant FLAG-ASH2L^{WT} restored H3K4me3 and transcription of the *β-globin* gene to that of wild-type level. Consistent with our binding and methyltransferase assays, FLAG-ASH2L^{Arg343Ala} and FLAG-ASH2L^{Pro356Ala} mutants failed to maintain maximal expression of the *β-globin* gene and rescue the loss of H3K4me3 (Figure 4.2C and D). Correlatively, transfection of FLAG-ASH2L^{Tyr359Val}, a mutant that exhibited activity similar to ASH2L^{WT}, restored H3K4me3 and *β-globin* gene expression levels similar to ASH2L^{WT}. Together, our findings strongly suggest that a functional ASH2L/RbBP5 heterodimer is pivotal for maintaining the differentiation potential of MEL cells.

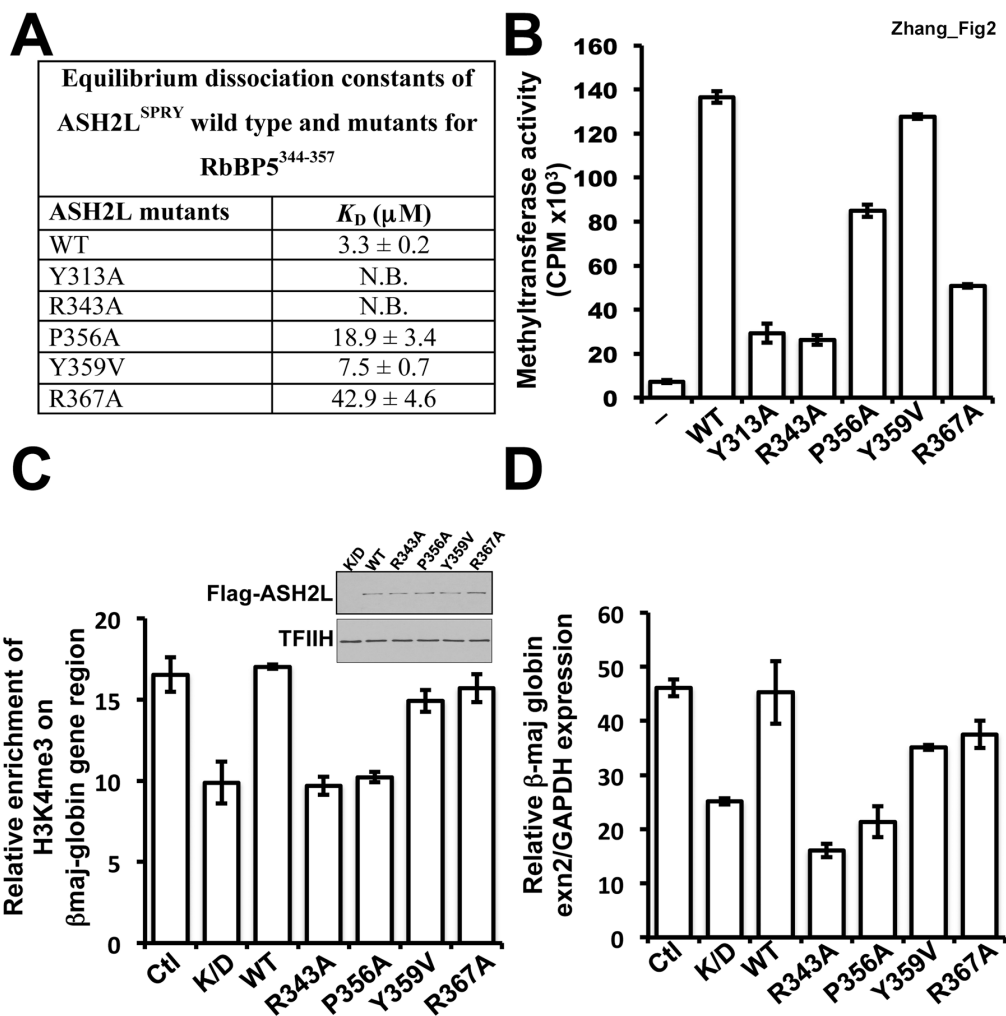


Figure 4.2. Interaction between ASH2L and RbBP5 is essential for terminal differentiation of erythroid cells.

A) Dissociation constants determined using ITC.

B) Methyltransferase assays performed with MLL1 3762-3969 alone (-), or in the presence of wild-type ASH2L (+) or the indicated mutants.

C) Mutation of ASH2L SPRY surface residues prevents maximal H3K4 tri-methylation at the *β-globin* LCR. Enrichment of H3K4me3 was measured by ChIP as previously described (Sarvan et al., 2011) with either the empty vector (K/D), or constructs corresponding to ASH2L WT or the ASH2L R343A, P356A, Y359V or R367A mutant. The inset illustrates a western blot of endogenous ASH2L knockdown and rescue with shRNA-resistant FLAG-tagged ASH2L WT or mutants in differentiated MEL cells in which TFIIF p89 is used as a loading control.

D) Interactions between ASH2L and RbBP5 are important for *β-globin* gene expression. Transcription of the *β-major globin* gene (*βmaj-globin*) vs glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was assessed using qRT-PCR as previously described (Demers et al., 2007).

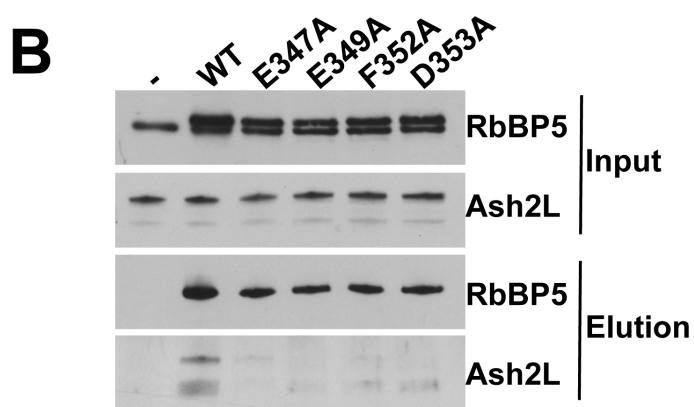
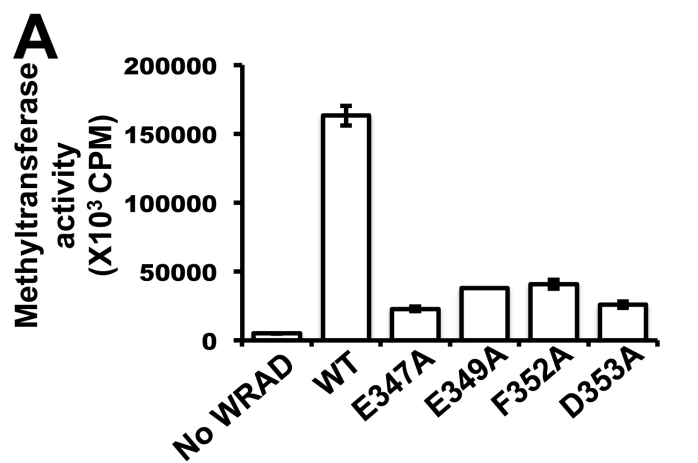


Figure S4.3. RbBP5 D/E box region is important for the stimulation of the SET1 family of lysine methyltransferases.

A) The D/E box within RbBP5 is important for the stimulation of MLL1 enzymatic activity. Methyltransferase assays showing the impact of ABM mutation. Error bars represent the standard deviation of three experiments performed in triplicate with three independent batches of RbBP5 wild-type or mutants.

B) Mutation of the D/E box impairs ASH2L/RbBP5 complex formation. Co-immunoprecipitation of ASH2L by FLAG-tagged RbBP5 wild-type or mutants were analyzed by immunoblots with the indicated antibodies.

Phosphorylation of RbBP5 on S350 Potentiates WRAD Assembly

MLL1 is tightly regulated by various mechanisms including allosteric regulation by the WRAD complex (Dou et al., 2006), deposition of other post-translational modifications on histone proteins (Southall et al., 2009) and phosphorylation of MLL1 by ATR (Liu et al., 2010). In the RbBP5 D/E box, an evolutionarily conserved serine residue (S350) is found in the center of the ASH2L SPRY concave surface (Figure 4.3A). Interestingly, three independent studies revealed that RbBP5 S350 is phosphorylated *in vivo* (Christensen et al., 2010; Phanstiel et al., 2011; Shiromizu et al., 2013). To determine the impact of RbBP5 phosphorylation on WRAD formation, we ectopically expressed constructs corresponding to either wild-type RbBP5 or an RbBP5 S350A mutant in fusion with a FLAG tag in HEK293 cells. While we observed enrichment of ASH2L following immunoprecipitation of wild-type FLAG-RbBP5, incubation of FLAG-RbBP5 S350A with M2 agarose beads failed to co-immunoprecipitate ASH2L (Figure 4.3B). Considering that S350 does not make significant interactions with ASH2L (Figure 4.3C) and that its substitution to alanine impairs WRAD assembly suggest that keeping the hydroxyl group on S350 is essential for high affinity interaction between ASH2L and RbBP5. We next used ITC to determine the impact of S350 phosphorylation on the binding of RbBP5 to ASH2L and found that the phosphorylated peptide RbBP5³⁴⁴⁻³⁵⁷ bound to ASH2L^{SPRYdel} with 15-fold higher affinity (Figure 4.3D), strongly suggesting that the ASH2L SPRY domain is a novel phospho-reader domain.

To understand the structural basis underlying the binding preference of ASH2L to RbBP5^{phos}, we solved the crystal structure of the ASH2L/RbBP5^{phos} complex. The ASH2L/RbBP5^{phos} complex aligns with the ASH2L/RbBP5 complex with a root mean square deviation of 0.192Å, suggesting that binding of RbBP5^{phos} does not induce large structural reorganization of the ASH2L SPRY domain compared to the unmodified complex.

However, the phosphate moiety displaces the Lys369 side chain of ASH2L to accommodate short water-mediated hydrogen bonds with the phosphate group (Figure 4.3E), demonstrating the ability of the ASH2L SPRY domain to read the phosphorylated form of RbBP5.

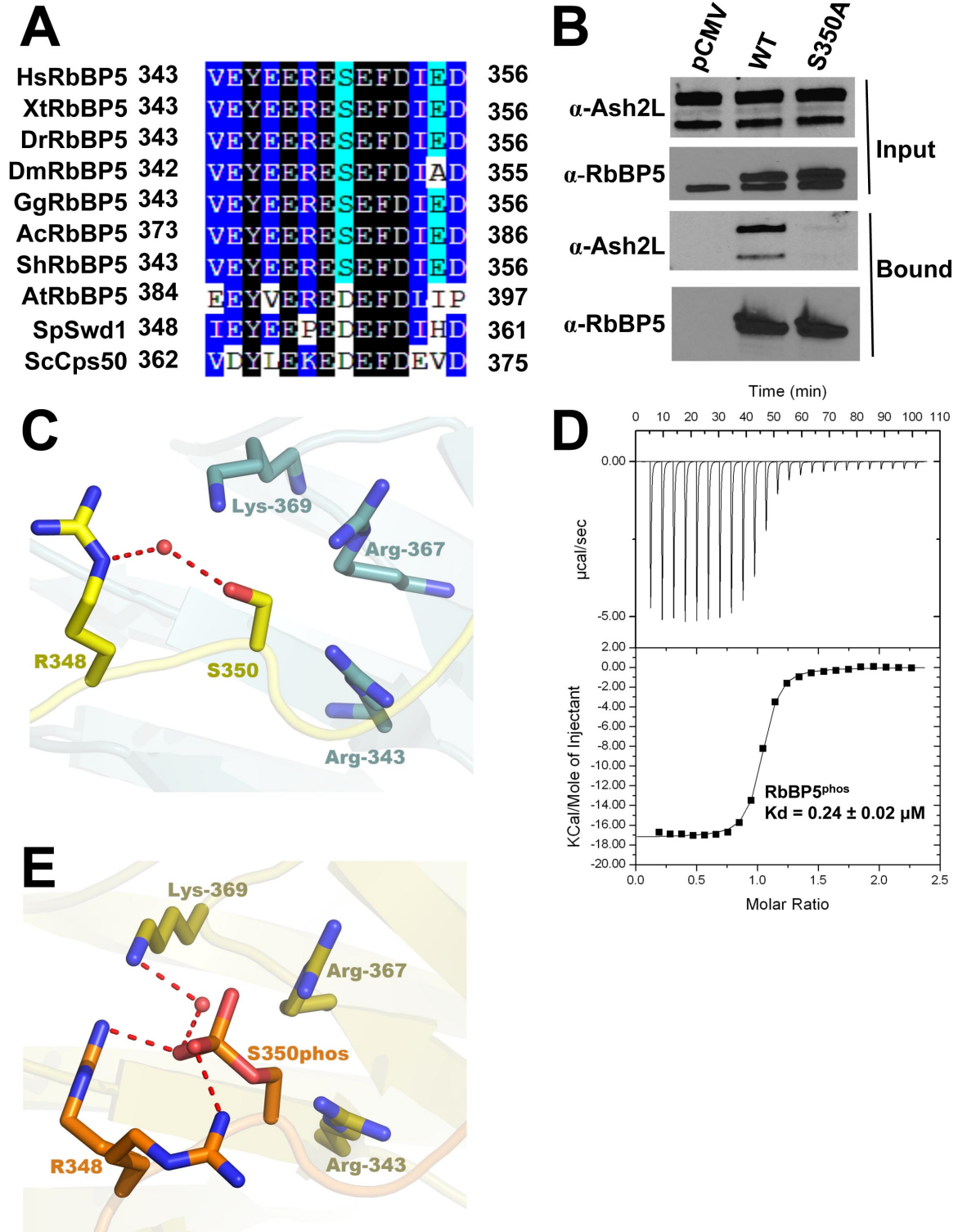


Figure 4.3. Phosphorylation of RbBP5 stimulates WRAD complex formation.

A) RbBP5 D/E box is evolutionarily conserved. Sequence alignment of RbBP5 D/E box from *Homo sapiens* (Hs), *Xenopus tropicalis* (Xt), *Dario rerio* (Dr), *Drosophila melanogaster* (Dm), *Gallus gallus* (Gg), *Anolis carolinensis* (Ac), *Sarcophilus harrisii* (Sh), *Arabidopsis thaliana* (At), *Schizosaccharomyces pombe* (Sp) and *Saccharomyces cerevisiae* (Sc). Positions with 100%, 99-75% and less than 75% of amino acid conservation are represented in black, blue and cyan, respectively.

B) Replacement of S350 to alanine decreases the interaction between RbBP5 and ASH2L. Immunoprecipitation of ectopically expressed FLAG-tagged constructs of RbBP5 wild-type and S350A with M2 agarose beads. RbBP5 and ASH2L were detected with the indicated antibodies.

C) Zoomed view on RbBP5 S350. Residues are colored as in Figure 4.1.

D) Phosphorylated RbBP5 binds ASH2L with higher affinity. Representative ITC experiment of RbBP5^{phos} binding to ASH2L. Data is shown as in Figure S4.1C.

E) Crystal structure of ASH2L in complex with RbBP5^{phos}. Zoomed view of phosphorylated S350 in which RbBP5^{phos} and ASH2L carbon atoms are rendered in orange and dark yellow, respectively. Hydrogen bonds are illustrated as in Figure 4.1A.

RbBP5 Phosphorylation: a Novel Regulatory Switch Controlling WRAD Assembly

With prior studies showing that the ASH2L C4-Winged-Helix (C4-WH) domain is important for binding to DNA (Y. Chen et al., 2011; Sarvan et al., 2011) and ubiquitin (L. Wu et al., 2013) and that its SDI motif is important for binding to DPY30 (Y. Chen et al., 2012; South, Fingerman, Mersman, Du, & Briggs, 2010), our results point to a model whereby ASH2L acts as a modulatory platform enabling the integration of a cascade of binding events that ultimately lead to the precise regulation of KMT2 methyltransferase activity. Herein we report that ASH2L also recognizes the phosphorylated form of RbBP5. Binding and structural studies show that the ASH2L SPRY domain binds RbBP5^{phos} with 15-fold more affinity and that the phosphate moiety induces local structural reorganization within ASH2L, suggesting that the ASH2L SPRY domain is a novel phospho-binding domain. However, the recognition of the phosphate moiety by ASH2L differs from other known phospho-readers. This is particularly apparent for 14-3-3 proteins, which engage in several electrostatic interactions with the phosphate moiety within a well-defined basic pocket (Rittinger et al., 1999). Consistently, Muslin et al. showed that 14-3-3 can only bind to a Ser259-phosphorylated form of Raf-1 peptide (Muslin, Tanner, Allen, & Shaw, 1996). Our observations that ASH2L engages in a relatively small number of contacts with the phosphate moiety of S350 and that it binds to both the unmodified and phosphorylated forms of RbBP5 suggest that this mode of phosphopeptide recognition serves as a rheostat modulating formation of the WRAD complex rather than an on/off switch assigned to other canonical phospho-readers.

RbBP5 Phosphorylation Controls Histone H3K4 Methylation by the KMT2 Enzymes

Our studies revealed that RbBP5 phosphorylation creates a better epitope for the binding of the ASH2L SPRY domain. However, close inspection of the structure revealed that the RbBP5 phosphate moiety is not entirely buried within the SPRY concave surface (Figure 4.4A), suggesting that it may potentially play a direct role in regulating the methyltransferase activity of the KMT2 enzymes. To address this question, we performed pull-down experiments with His-SUMO-tagged MLL3 bound to TALON beads and ASH2L/RbBP5 or ASH2L/RbBP5^{phos}. Following several washes, TALON-bound protein complexes were eluted with sample loading buffer, resolved on SDS-PAGE and stained with coomassie. Consistent with recent binding studies (Cao et al., 2010), we observed binding of the ASH2L/RbBP5 heterodimer to the MLL3 SET domain. Interestingly, a 5-fold increase in binding was observed when the ASH2L/RbBP5^{phos} complex was incubated with His-SUMO-MLL3 (Figure 4.4B), suggesting that the ASH2L/RbBP5^{phos} dimer serves as a better interacting platform for the binding of the MLL3 SET domain.

Based on these observations, we surmised that ASH2L/RbBP5^{phos} might modulate the methyltransferase activity of the KMT2 enzymes. To confirm this hypothesis, enzymatic assays were performed with different concentrations of the MLL3 SET domain incubated with stoichiometric amounts of ASH2L/RbBP5 or ASH2L/RbBP5^{phos}. As shown in Figure 4.4C and consistent with previous studies (P. Zhang et al., 2012), both complexes stimulated MLL3 methyltransferase activity at 1 μ M. However, upon dilution of the complex, ASH2L/RbBP5 failed to stimulate the activity of MLL3 while ASH2L/RbBP5^{phos} retained full activity of MLL3, demonstrating that RbBP5 phosphorylation serves as a rheostat increasing MLL3 kinetics.

After determining the impact of RbBP5 phosphorylation on MLL3 kinetics, we

sought to determine the degree of K4 methylation catalyzed by MLL1 and MLL3 in the presence of the ASH2L/RbBP5 heterodimer reconstituted with RbBP5 or RbBP5^{phos}. We conducted enzymatic assays and subjected aliquots of the reactions to electrospray ionization mass spectrometry (ESI-MS). In comparison with the control reactions (Figures S4.4 and 4.4D), a shift in the mass/charge (m/z) ratio from 2346 to 2360 is measured for MLL1 and MLL3 in the presence of the ASH2L/RbBP5 heterodimer, corresponding to the transfer of a single methyl group to the ϵ -amine of K4. However, in contrast to the assays performed with unmodified RbBP5, we observed a sharp increase in H3K4me1 when the assays were performed with the ASH2L/RbBP5 heterodimer reconstituted with RbBP5^{phos} (Figure 4.4D). Time course of the methylation reactions followed by ESI-MS further showed that the MLL3/ASH2L/RbBP5^{phos} robustly methylates a histone H3 peptide when compared to MLL3 incubated with the unphosphorylated ASH2L/RbBP5 heterodimer (Figure 4.4D). Interestingly, we also observed detectable levels of H3K4me2 for both MLL1 and MLL3 (Figures S4.4 and 4.4D), suggesting that the enhancement of MLL3 catalytic activity, a predominant histone H3K4 mono-methyltransferase, by the ASH2L/RbBP5^{phos} complex pushes the reaction further to observe H3K4me2. Intriguingly, methyltransferase assays performed with a higher concentration of MLL3 reconstituted with the ASH2L/RbBP5 or ASH2L/RbBP5^{phos} showed that both complexes efficiently tri-methylate H3K4 but failed to show increased rates of di- and tri-methylation of histone H3K4 by the MLL3/ASH2L/RbBP5^{phos} complex (Figure S4.5). Overall, our observations strongly suggest that RbBP5 phosphorylation couples the assembly of the WRAD complex to the allosteric regulation of KMT2 enzymes.

Enzymatic assays revealed that MLL3 mono-methylates H3K4 in the presence of ASH2L/RbBP5 reconstituted with unmodified RbBP5. These observations are consistent

with recent studies showing that COMPASS-like MLL3/MLL4 complexes predominantly mono-methylate H3K4 at enhancer regions and specific promoter regions (J. Cheng et al., 2014; Herz et al., 2012; Hu et al., 2013a; Morgan & Shilatifard, 2013). Interestingly, upon incubation of the MLL3 SET domain with the ASH2L/RbBP5 complex reconstituted with RbBP5^{phos}, peaks corresponding to H3K4me1 and H3K4me2 were observed. In addition, a peak corresponding to H3K4me3 was also observed when higher concentration experiments were performed. These observations are also consistent with recent studies showing that deletion of MLL3 in NIH3T3-L1 cells results in a significant loss of H3K4me3 at the promoter region of the adipogenic marker gene *aP2* (J. Lee et al., 2008). Moreover, B-cell specific knockout of PTIP, a subunit associating with MLL3/MLL4 complexes (Y. W. Cho et al., 2007; Issaeva et al., 2007), results in a loss of H3K4me3 at specific Igh switch regions upon LPS stimulation (Daniel et al., 2010). These seemingly contrasting results potentially point to a model whereby RbBP5 phosphorylation could act as a switch increasing MLL3 kinetics facilitating formation of H3K4me1 that could potentially be further methylated to ultimately form H3K4me2/3. Analogous to the differences in activity between members of the KMT2 family of enzyme, our observations suggest that at least two populations of WRAD complex exist in cells tailored to performed distinct functions.

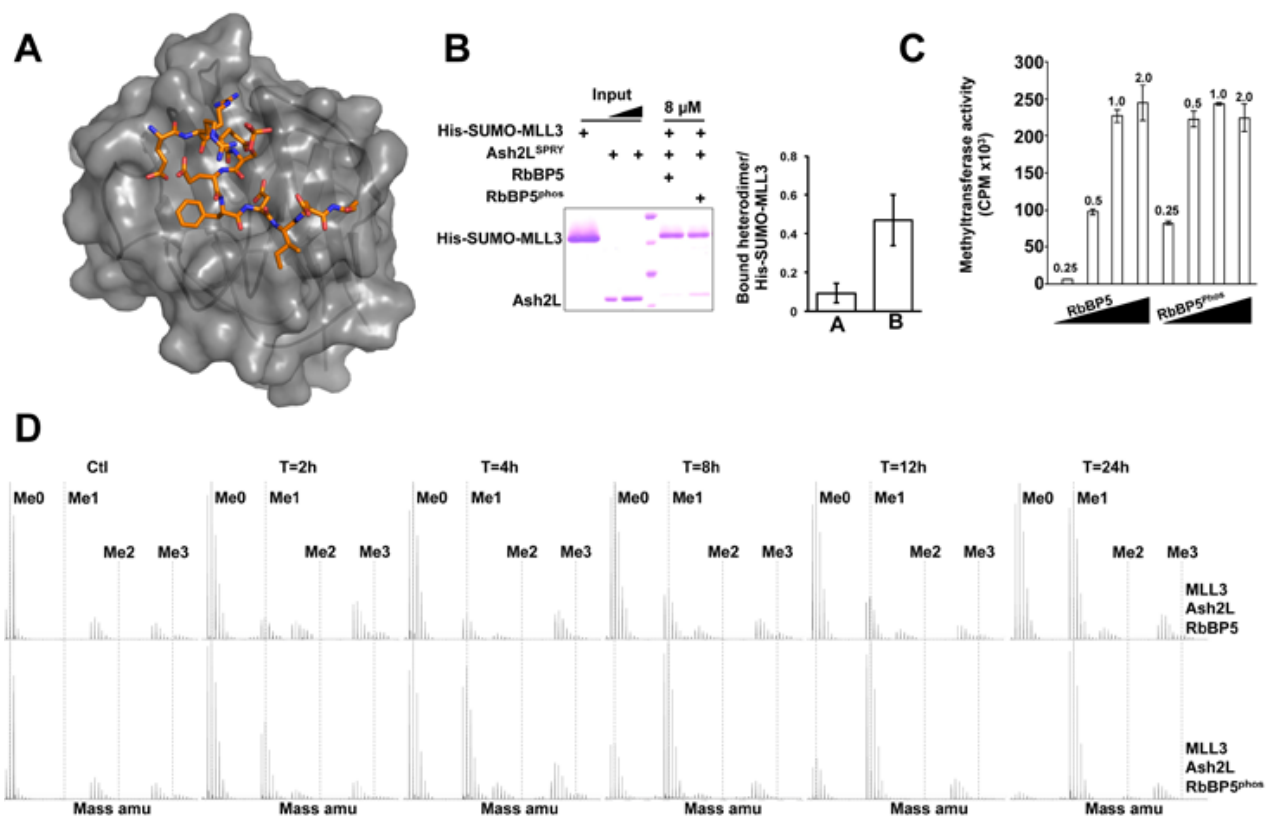


Figure 4.4. RbBP5 S350 phosphorylation switches MLL3 from a H3K4 mono-methyltransferase to a di/tri-methyltransferase.

A) Surface representation of the ASH2L SPRY domain in complex with RbBP5^{phos}. The ASH2L surface is highlighted in grey and RbBP5 is colored as in Figure 4.3E.

B) Pull-down assays of the ASH2L/RbBP5 or ASH2L/RbBP5^{phos} complexes by the MLL3 SET domain. Bound proteins were separated on SDS-PAGE and detected by coomassie staining. A representative coomassie stained SDS-PAGE gel is shown on the left and the quantified mean of bound ASH2L/RbBP5 (A) or ASH2L/RbBP5^{phos} (B) complexes normalized to MLL3 is shown on the right (n = 3 experiments; p < 0.05).

C) Methyltransferase assays were performed with increasing amount (indicated at the top of each graph bar (μM)) of MLL3 and ASH2L/RbBP5 or ASH2L/RbBP5^{phos}. Assays were performed as in Figure S4.1B.

D) Representative spectra of ESI-MS experiments performed with MLL3 incubated with ASH2L/RbBP5 (top) or ASH2L/RbBP5^{phos} (bottom) complexes. Duration of the experiments is indicated on top of each panel.

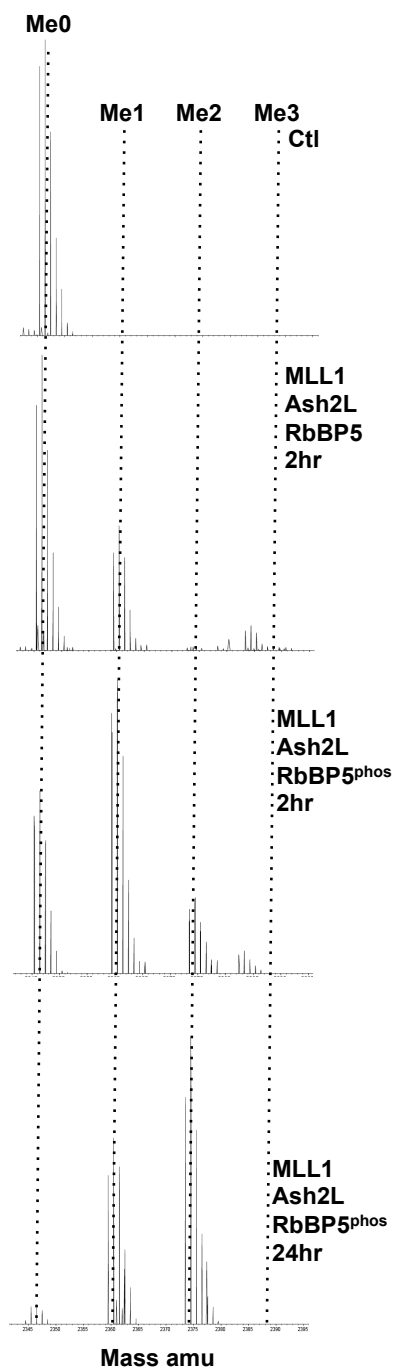


Figure S4.4. RbBP5 S350 phosphorylation increases the rate of MLL1 methylation.

Electrospray ionization mass spectrometry (ESI-MS) experiments performed with 10 μ M of MLL1/ASH2L/RbBP5 vs MLL1/ASH2L/RbBP5^{Phos} complexes. The top panel is the unreacted peptide while the other panels correspond to the methylation reactions performed with MLL1/ASH2L/RbBP5 or MLL1/ASH2L/RbBP5^{Phos} for 2 or 24 hours.

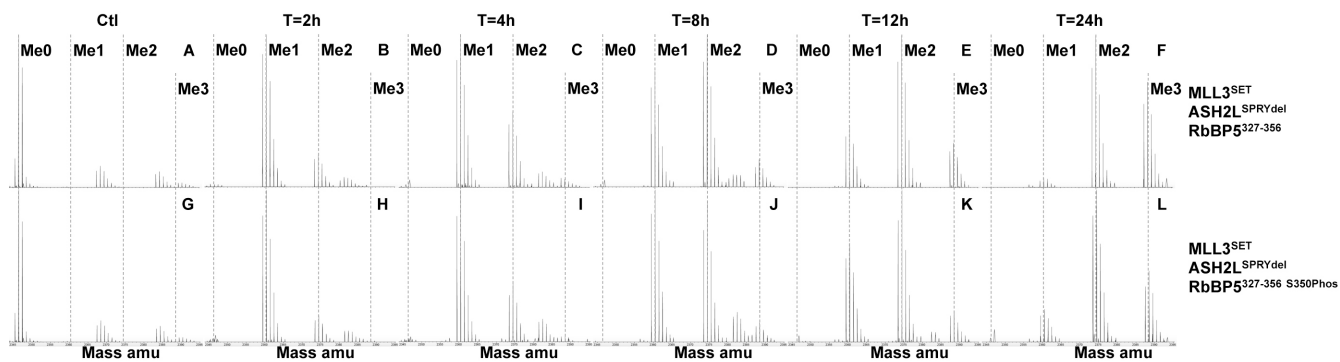


Figure S4.5. RbBP5/ASH2L/MLL3 and RbBP5^{phos}/ASH2L/MLL3 complex display the same kinetics for H3K4 di- and tri-methylation.

Representative spectra of ESI-MS experiments performed with 10 μ M the MLL3 SET domain co-incubated with ASH2L/RbBP5 (top) or ASH2L/RbBP5^{phos} (bottom) complexes. Duration of the experiments is indicated on top of the panels.

MATERIALS AND METHODS

Protein Crystallization and Structure Determination

Recombinantly purified ASH2L^{SPRYdel} (5-10 mg/ml) (see supplementary material and methods) was incubated with equimolar amounts of RbBP5 344-357 on ice for one hour and crystallized using the sitting-drop vapour diffusion method at 18°C. Diffraction quality crystals were obtained in 0.2 M magnesium chloride hexahydrate, 0.1 M BIS-TRIS pH 5.5 and 25% w/v polyethylene glycol. The crystals were sequentially soaked in the mother liquor supplemented with an increasing amount (5-20%) of glycerol, harvested and flash frozen in liquid nitrogen. The structure was solved by molecular replacement and model building was performed as detailed in supplementary material and methods.

Isothermal Titration Calorimetry, *In Vitro* Methyltransferase Assays and Electrospray Ionization Mass Spectrometry (ESI-MS)

Isothermal titration calorimetry (ITC) experiments and enzymatic assays were performed as previously described (P. Zhang et al., 2012). ESI-MS analysis was performed at the SPARC BioCentre using a QSTAR Elite and detailed in supplementary material and methods.

MEL Cells

Mouse Erythroleukemia (MEL) cells were transfected with plasmids expressing FLAG-only, FLAG-ASH2L WT, FLAG-ASH2L Y313A, R343A, P356A, Y359V and R367A by electroporation. 12 hours post transfection, differentiation was induced with DMSO as previously described (Demers et al., 2007). After two days, cells were pelleted by centrifugation, resuspended and cross-linked as previously described (Demers et al., 2007). Chromatin extraction and immunoprecipitation experiments were performed as previously

described (Sarvan et al., 2011) and quantified as detailed in supplementary material and methods.

ACKNOWLEDGEMENTS

Pamela Zhang is supported by a Canadian Institutes of Health Research (CIHR) Banting and Best scholarship. Dr. Jean-Francois Couture is supported by a CIHR grant (MOP-136816). This study was also supported by grants from CIHR to Marjorie Brand (MOP-89834), NIH to Dr. Shilatifard (R01GM069905). Georgios Skiniotis acknowledges the support from the Pew Scholars Program in Biomedical Sciences.

REFERENCES

- Allis CD, Berger SL, Cote J, Dent S, Jenuwien T, Kouzarides T, Pillus L, Reinberg D, Shi Y, Shiekhata R et al. 2007. New nomenclature for chromatin-modifying enzymes. *Cell* **131**: 633-636.
- Avdic V, Zhang P, Lanouette S, Groulx A, Tremblay V, Brunzelle J, Couture JF. 2011. Structural and biochemical insights into MLL1 core complex assembly. *Structure* **19**: 101-108.
- Cao F, Chen Y, Cierpicki T, Liu Y, Basrur V, Lei M, Dou Y. 2010. An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain. *PLoS One* **5**: e14102.
- Chen Y, Cao F, Wan B, Dou Y, Lei M. 2012. Structure of the SPRY domain of human Ash2L and its interactions with RbBP5 and DPY30. *Cell Res* **22**: 598-602.
- Chen Y, Wan B, Wang KC, Cao F, Yang Y, Protacio A, Dou Y, Chang HY, Lei M. 2011. Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding. *EMBO Rep* **12**: 797-803.
- Cheng J, Blum R, Bowman C, Hu D, Shilatifard A, Shen S, Dynlacht BD. 2014. A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. *Mol Cell* **53**: 979-992.
- Cho YW, Hong T, Hong S, Guo H, Yu H, Kim D, Guszczynski T, Dressler GR, Copeland TD, Kalkum M et al. 2007. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem* **282**: 20395-20406.
- Christensen GL, Kelstrup CD, Lyngso C, Sarwar U, Bogebo R, Sheikh SP, Gammeltoft S, Olsen JV, Hansen JL. 2010. Quantitative phosphoproteomics dissection of seven-transmembrane receptor signaling using full and biased agonists. *Molecular & cellular proteomics : MCP* **9**: 1540-1553.
- Couture JF, Skiniotis G. 2013. Assembling a COMPASS. *Epigenetics* **8**.

- Daniel JA, Santos MA, Wang Z, Zang C, Schwab KR, Jankovic M, Filsuf D, Chen HT, Gazumyan A, Yamane A et al. 2010. PTIP promotes chromatin changes critical for immunoglobulin class switch recombination. *Science* **329**: 917-923.
- Dehe PM, Dichtl B, Schaft D, Roguev A, Pamblanco M, Lebrun R, Rodriguez-Gil A, Mkandawire M, Landsberg K, Shevchenko A et al. 2006. Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J Biol Chem* **281**: 35404-35412.
- Demers C, Chaturvedi CP, Ranish JA, Juban G, Lai P, Morle F, Aebersold R, Dilworth FJ, Groudine M, Brand M. 2007. Activator-mediated recruitment of the MLL2 methyltransferase complex to the beta-globin locus. *Mol Cell* **27**: 573-584.
- Dharmarajan V, Lee JH, Patel A, Skalnik DG, Cosgrove MS. 2012. Structural basis for WDR5 interaction (Win) motif recognition in human SET1 family histone methyltransferases. *J Biol Chem* **287**: 27275-27289.
- Dou Y, Milne TA, Ruthenburg AJ, Lee S, Lee JW, Verdine GL, Allis CD, Roeder RG. 2006. Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol* **13**: 713-719.
- Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, Voets O, Verrijzer CP, Shilatifard A. 2012. Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes Dev* **26**: 2604-2620.
- Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. 2013. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol Cell Biol* **33**: 4745-4754.
- Issaeva I, Zonis Y, Rozovskaia T, Orlovsky K, Croce CM, Nakamura T, Mazo A, Eisenbach L, Canaani E. 2007. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol Cell Biol* **27**: 1889-1903.
- Krogan NJ, Dover J, Khorrami S, Greenblatt JF, Schneider J, Johnston M, Shilatifard A. 2002. COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem* **277**: 10753-10755.
- Lee J, Saha PK, Yang QH, Lee S, Park JY, Suh Y, Lee SK, Chan L, Roeder RG, Lee JW. 2008. Targeted inactivation of MLL3 histone H3-Lys-4 methyltransferase activity in the mouse reveals vital roles for MLL3 in adipogenesis. *Proc Natl Acad Sci U S A* **105**: 19229-19234.
- Liu H, Takeda S, Kumar R, Westergard TD, Brown EJ, Pandita TK, Cheng EHY, Hsieh JJD. 2010. Phosphorylation of MLL by ATR is required for execution of mammalian S-phase checkpoint. *Nature* **467**: 343-346.
- Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF, Shilatifard A. 2001. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A* **98**: 12902-12907.
- Morgan MA, Shilatifard A. 2013. Drosophila SETs its sights on cancer: Trl/MLL3/4 COMPASS-like complexes in development and disease. *Mol Cell Biol* **33**: 1698-1701.
- Muslin AJ, Tanner JW, Allen PM, Shaw AS. 1996. Interaction of 14-3-3 with signaling proteins is mediated by the recognition of phosphoserine. *Cell* **84**: 889-897.
- Odho Z, Southall SM, Wilson JR. 2010. Characterization of a novel WDR5-binding site that recruits RbBP5 through a conserved motif to enhance methylation of histone H3 lysine 4 by mixed lineage leukemia protein-1. *J Biol Chem* **285**: 32967-32976.

- Patel A, Dharmarajan V, Vought VE, Cosgrove MS. 2009. On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem* **284**: 24242-24256.
- Phanstiel DH, Brumbaugh J, Wenger CD, Tian S, Probasco MD, Bailey DJ, Swaney DL, Tervo MA, Bolin JM, Ruotti V et al. 2011. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nature methods* **8**: 821-827.
- Rittinger K, Budman J, Xu J, Volinia S, Cantley LC, Smerdon SJ, Gambin SJ, Yaffe MB. 1999. Structural analysis of 14-3-3 phosphopeptide complexes identifies a dual role for the nuclear export signal of 14-3-3 in ligand binding. *Mol Cell* **4**: 153-166.
- Roguev A, Schaft D, Shevchenko A, Pijnappel WW, Wilm M, Aasland R, Stewart AF. 2001. The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J* **20**: 7137-7148.
- Sarvan S, Avdic V, Tremblay V, Chaturvedi CP, Zhang P, Lanouette S, Blais A, Brunzelle JS, Brand M, Couture JF. 2011. Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain. *Nat Struct Mol Biol* **18**: 857-859.
- Shiromizu T, Adachi J, Watanabe S, Murakami T, Kuga T, Muraoka S, Tomonaga T. 2013. Identification of missing proteins in the neXtProt database and unregistered phosphopeptides in the PhosphoSitePlus database as part of the Chromosome-centric Human Proteome Project. *Journal of proteome research* **12**: 2414-2421.
- South PF, Fingerhman IM, Mersman DP, Du HN, Briggs SD. 2010. A conserved interaction between the SDI domain of Bre2 and the DPY30 domain of Sdc1 is required for histone methylation and gene expression. *J Biol Chem* **285**: 595-607.
- Southall SM, Wong PS, Odho Z, Roe SM, Wilson JR. 2009. Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol Cell* **33**: 181-191.
- Steward MM, Lee JS, O'Donovan A, Wyatt M, Bernstein BE, Shilatifard A. 2006. Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat Struct Mol Biol* **13**: 852-854.
- Wu L, Lee SY, Zhou B, Nguyen UT, Muir TW, Tan S, Dou Y. 2013. ASH2L regulates ubiquitylation signaling to MLL: trans-regulation of H3 K4 methylation in higher eukaryotes. *Mol Cell* **49**: 1108-1120.
- Wu M, Wang PF, Lee JS, Martin-Brown S, Florens L, Washburn M, Shilatifard A. 2008. Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol Cell Biol* **28**: 7337-7344.
- Yu BD, Hess JL, Horning SE, Brown GA, Korsmeyer SJ. 1995. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**: 505-508.
- Zhang P, Lee H, Brunzelle JS, Couture JF. 2012. The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Res* **40**: 4237-4246.

Supporting Online Material: Zhang et al.

Pamela Zhang¹, Chandra-Prakash Chaturvedi^{3,4}, Joseph Brunzelle⁶, Georgios Skiniotis⁵, Marjorie Brand^{3,4}, Ali Shilatifard² and Jean-François Couture¹

SUPPLEMENTARY RESULTS

Deconstruction of the WRAD Complex and its Allosteric Effect on KMT2 Enzymes

In our first study on the dissection of the WRAD complex, we determined that a fragment of 50 residues on RbBP5 (330-380) stimulates MLL1 enzymatic activity to the same extent as full-length RbBP5 (Avdic et al., 2011a). Given that amino acids 371-380 of RbBP5 are critical for binding WDR5, we reasoned that residues conferring binding to ASH2L and direct stimulation of MLL1 enzymatic activity would be located N-terminally to this motif (Figure S4.1A). To determine this minimal sequence, we assembled complexes with peptides spanning residues 328-379 (RbBP5³²⁸⁻³⁷⁹), 330-356 (RbBP5³³⁰⁻³⁵⁶) or 344-364 (RbBP5³⁴⁴⁻³⁶⁴) of RbBP5 with ASH2L^{SPRYdel} and the MLL1 SET domain and performed radiometric methyltransferase assays with histone H3 and tritium-labeled AdoMet. As shown in Figure S4.1B, complexes reconstituted with RbBP5³²⁸⁻³⁷⁹ or RbBP5³³⁰⁻³⁵⁶ recapitulated the activity observed with full-length RbBP5 while RbBP5³⁴⁴⁻³⁶⁴ failed to stimulate MLL1 methyltransferase activity. Isothermal titration calorimetry (ITC) experiments performed with RbBP5³²⁷⁻³⁵⁶ and RbBP5³⁴⁴⁻³⁵⁷ showed that these peptides bind with similar equilibrium dissociation constants (Figure S4.1C), suggesting that residues 344-357 are sufficient for the interaction of RbBP5 with ASH2L but residues 330-343 are required for the stimulation of MLL1 methyltransferase activity.

Deconvolution of ASH2L/RbBP5 Heterodimer Formation in MLL3 Interaction

At the surface of the ASH2L SPRY domain, two tyrosine residues, Tyr313 and Tyr475, form the RbBP5-binding interface. Tyr313 interacts with the middle section of RbBP5 and Tyr475 is in close proximity to the N-terminal end of the peptide and each tyrosine residue makes several contacts with RbBP5. Consistently, as shown in Figure 4.2A, substitution of each aromatic residue to alanine completely impairs the binding of ASH2L to RbBP5. Consistent with the ITC data, enzymatic assays with WRAD reconstituted with the ASH2L Tyr313Ala mutant fails to stimulate MLL1 enzymatic activity (Figure 4.2B). Surprisingly, despite the absence of binding of the ASH2L^{SPRY} Tyr475Ala mutant to RbBP5, methyltransferase assays performed with the WRAD complex reconstituted with the ASH2L Tyr475Ala mutant stimulates MLL1 methyltransferase activity similar to ASH2L wild-type. One possible reason explaining the differences between the results obtained with ITC and the methyltransferase assay may originate from the differences in the constructs used to perform these experiments. ITC was performed with the SPRY domain of ASH2L and an RbBP5 peptide while the methyltransferase assays were performed with full-length RbBP5, WDR5 and ASH2L subunits. These observations suggest that WDR5 is also important for the stability of the complex when assays are performed with full-length RbBP5 and ASH2L and the inclusion of this subunit presumably triggers a conformational change in RbBP5 rendering Tyr475 dispensable for the formation of a MLL1/WRAD complex. Alternatively, binding of the MLL1 SET domain to the RbBP5 N-terminus, the ABM region that is in close proximity of Tyr475, induce a structural reorganization of RbBP5 in a conformation non-permissive to make contacts with the aromatic residue. This hypothesis is consistent with the observation that RbBP5³³⁰⁻³⁴³ (Figure S4.1B), residues located at the N-terminus of the RbBP5 ABM, are essential for the stimulation of MLL1 methyltransferase activity but

dispensable for the interaction with ASH2L, raising the possibility that this region of RbBP5 directly contacts the MLL1 SET domain. This model is further supported by recent cryo-EM studies of COMPASS showing that the RbBP5 ABM is in close proximity of the SET1 catalytic domain (Takahashi et al., 2011). However, only a crystal structure of the WDR5/RbBP5/ASH2L/MLL1 complex will demonstrate the role, or lack thereof, of Tyr475 in the assembly of the WRAD complex and the enzymatic stimulation of the KMT2 enzymes.

An Evolutionarily Conserved D/E Box on RbBP5 is Important for Binding to ASH2L

After solving the crystal structure of the ASH2L/RbBP5 complex, we sought to identify residues on RbBP5 important for conferring binding to ASH2L. We first performed mutational analysis of RbBP5 residues binding to ASH2L focusing on the residues forming the bulge structure. Substitution of E347 and E349, two residues forming an extensive network of hydrogen bonds with ASH2L, to alanine led to approximately five- and three-fold reduction in MLL1 activity, respectively (Figure S4.3A). Similarly, substitution of F352, a residue that makes several hydrophobic interactions with ASH2L, to alanine resulted in a three-fold loss of MLL1 enzymatic activity. Finally, an RbBP5 construct harbouring a D353A mutation, a residue that forms a salt bridge with Arg343, failed to stimulate MLL1 methyltransferase activity to the same extent as wild-type RbBP5 (Figure S4.3A). Consistent with our methyltransferase assays, these mutants failed to immunoprecipitate ASH2L when expressed in HEK293 cells (Figure S4.3B). These observations strongly suggest that the D/E box of RbBP5 is important for WRAD assembly and stimulation of MLL1 methyltransferase activity.

SUPPLEMENTARY MATERIALS AND METHODS

Peptides and Antibodies

RbBP5 peptide corresponding to residues 344-357 was purchased from New England Peptides while the S350-phosphorylated form of the same peptide was purchased from GenScript. RbBP5 peptides corresponding to residues 327-356 unmodified or S350-phosphorylated were purchased from Synpeptide. All peptides were resuspended in water at 5 mM and stored at -20°C. The commercial antibody for RbBP5 is from Bethyl and the characterization of the ASH2L antibody was previously reported (Demers et al., 2007).

Protein Purification

A construct corresponding to the SPRY domain of ASH2L (residues 286-504) was cloned in a home-made His-SUMO vector, a plasmid allowing the expression of the protein in fusion with a His-tagged SUMO protein. To obtain diffraction quality crystals, residues 402-445, a fragment predicted to fold as a long, disordered loop, was replaced by a linker peptide corresponding to residues Ile-Ser-Gly-Arg-Gly-Ser, resulting in a final construct referred to as ASH2L^{SPRYdel}. The fusion protein was overexpressed in Rosetta cells (Novagen) with 0.1 mM IPTG for 16 hours at 18°C. Cells were harvested in 50 mM sodium phosphate (NaPi) pH 7.0, 500 mM sodium chloride (NaCl) and 5 mM β-mercaptoethanol (BME), lysed by sonication, clarified by centrifugation and filtration, and purified by TALON metal affinity chromatography (Clontech). The tagged protein was cleaved on beads using ULP1 for 16 hours at 4°C and the cleaved protein was further purified by size exclusion chromatography (Superdex 75, GE Healthcare) pre-equilibrated in a buffer composed of 20 mM Tris-HCl pH 7.0, 150 mM NaCl and 5 mM BME.

Point mutations were introduced into ASH2L^{SPRYdel}, full-length ASH2L (ASH2L^{FL}) and

RbBP5 (RbBP5^{FL}) by PCR site-directed mutagenesis (Agilent Technologies). The ASH2L SPRY mutants were purified using the same protocol as for the wild-type (WT). The catalytic domain of MLL1 (3762-3969) and MLL3 (4706-4911), WDR5 (22–334), RbBP5^{FL} and ASH2L^{FL} were overexpressed and purified as previously described (Avdic et al., 2011a; Avdic et al., 2011b).

Protein Crystallization and Structure determination

Once transferred in the cryoprotective solution, a single complete data set was collected on a HomeSource MicroMax-007 (Rigaku) equipped with a Raxis IV image plate detector. Reflections were indexed and scaled using d*TREK (Pflugrath, 1999). The structure was solved by molecular replacement using Phaser and using the apo-ASH2L^{SPRY} structure (Y. Chen et al., 2012) as a search model (rcsb 3TOJ.pdb). Model building was performed using Coot (Emsley, Lohkamp, Scott, & Cowtan, 2010) and the structure was further refined using iterative cycles of energy minimization by BUSTER 2.10.0 (Blanc et al., 2004) and Phenix. The quality of the structure was assessed using MolProbity (Headd et al., 2009).

Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) experiments were performed with a VP-ITC. Briefly, RbBP5 344-357 (1 mM) was injected into a solution composed of 25 mM Tris-HCl pH 8.0, 150 mM NaCl and 5 mM BME containing WT or mutant ASH2L^{SPRYdel} (0.06 mM). Titrations were performed at 19°C and the data was analyzed using Origin 7.0 (OriginLab) and showed that the RbBP5 peptide binds ASH2L with binding stoichiometries (N-values) between 0.8 and 1.0. The K_D, the error and the N-value were determined by fitting the integrated data to a standard single-site binding model.

***In Vitro* Methyltransferase Assay**

Methyltransferase assays were performed as previously described (Avdic et al., 2011a). MLL1 WinSET was incubated with equimolar amounts of WT or mutants of ASH2L^{FL} and RbBP5^{FL}, WDR5 and 10 μ M of recombinantly purified histone H3 in 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 3 mM dithiothreitol (DTT), 5 mM magnesium chloride (MgCl₂) and 5% glycerol. Reactions were initiated by adding 1 μ Ci of radiolabeled *S*-Adenosyl-L-methionine (AdoMet) and incubated for two hours at room temperature. Assays were stopped by spotting the reactions onto Whatman P81 filter papers and free AdoMet was removed by washing the filter papers four times in 250 ml of 50 mM sodium bicarbonate pH 9.0 and MLL1 activity was quantified by liquid scintillation counting.

ESI-MS

ESIMS analysis was performed at the SPARC BioCentre using a QSTAR Elite. MLL1 (10 μ M) or MLL3 SET (0.5 or 10 μ M) was incubated with equimolar amounts of ASH2L^{SPRYdel}, RbBP5 327-356 unmodified or S350-phosphorylated and 10 μ M of H3 20mer peptide (GenScript) in 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 3 mM DTT, 5 mM MgCl₂ and 5% glycerol. Reactions were initiated by adding 0.5 mM of *S*-Adenosyl-L-methionine (AdoMet) and incubated for two hours or 24 hours at room temperature. Reactions were stopped by flash freezing the samples in liquid nitrogen.

MEL Cells

Binding of FLAG-tagged proteins to the HS2 site of the *β -globin* LCR was measured by quantitative PCR as previously described (Demers et al., 2007) and reported as a fraction of the input and normalized to an intergenic region (between HS1 and the Ey promoter) of the

β-globin locus where ASH2L does not bind in MEL cells (Demers et al., 2007).

FLAG-tagged Immunoprecipitation

FLAG-tagged constructs corresponding to RbBP5 WT or mutants E347A, E349A, F352A and D353A were transfected into HEK293 cells. After 24 hours, the cells were harvested and lysed in RIPA buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 0.25% sodium deoxycholate), and 400 µg of each lysate was bound onto anti-Flag M2 Magnetic Beads (Sigma) for two hours at 4°C, washed with RIPA and eluted with a FLAG peptide. Bound proteins were detected using immunoblotting.

ASH2L Pull-Down

His-SUMO-tagged MLL3 SET was first incubated with 25 µl TALON metal affinity resin in 50 mM NaPi pH 7.0, 500 mM NaCl, 5 mM BME and 0.05% Triton X-100 for one hour at 4°C. After three washing steps with the binding buffer, 8 µM concentrations of unmodified or S350^{Phos} ASH2L^{SPRYdel}-RbBP5³²⁷⁻³⁵⁶ were incubated with the bound MLL3 in the presence of a H3 N-terminal peptide (20mer) and SAM for one hour at 4°C in 20 mM Tris-HCl pH 7.0, 150 mM NaCl, 5 mM BME, 50 mM imidazole and 0.05% Triton X-100. After extensively washing the beads with the same buffer, bound proteins were eluted using 2× sample loading buffer and separated on a SDS-PAGE. Binding was quantified using the ImageJ software.

SUPPLEMENTARY REFERENCES

Avdic V, Zhang P, Lanouette S, Groulx A, Tremblay V, Brunzelle J, Couture JF. 2011a. Structural and biochemical insights into MLL1 core complex assembly. *Structure* 19: 101-108.

- Avdic V, Zhang P, Lanouette S, Voronova A, Skerjanc I, Couture JF. 2011b. Fine-tuning the stimulation of MLL1 methyltransferase activity by a histone H3-based peptide mimetic. *FASEB J* 25: 960-967.
- Blanc E, Roversi P, Vonnrhein C, Flensburg C, Lea SM, Bricogne G. 2004. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr* 60: 2210-2221.
- Chen Y, Cao F, Wan B, Dou Y, Lei M. 2012. Structure of the SPRY domain of human Ash2L and its interactions with RbBP5 and DPY30. *Cell Res* 22: 598-602.
- Demers C, Chaturvedi CP, Ranish JA, Juban G, Lai P, Morle F, Aebersold R, Dilworth FJ, Groudine M, Brand M. 2007. Activator-mediated recruitment of the MLL2 methyltransferase complex to the beta-globin locus. *Mol Cell* 27: 573-584.
- Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486-501.
- Headd JJ, Immormino RM, Keedy DA, Emsley P, Richardson DC, Richardson JS. 2009. Autofix for backward-fit sidechains: using MolProbity and real-space refinement to put misfits in their place. *J Struct Funct Genomics* 10: 83-93.
- Pflugrath JW. 1999. The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr* 55: 1718-1725.
- Takahashi YH, Westfield GH, Oleskie AN, Trievel RC, Shilatifard A, Skiniotis G. 2011. Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. *Proc Natl Acad Sci U S A* 108: 20526-20531.

CHAPTER 5. DISCUSSION

The mammalian KMT2 enzymes associate with the WRAD complex to carry out H3K4 methylation (Dou et al., 2006), however, when I started my PhD, the interactions underlying KMT2-WRAD complex assembly and how WRAD regulates the activity of the KMT2 enzymes were not well investigated. During my PhD, I have demonstrated that 1) WDR5 uses two peptide-binding clefts on opposite sides of its β -propeller domain to bridge MLL1 and the other KMT2 enzymes to the regulatory subunit RbBP5, 2) the WDR5 peptidyl-arginine-binding cleft exhibits plasticity in binding the divergent C-terminal ends of the KMT2 Win peptides and 3) RbBP5 S350 phosphorylation stimulates formation of the RbBP5-ASH2L complex and H3K4 methylation by the KMT2 enzymes. By combining structural studies with structure-guided biochemical studies, my research has highlighted the important interactions governing KMT2-WRAD assembly and how these interactions regulate the H3K4 methyltransferase activity of the KMT2 enzymes.

5.1 WDR5, a Scaffolding Protein of Chromatin Complexes

Our studies on the scaffolding role of WDR5 within the core complex have uncovered two binding clefts located on opposite sides of its β -propeller domain. Particularly for the peptidyl-arginine-binding cleft, we have shown that WDR5 exhibits plasticity in binding the divergent C-terminal ends of the KMT2 Win peptides by using three different binding modes and undergoing subtle conformational changes to accommodate the binding of all Win motifs. On the opposite side, we have shown that the valine-aspartate-valine (VDV) motif of RbBP5 is required for recognition by the WDR5 V-shaped cleft. Altogether, these results suggest that WDR5 scaffolds protein complexes using at least two peptide-

binding clefts. These findings are consistent with the observation that the WD40 domain of WDR5 harbors many exposed surfaces (Xu & Min, 2011), enabling the β -propeller protein to bind a diverse range of partners. Aside from its role within the core complex, WDR5 binds to several proteins and long noncoding RNAs (lncRNAs) for transcription regulation. A recent quantitative analysis using HeLa nuclear extracts has revealed that WDR5 is approximately 10-fold more abundant than RbBP5 and ASH2L (van Nuland et al., 2013). The same study also showed that GFP-WDR5 directly or indirectly interacts with almost 200 different proteins, including subunits of previously known WDR5-interacting complexes as well as newly identified complexes. This study supports prior studies showing that WDR5 interacts with histone deacetylase 3 (HDAC3) (M. Z. Wu et al., 2011), pituitary homeobox 2 (Pitx2) (Gan et al., 2011), the mitochondrial antiviral-signaling protein VISA (Y. Y. Wang et al., 2010), the pluripotency transcription factor Oct4 (Ang et al., 2011) and the lncRNAs *HOTTIP* (K. C. Wang et al., 2011) and *NeST* (Gomez et al., 2013) for the expression of their responsive genes. In addition to the human Ada Two-A containing (ATAC) acetyltransferase complex (Y. L. Wang et al., 2008), WDR5 is a subunit of the non-specific lethal (NSL) and male-specific lethal (MSL) histone acetyltransferase complexes (Cai et al., 2010). Altogether, these studies demonstrate that WDR5 is a central hub in chromatin and transcription regulatory networks.

Our results showing that WDR5 harbors two binding clefts provide some mechanistic insights into its ability to bind a diverse range of partners. Based on the above studies showing that WDR5 is important for H3K4 methylation and transcription activation of specific genes, it is tempting to propose a stepwise model by which WDR5 uses its two binding clefts to orchestrate a series of events leading up to gene transcription. In this regard, WDR5 is first recruited to a specific gene promoter through its interaction with chromatin, a

transcription factor or an lncRNA using one of its binding clefts. Once tethered to the promoter, WDR5 assembles a KMT2 complex and cooperates with other KMT2 subunits to recruit additional factors required for transcription activation.

Coincidentally, it was recently shown that WDR5 directly interacts with the lncRNAs *HOTTIP* (K. C. Wang et al., 2011) and *NeST* (Gomez et al., 2013), both of which are involved in the recruitment of KMT2 complexes via WDR5 for H3K4 methylation and transcription activation of their respective loci. WDR5 has also been shown to bind over 1000 RNAs in embryonic stem cells (ESCs) (Yang et al., 2014). Intriguingly, WDR5 interacts with *HOTTIP* via a newly identified RNA-binding pocket that partially overlaps with its V-shaped cleft, suggesting that *HOTTIP* and RbBP5 may compete for the same binding cleft (Yang et al., 2014). The same study also showed that a key WDR5 phenylalanine residue is important in selectively binding *HOTTIP* without disrupting WDR5-RbBP5 interaction and MLL1 activity, reconciling the requirement for WDR5 to bind both *HOTTIP* and RbBP5 during transcription activation. Nevertheless, *NeST* and the other RNA interactors may also use the same RNA-binding pocket to bind WDR5 or bind via a distinct unidentified pocket. In the former case, the relevant question remains how the WDR5 RNA-binding pocket discriminates between *HOTTIP*, *NeST* and the other RNA interactors during transcription regulation by the KMT2 complexes. Analogous to the three different binding modes accommodated by the WDR5 peptidyl-arginine-binding cleft, the WDR5 RNA-binding cleft may facilitate discrimination between the different lncRNAs by using different sets of residues for their binding. Alternatively, if WDR5 harbors a single lncRNA-binding site, we anticipate that competition will take place between these lncRNAs for the recruitment of WDR5 along with a specific KMT2 enzyme to a punctate chromatin region. Obtaining the crystal structure of WDR5 in complex with several RNA molecules

will provide pivotal insights into WDR5 recognition and discrimination between the different RNA interactors.

5.2 WDR5, a Scaffolding Protein of KMT2 Complexes

A recent study comparing WRAD involvement in the stimulation of the KMT2 enzymes have shown that while WDR5 is required for stimulating H3K4me2 and H3K4me3 by MLL1 and SET1A, respectively, exclusion of WDR5 from *in vitro* enzymatic assays is dispensable for stimulating H3K4me1 by all KMT2 enzymes and further methylation by MLL2, MLL3, MLL4 and SET1B (Shinsky et al., 2015). This finding opposes the scaffolding role of WDR5 within the core complex and contradicts *in vivo* data showing that knockdown of WDR5 results in global loss of H3K4 methylation, suggesting that it is important for H3K4 methylation by all KMT2 enzymes (Dou et al., 2006). The results of Shinsky et al. are also in stark contrast with data obtained by our group showing that WDR5 is important for the stimulation of all KMT2 enzymes. The discrepancies between these results may stem from the use of different histone H3 substrates as well as the use of different KMT2 complex concentrations in the enzymatic assays. First, while we used full-length H3, Shinsky et al. used H3 20mer peptides. A recent study showing that RbBP5 binds a region nearby the histone H3 K27 residue (D. H. Kim et al., 2013), which is missing in the peptide substrates used by Shinsky et al., suggest that Shinsky et al. may not be using the optimal substrate for elucidating the impact of WDR5 on the stimulation of the KMT2 enzymes. Second, Shinsky et al. used 3 or 7 μM of the KMT2 complexes in their assays while we showed that the concentrations at which WDR5 is stimulatory is maximal at lower concentrations of the complexes (0.5 to 1 μM), corroborating with data obtained by Cao et al. showing that the stimulatory role of WDR5 is maximal at 0.1 to 0.5 μM of the MLL1

complex and decreases with increasing complex concentrations (Cao et al., 2010). In addition, Cao et al. have shown that in the absence of WDR5, the ASH2L-RbBP5 heterodimer stimulates MLL1 methyltransferase activity through transient interactions with the MLL1 SET domain, which are more likely to occur at higher concentrations of the complex and may mask the stimulatory role of WDR5 at these concentrations. Thus, to discriminate between the stimulatory roles of the ASH2L-RbBP5 heterodimer and WDR5 and to ultimately observe the impact of adding WDR5, enzymatic assays should be performed at lower concentrations of the KMT2 complexes and with histone H3 substrates that include at least the first 30 amino acids of the histone H3 tail.

5.3 RbBP5 S350 Phosphorylation, an Additional Regulatory Mechanism

Controlling H3K4 Methylation

Due to its involvement in cancer, H3K4 methylation is tightly regulated at various stages, from assembly of the KMT2 complexes to their enzymatic activity to their recruitment to chromatin. Our investigation on the pairwise interactions within the KMT2 core complex has shown that key residues at the WDR5-RbBP5, WDR5-KMT2 and RbBP5-ASH2L binding interfaces control the formation of the KMT2 complexes and subsequent H3K4 methylation. Our results also demonstrated, for the first time, that phosphorylation on the core subunit RbBP5 of the KMT2 complexes provides another level of control in regulating H3K4 methyltransferase activity. Intriguingly, sequence alignment of the RbBP5 D/E box from different species shows that the serine residue is only conserved in higher eukaryotes, suggesting that metazoans have implemented additional controls for the methylation of histone H3K4. However, the enzyme responsible for phosphorylating RbBP5 S350 is yet to be identified. Kinase profiling assays against the 518 known human kinases (J.

Zhang, Yang, & Gray, 2009) using RbBP5 wild-type or the S350A mutant as substrates will provide a starting point for determining the potential kinases. In addition, the motif surrounding S350 may be used to predict the potential kinases. Nevertheless, a similar mechanism has been demonstrated in yeast whereby ubiquitylation of Cps35 is required for the incorporation of Cps35 in COMPASS and H2Bub-mediated H3K4me3 (Vitaliano-Prunier et al., 2008). Both mechanisms, albeit different, do not turn on/off H3K4 methylation but rather act as molecular switches fine-tuning H3K4 methylation.

5.4 ASH2L SPRY, a New Unique Phospho-reader

Our results have demonstrated that the ASH2L SPRY domain binds RbBP5 S350^{phos} by undergoing local structural reorganization to make additional interactions with the phosphate moiety, suggesting that ASH2L SPRY is a new phospho-reader. However, in contrast with other canonical phospho-readers, such as the 14-3-3 family of proteins (Rittinger et al., 1999), WW domains (Verdecia, Bowman, Lu, Hunter, & Noel, 2000) and polo-box domains (K. Y. Cheng, Lowe, Sinclair, Nigg, & Johnson, 2003), which form an extensive network of contacts with the phosphorylated motif of their respective targets, the number of contacts between the ASH2L SPRY domain and RbBP5 S350^{phos} are far less than the canonical phospho-readers. Accordingly, the phosphate group of RbBP5 is solvent exposed and does not bind within a well-defined binding pocket (Figure 5.1). Our enzymatic assays have shown that both ASH2L-RbBP5 and ASH2L-RbBP5^{phos} stimulate the activity of the KMT2 enzymes, but to different extent. These observations suggest that phosphorylation of RbBP5 acts as a tuner of KMT2 turnover numbers. Interestingly, we have also shown that ASH2L-RbBP5^{phos} binds the MLL3 SET domain with five-fold higher affinity than ASH2L-RbBP5, suggesting that it creates a better binding platform for interaction with MLL3 and

that residues on the MLL3 SET domain interact with the phosphate moiety. It is tempting to speculate that the crystal structure of ASH2L-RbBP5^{phos} only provides a view of the residues forming half of the phospho-binding site. Additional structural studies of the ASH2L-RbBP5^{phos}-MLL3 complex will provide evidence for the interplay between RbBP5 phosphorylation and H3K4 methylation by the KMT2 enzymes.

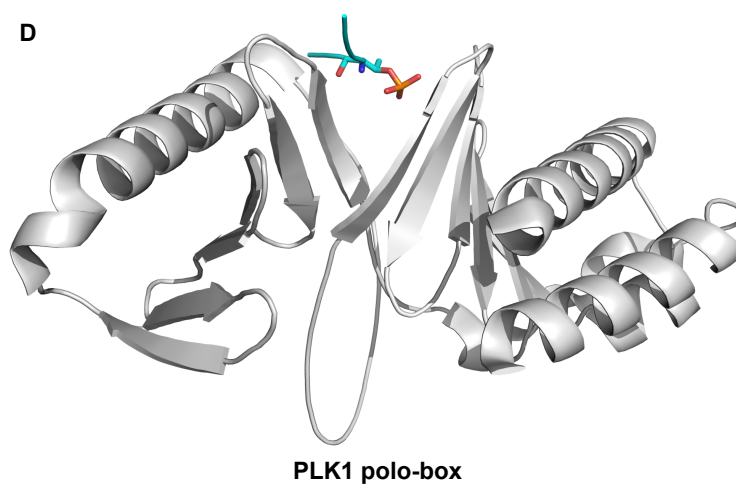
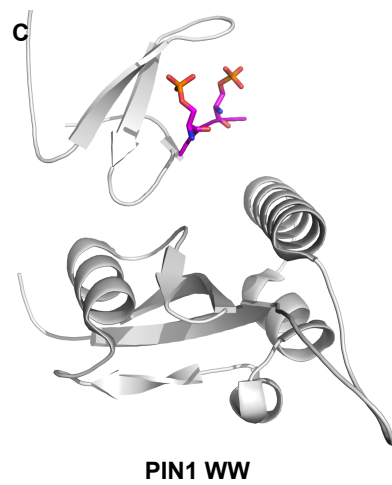
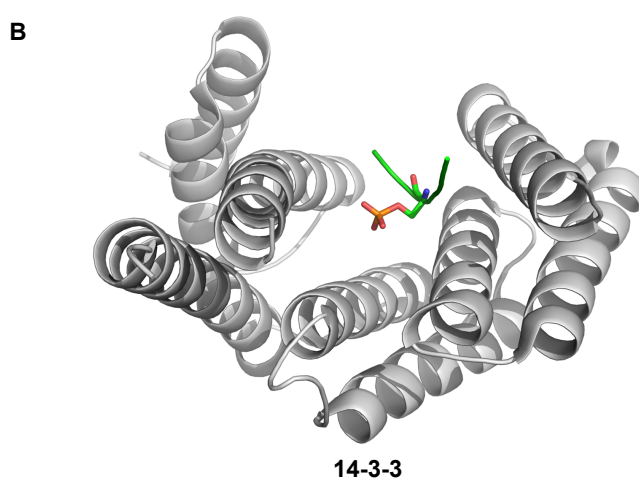


Figure 5.1. Structural comparison of ASH2L SPRY and the known phospho-binding domains.

The domains are shown in gray and their respective peptides are shown as colored lines with the phosphate groups shown as sticks.

Protein Data Bank (PDB) codes: ASH2L SPRY (4X8N), 14-3-3 (1QJB), PIN1 WW (1F8A), PLK1 polo-box (1Q4K)

5.5 Limitations

As our studies only focus on the minimal domains required for KMT2-WRAD complex formation, it is very likely that other domains or a combination of domains are involved in controlling KMT2 complex assembly and activity. For example, in the presence of the MLL1 SET domain, WDR5 binds MLL1 with 14-fold more affinity than the Win motif alone, suggesting that WDR5 directly interacts with the MLL1 SET domain to facilitate complex assembly (A. Patel et al., 2009). Therefore, structural data on the full-length proteins and complexes will elucidate the entire spectrum of interactions controlling complex formation and activity. In addition, crystal structure provides a static picture of the protein complex when it is highly possible that the complex can adopt various conformations. Therefore, the use of techniques that can monitor dynamics, such as deuterium exchange mass spectroscopy and fluorescence resonance energy transfer (FRET), can be used in conjunction with structural studies to determine the conformations adopted by the WRAD complex upon interaction with the different KMT2 enzymes. Another potential pitfall of my approach is the type of substrates I have used. Histone H3 peptides and full-length histone H3 do not reflect the native substrate found in cells; therefore, it will be more informative to use reconstituted nucleosomes in *in vitro* enzymatic assays and structural studies, especially for investigating crosstalk between H3K4 methylation and other modifications deposited adjacent to the H3K4 methylation site or on other histone proteins (L. Wu et al., 2013).

5.6 Targeting MLL1 Complex Assembly for Leukemia Drug Development

Our understanding of the underlying factors governing MLL1 complex assembly and activity provides a therapeutic platform for the design of novel drug candidates that can treat

MLL1-fusion leukemia, on the premise that the H3K4 methyltransferase activity of MLL1 is a key factor contributing to leukemogenesis (Thiel et al., 2010). Through structural studies on the interactions underlying formation of the MLL1 complex, we have gained a wealth of knowledge on the key residues controlling complex assembly. Thus, design of synthetic peptide mimetics that can inhibit these interactions may be used as drug candidates for targeting the assembly and the activity of the enzyme. In this regard, we have shown that MLL1-WDR5 interaction can be inhibited by an N α -acetylated H3 peptide mimetic that has higher binding affinity for WDR5 than both MLL1 and H3 (Avdic et al., 2011b). Also targeting the MLL1-WDR5 interaction, Cao et al. have developed a compound inhibitor that only disrupts MLL1 activity but not the activity of the other KMT2 enzymes (Cao et al., 2014). In addition, Karatas et al. have designed a library of potent high-affinity peptidomimetics targeting MLL1-WDR5 interaction and have provided the first proof-of-concept showing that inhibition of MLL1-WDR5 interaction by one such peptidomimetic decreases the expression of the key hematopoietic genes *HOXA9* and *MEIS1* involved in MLL1-fusion leukemia, inhibits cell growth and induces apoptosis of bone marrow leukemia cells (Karatas et al., 2013). This important finding, in addition to the study by Cao et al. showing that the specific inhibitor of MLL1-WDR5 interaction does not induce toxicity in normal cells, demonstrates that inhibitors disrupting MLL1 complex assembly is a genuine strategy in combating MLL1-fusion leukemia.

As WDR5 is involved in other chromatin complexes and disruption of its peptidyl-arginine-binding cleft may also disrupt its interaction with other complexes, our studies on RbBP5 S350 phosphorylation have uncovered another target that can be exploited. We have shown that RbBP5 S350 phosphorylation stimulates MLL1 complex assembly and activity, suggesting that leukemia cells displaying constitutive MLL1 activity may be a result of

constitutive RbBP5 S350 phosphorylation. Thus, identification of the kinase and development of kinase inhibitors that selectively disrupt RbBP5 S350 phosphorylation without disrupting RbBP5-ASH2L interaction in normal cells may present an attractive therapeutic window. Indeed, the use and validity of several kinase inhibitors as cancer treatments has already been proven on the market (J. Zhang et al., 2009). Therefore, our studies on the RbBP5-ASH2L interaction have provided a lead for the development of kinase inhibitors against leukemia.

5.7 Concluding Remarks

During my PhD, I have shown that 1) WDR5 binds the RbBP5 VDV motif using its V-shaped cleft and WDR5 scaffolds the KMT2 complexes using two binding clefts on opposite sides of its β -propeller domain, 2) WDR5 binds all KMT2 enzymes by exhibiting plasticity within its peptidyl-arginine-binding cleft and WDR5 is required for the optimal stimulation of all KMT2 enzymes and 3) RbBP5 S350 phosphorylation at the RbBP5-ASH2L binding interface is a new mechanism regulating KMT2 complex assembly and activity and the ASH2L SPRY domain is a new phospho-reader for which the mode of phosphopeptide recognition is the first of its kind. Altogether, my PhD thesis on the structural and biochemical dissection of the interactions underlying KMT2 complex formation and activity has increased our understanding on the structure-function relationship between KMT2 complex assembly and activity as well as established the foundation for future investigation and the design of cancer drug candidates targeting these interactions.

REFERENCES (FOR CHAPTERS 1 AND 5)

- Ambler, R. P., & Rees, M. W. (1959). [epsilon]-N-Methyl-lysine in Bacterial Flagellar Protein. *Nature*, 184(4679), 56-57.
- Amoros, M., Corominas, M., Deak, P., & Serras, F. (2002). The ash2 gene is involved in *Drosophila* wing development. *Int J Dev Biol*, 46(3), 321-324.
- Ananthanarayanan, M., Li, Y., Surapureddi, S., Balasubramanian, N., Ahn, J., Goldstein, J. A., & Suchy, F. J. (2011). Histone H3K4 trimethylation by MLL3 as part of ASCOM complex is critical for NR activation of bile acid transporter genes and is downregulated in cholestasis. *Am J Physiol Gastrointest Liver Physiol*, 300(5), G771-781. doi: 10.1152/ajpgi.00499.2010
- Ang, Y. S., Tsai, S. Y., Lee, D. F., Monk, J., Su, J., Ratnakumar, K., . . . Lemischka, I. R. (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell*, 145(2), 183-197. doi: 10.1016/j.cell.2011.03.003
- Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., . . . Korsmeyer, S. J. (2002). MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet*, 30(1), 41-47. doi: 10.1038/ng765
- Avdic, V., Zhang, P., Lanouette, S., Groulx, A., Tremblay, V., Brunzelle, J., & Couture, J. F. (2011a). Structural and biochemical insights into MLL1 core complex assembly. *Structure*, 19(1), 101-108. doi: 10.1016/j.str.2010.09.022
- Avdic, V., Zhang, P., Lanouette, S., Voronova, A., Skerjanc, I., & Couture, J. F. (2011b). Fine-tuning the stimulation of MLL1 methyltransferase activity by a histone H3-based peptide mimetic. *Faseb j*, 25(3), 960-967. doi: 10.1096/fj.10-171959
- Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res*, 21(3), 381-395. doi: cr201122 [pii] 10.1038/cr.2011.22
- Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., . . . Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, 129(4), 823-837. doi: <http://dx.doi.org/10.1016/j.cell.2007.05.009>
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., . . . Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2), 315-326. doi: 10.1016/j.cell.2006.02.041
- Bertani, S., Sauer, S., Bolotin, E., & Sauer, F. (2011). The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol Cell*, 43(6), 1040-1046. doi: 10.1016/j.molcel.2011.08.019
- Broeker, P. L., Harden, A., Rowley, J. D., & Zeleznik-Le, N. (1996). The mixed lineage leukemia (MLL) protein involved in 11q23 translocations contains a domain that binds cruciform DNA and scaffold attachment region (SAR) DNA. *Curr Top Microbiol Immunol*, 211, 259-268.
- Cai, Y., Jin, J., Swanson, S. K., Cole, M. D., Choi, S. H., Florens, L., . . . Conaway, R. C. (2010). Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. *J Biol Chem*, 285(7), 4268-4272. doi: 10.1074/jbc.C109.087981
- Cao, F., Chen, Y., Cierpicki, T., Liu, Y., Basrur, V., Lei, M., & Dou, Y. (2010). An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through

- coordinated substrate interactions with the MLL1 SET domain. *PLoS One*, 5(11), e14102. doi: 10.1371/journal.pone.0014102
- Cao, F., Townsend, E. C., Karatas, H., Xu, J., Li, L., Lee, S., . . . Dou, Y. (2014). Targeting MLL1 H3 K4 methyltransferase activity in MLL leukemia. *Mol Cell*, 53(2), 247-261. doi: 10.1016/j.molcel.2013.12.001
- Caslini, C., Alarcon, A. S., Hess, J. L., Tanaka, R., Murti, K. G., & Biondi, A. (2000). The amino terminus targets the mixed lineage leukemia (MLL) protein to the nucleolus, nuclear matrix and mitotic chromosomal scaffolds. *Leukemia*, 14(11), 1898-1908.
- Caslini, C., Yang, Z., El-Osta, M., Milne, T. A., Slany, R. K., & Hess, J. L. (2007). Interaction of MLL amino terminal sequences with menin is required for transformation. *Cancer Res*, 67(15), 7275-7283. doi: 10.1158/0008-5472.can-06-2369
- Chen, J., Santillan, D. A., Koonce, M., Wei, W., Luo, R., Thirman, M. J., . . . Diaz, M. O. (2008). Loss of MLL PHD-finger-3 is necessary for MLL-ENL-induced hematopoietic stem cell immortalization. *Cancer Res*, 68(15), 6199-6207. doi: 10.1158/0008-5472.can-07-6514
- Chen, Y., Cao, F., Wan, B., Dou, Y., & Lei, M. (2012). Structure of the SPRY domain of human Ash2L and its interactions with RbBP5 and DPY30. *Cell Res*, 22(3), 598-602. doi: 10.1038/cr.2012.9
- Chen, Y., Wan, B., Wang, K. C., Cao, F., Yang, Y., Protacio, A., . . . Lei, M. (2011). Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding. *EMBO Rep*, 12(8), 797-803. doi: 10.1038/embo.2011.101
- Cheng, K. Y., Lowe, E. D., Sinclair, J., Nigg, E. A., & Johnson, L. N. (2003). The crystal structure of the human polo-like kinase-1 polo box domain and its phospho-peptide complex. *Embo j*, 22(21), 5757-5768. doi: 10.1093/emboj/cdg558
- Cho, Y. W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., . . . Ge, K. (2007). PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem*, 282(28), 20395-20406. doi: M701574200 [pii] 10.1074/jbc.M701574200
- Collins, R. E., Tachibana, M., Tamaru, H., Smith, K. M., Jia, D., Zhang, X., . . . Cheng, X. (2005). In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *J Biol Chem*, 280(7), 5563-5570. doi: 10.1074/jbc.M410483200
- Couture, J. F., Collazo, E., Brunzelle, J. S., & Trievel, R. C. (2005). Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes Dev*, 19(12), 1455-1465. doi: 10.1101/gad.1318405
- Couture, J. F., Collazo, E., & Trievel, R. C. (2006). Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat Struct Mol Biol*, 13(8), 698-703. doi: 10.1038/nsmb1116
- Dehe, P. M., Dichtl, B., Schaft, D., Roguev, A., Pamblanco, M., Lebrun, R., . . . Geli, V. (2006). Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J Biol Chem*, 281(46), 35404-35412. doi: M603099200 [pii] 10.1074/jbc.M603099200
- Demers, C., Chaturvedi, C. P., Ranish, J. A., Juban, G., Lai, P., Morle, F., . . . Brand, M. (2007). Activator-mediated recruitment of the MLL2 methyltransferase complex to the beta-globin locus. *Mol Cell*, 27(4), 573-584. doi: 10.1016/j.molcel.2007.06.022

- Dillon, S. C., Zhang, X., Trievel, R. C., & Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol*, 6(8), 227. doi: 10.1186/gb-2005-6-8-227
- Dou, Y., Milne, T. A., Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., . . . Roeder, R. G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol*, 13(8), 713-719. doi: 10.1038/nsmb1128
- Dou, Y., Milne, T. A., Tackett, A. J., Smith, E. R., Fukuda, A., Wysocka, J., . . . Roeder, R. G. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell*, 121(6), 873-885. doi: 10.1016/j.cell.2005.04.031
- Dover, J., Schneider, J., Tawiah-Boateng, M. A., Wood, A., Dean, K., Johnston, M., & Shilatifard, A. (2002). Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. *J Biol Chem*, 277(32), 28368-28371. doi: 10.1074/jbc.C200348200
- Ernst, P., & Vakoc, C. R. (2012). WRAD: enabler of the SET1-family of H3K4 methyltransferases. *Brief Funct Genomics*, 11(3), 217-226. doi: 10.1093/bfpg/els017
- Finch, J. T., & Klug, A. (1976). Solenoidal model for superstructure in chromatin. *Proc Natl Acad Sci U S A*, 73(6), 1897-1901.
- Fossati, A., Dolfini, D., Donati, G., & Mantovani, R. (2011). NF-Y recruits Ash2L to impart H3K4 trimethylation on CCAAT promoters. *PLoS One*, 6(3), e17220. doi: 10.1371/journal.pone.0017220
- Franklin, S. G., & Zweidler, A. (1977). Non-allelic variants of histones 2a, 2b and 3 in mammals. *Nature*, 266(5599), 273-275.
- Gan, Q., Thiebaud, P., Theze, N., Jin, L., Xu, G., Grant, P., & Owens, G. K. (2011). WD repeat-containing protein 5, a ubiquitously expressed histone methyltransferase adaptor protein, regulates smooth muscle cell-selective gene activation through interaction with pituitary homeobox 2. *J Biol Chem*, 286(24), 21853-21864. doi: M111.233098 [pii] 10.1074/jbc.M111.233098
- Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K. R., . . . Stewart, A. F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development*, 133(8), 1423-1432. doi: 10.1242/dev.02302
- Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell*, 128(4), 635-638. doi: 10.1016/j.cell.2007.02.006
- Gomez, J. A., Wapinski, Orly L., Yang, Yul W., Bureau, J.-F., Gopinath, S., Monack, Denise M., . . . Kirkegaard, K. (2013). The NeST Long ncRNA Controls Microbial Susceptibility and Epigenetic Activation of the Interferon- γ Locus. *Cell*, 152(4), 743-754. doi: <http://dx.doi.org/10.1016/j.cell.2013.01.015>
- Goo, Y. H., Sohn, Y. C., Kim, D. H., Kim, S. W., Kang, M. J., Jung, D. J., . . . Lee, J. W. (2003). Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. *Mol Cell Biol*, 23(1), 140-149.
- Guccione, E., Bassi, C., Casadio, F., Martinato, F., Cesaroni, M., Schuchlantz, H., . . . Amati, B. (2007). Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. *Nature*, 449(7164), 933-937. doi: 10.1038/nature06166

- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., . . . Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*, 39(3), 311-318. doi: 10.1038/ng1966
- Hsu, D. R., & Meyer, B. J. (1994). The DPY30 gene encodes an essential component of the *Caenorhabditis elegans* dosage compensation machinery. *Genetics*, 137(4), 999-1018.
- Hu, D., Gao, X., Morgan, M. A., Herz, H. M., Smith, E. R., & Shilatifard, A. (2013a). The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol Cell Biol*, 33(23), 4745-4754. doi: 10.1128/MCB.01181-13
MCB.01181-13 [pii]
- Hu, D., Garruss, A. S., Gao, X., Morgan, M. A., Cook, M., Smith, E. R., & Shilatifard, A. (2013b). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nat Struct Mol Biol*, 20(9), 1093-1097. doi: 10.1038/nsmb.2653
- Hughes, C. M., Rozenblatt-Rosen, O., Milne, T. A., Copeland, T. D., Levine, S. S., Lee, J. C., . . . Meyerson, M. (2004). Menin associates with a trithorax family histone methyltransferase complex and with the *hoxc8* locus. *Mol Cell*, 13(4), 587-597.
- Huntsman, D. G., Chin, S. F., Muleris, M., Batley, S. J., Collins, V. P., Wiedemann, L. M., . . . Caldas, C. (1999). MLL2, the second human homolog of the *Drosophila* trithorax gene, maps to 19q13.1 and is amplified in solid tumor cell lines. *Oncogene*, 18(56), 7975-7984. doi: 10.1038/sj.onc.1203291
- Hyllus, D., Stein, C., Schnabel, K., Schiltz, E., Imhof, A., Dou, Y., . . . Bauer, U. M. (2007). PRMT6-mediated methylation of R2 in histone H3 antagonizes H3 K4 trimethylation. *Genes Dev*, 21(24), 3369-3380. doi: 10.1101/gad.447007
- Jiang, H., Shukla, A., Wang, X., Chen, W. Y., Bernstein, B. E., & Roeder, R. G. (2011). Role for DPY30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell*, 144(4), 513-525. doi: 10.1016/j.cell.2011.01.020
- Jones, R. S., & Gelbart, W. M. (1993). The *Drosophila* Polycomb-group gene *Enhancer of zeste* contains a region with sequence similarity to trithorax. *Mol Cell Biol*, 13(10), 6357-6366.
- Karatas, H., Townsend, E. C., Cao, F., Chen, Y., Bernard, D., Liu, L., . . . Wang, S. (2013). High-affinity, small-molecule peptidomimetic inhibitors of MLL1/WDR5 protein-protein interaction. *J Am Chem Soc*, 135(2), 669-682. doi: 10.1021/ja306028q
- Kawabe, Y., Wang, Y. X., McKinnell, I. W., Bedford, M. T., & Rudnicki, M. A. (2012). *Carm1* regulates Pax7 transcriptional activity through MLL1/2 recruitment during asymmetric satellite stem cell divisions. *Cell Stem Cell*, 11(3), 333-345. doi: 10.1016/j.stem.2012.07.001
- Kim, D. H., Tang, Z., Shimada, M., Fierz, B., Houck-Loomis, B., Bar-Dagen, M., . . . Lee, J. W. (2013). Histone H3K27 trimethylation inhibits H3 binding and function of SET1-like H3K4 methyltransferase complexes. *Mol Cell Biol*, 33(24), 4936-4946. doi: 10.1128/mcb.00601-13
- Kim, T., & Buratowski, S. (2009). Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell*, 137(2), 259-272. doi: 10.1016/j.cell.2009.02.045
- Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science*, 184(4139), 868-871.

- Krivtsov, A. V., & Armstrong, S. A. (2007). MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer*, 7(11), 823-833. doi: 10.1038/nrc2253
- Krogan, N. J., Dover, J., Khorrami, S., Greenblatt, J. F., Schneider, J., Johnston, M., & Shilatifard, A. (2002). COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem*, 277(13), 10753-10755. doi: 10.1074/jbc.C200023200
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., & Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & Development*, 16(22), 2893-2905. doi:10.1101/gad.1035902
- Lanouette, S., Mongeon, V., Figeys, D., & Couture, J. F. (2014). The functional diversity of protein lysine methylation. *Mol Syst Biol*, 10, 724. doi: 10.1002/msb.134974
- Lechner, M. S., Levitan, I., & Dressler, G. R. (2000). PTIP, a novel BRCT domain-containing protein interacts with Pax2 and is associated with active chromatin. *Nucleic Acids Res*, 28(14), 2741-2751.
- Lee, J. H., & Skalnik, D. G. (2008). Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol Cell Biol*, 28(2), 609-618. doi: MCB.01356-07 [pii] 10.1128/MCB.01356-07
- Lee, S., Lee, D. K., Dou, Y., Lee, J., Lee, B., Kwak, E., . . . Lee, J. W. (2006). Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. *Proc Natl Acad Sci U S A*, 103(42), 15392-15397. doi: 10.1073/pnas.0607313103
- Lee, S., Lee, J., Lee, S. K., & Lee, J. W. (2008). Activating signal cointegrator-2 is an essential adaptor to recruit histone H3 lysine 4 methyltransferases MLL3 and MLL4 to the liver X receptors. *Mol Endocrinol*, 22(6), 1312-1319. doi: 10.1210/me.2008-0012
- Lim, D. A., Huang, Y. C., Swigut, T., Mirick, A. L., Garcia-Verdugo, J. M., Wysocka, J., . . . Alvarez-Buylla, A. (2009). Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells. *Nature*, 458(7237), 529-533. doi: 10.1038/nature07726
- Martin, C., & Zhang, Y. (2005). The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol*, 6(11), 838-849. doi: 10.1038/nrm1761
- McKinnell, I. W., Ishibashi, J., Le Grand, F., Punch, V. G., Addicks, G. C., Greenblatt, J. F., . . . Rudnicki, M. A. (2008). Pax7 activates myogenic genes by recruitment of a histone methyltransferase complex. *Nat Cell Biol*, 10(1), 77-84. doi: 10.1038/ncb1671
- Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., . . . Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A*, 98(23), 12902-12907. doi: 10.1073/pnas.231473398
- Milne, T. A., Kim, J., Wang, G. G., Stadler, S. C., Basrur, V., Whitcomb, S. J., . . . Allis, C. D. (2010). Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. *Mol Cell*, 38(6), 853-863. doi: 10.1016/j.molcel.2010.05.011

- Mohan, M., Herz, H. M., Smith, E. R., Zhang, Y., Jackson, J., Washburn, M. P., . . . Shilatifard, A. (2011). The COMPASS family of H3K4 methylases in *Drosophila*. *Mol Cell Biol*, 31(21), 4310-4318. doi: 10.1128/mcb.06092-11
- Muller, H. J. (1930). Types of visible variations induced by X-rays in *Drosophila*. *Journal of Genetics*, 22(3), 299-334. doi: 10.1007/BF02984195
- Murray, K. (1964). The Occurrence of ϵ -N-Methyl Lysine in Histones. *Biochemistry*, 3(1), 10-15. doi: 10.1021/bi00889a003
- Nakagawa, T., Kajitani, T., Togo, S., Masuko, N., Ohdan, H., Hishikawa, Y., . . . Ito, T. (2008). Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes Dev*, 22(1), 37-49. doi: 10.1101/gad.1609708
- Odho, Z., Southall, S. M., & Wilson, J. R. (2010). Characterization of a novel WDR5-binding site that recruits RbBP5 through a conserved motif to enhance methylation of histone H3 lysine 4 by mixed lineage leukemia protein-1. *J Biol Chem*, 285(43), 32967-32976. doi: 10.1074/jbc.M110.159921
- Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V. M., . . . Zhang, Y. (2005). hDOT1L links histone methylation to leukemogenesis. *Cell*, 121(2), 167-178. doi: 10.1016/j.cell.2005.02.020
- Pang, C. N., Gasteiger, E., & Wilkins, M. R. (2010). Identification of arginine- and lysine-methylation in the proteome of *Saccharomyces cerevisiae* and its functional implications. *BMC Genomics*, 11, 92. doi: 10.1186/1471-2164-11-92
- Patel, A., Dharmarajan, V., & Cosgrove, M. S. (2008a). Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. *J Biol Chem*, 283(47), 32158-32161. doi: 10.1074/jbc.C800164200
- Patel, A., Dharmarajan, V., Vought, V. E., & Cosgrove, M. S. (2009). On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem*, 284(36), 24242-24256. doi: 10.1074/jbc.M109.014498
- Patel, A., Vought, V. E., Dharmarajan, V., & Cosgrove, M. S. (2008b). A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. *J Biol Chem*, 283(47), 32162-32175. doi: 10.1074/jbc.M806317200
- Patel, A., Vought, V. E., Dharmarajan, V., & Cosgrove, M. S. (2011). A novel non-SET domain multi-subunit methyltransferase required for sequential nucleosomal histone H3 methylation by the mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem*, 286(5), 3359-3369. doi: 10.1074/jbc.M110.174524
- Patel, S. R., Kim, D., Levitan, I., & Dressler, G. R. (2007). The BRCT-domain containing protein PTIP links PAX2 to a histone H3, lysine 4 methyltransferase complex. *Dev Cell*, 13(4), 580-592. doi: 10.1016/j.devcel.2007.09.004
- Peterson, C. L., & Herskowitz, I. (1992). Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. *Cell*, 68(3), 573-583.
- Porter, I. M., McClelland, S. E., Khoudoli, G. A., Hunter, C. J., Andersen, J. S., McAinsh, A. D., . . . Swedlow, J. R. (2007). Bod1, a novel kinetochore protein required for chromosome biorientation. *J Cell Biol*, 179(2), 187-197. doi: 10.1083/jcb.200704098
- Qian, C., & Zhou, M. M. (2006). SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell Mol Life Sci*, 63(23), 2755-2763. doi: 10.1007/s00018-006-6274-5

- Rampalli, S., Li, L., Mak, E., Ge, K., Brand, M., Tapscott, S. J., & Dilworth, F. J. (2007). p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation. *Nat Struct Mol Biol*, 14(12), 1150-1156. doi: 10.1038/nsmb1316
- Razin, A., & Cedar, H. (1977). Distribution of 5-methylcytosine in chromatin. *Proceedings of the National Academy of Sciences*, 74(7), 2725-2728.
- Rittinger, K., Budman, J., Xu, J., Volinia, S., Cantley, L. C., Smerdon, S. J., . . . Yaffe, M. B. (1999). Structural analysis of 14-3-3 phosphopeptide complexes identifies a dual role for the nuclear export signal of 14-3-3 in ligand binding. *Mol Cell*, 4(2), 153-166. doi: S1097-2765(00)80363-9 [pii]
- Roguev, A., Schaff, D., Shevchenko, A., Pijnappel, W. W., Wilm, M., Aasland, R., & Stewart, A. F. (2001). The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *Embo j*, 20(24), 7137-7148. doi: 10.1093/emboj/20.24.7137
- Ruault, M., Brun, M. E., Ventura, M., Roizes, G., & De Sario, A. (2002). MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a chromosome region frequently deleted in myeloid leukaemia. *Gene*, 284(1-2), 73-81.
- Ruthenburg, A. J., Wang, W., Graybosch, D. M., Li, H., Allis, C. D., Patel, D. J., & Verdine, G. L. (2006). Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nat Struct Mol Biol*, 13(8), 704-712. doi: 10.1038/nsmb1119
- Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., . . . Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature*, 419(6905), 407-411. doi: 10.1038/nature01080
- Sarvan, S., Avdic, V., Tremblay, V., Chaturvedi, C. P., Zhang, P., Lanouette, S., . . . Couture, J. F. (2011). Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain. *Nat Struct Mol Biol*, 18(7), 857-859. doi: 10.1038/nsmb.2093
- Schneider, J., Wood, A., Lee, J. S., Schuster, R., Dueker, J., Maguire, C., . . . Shilatifard, A. (2005). Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Mol Cell*, 19(6), 849-856. doi: S1097-2765(05)01521-2 [pii]
10.1016/j.molcel.2005.07.024
- Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C., & Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol*, 6(1), 73-77. doi: 10.1038/ncb1076
- Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu Rev Biochem*, 81, 65-95. doi: 10.1146/annurev-biochem-051710-134100
- Shinsky, S. A., Monteith, K. E., Viggiano, S., & Cosgrove, M. S. (2015). Biochemical Reconstitution and Phylogenetic Comparison of Human SET1 Family Core Complexes Involved in Histone Methylation. *J Biol Chem*, 290(10), 6361-6375. doi: 10.1074/jbc.M114.627646
- Sims, R. J., 3rd, Chen, C. F., Santos-Rosa, H., Kouzarides, T., Patel, S. S., & Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J Biol Chem*, 280(51), 41789-41792. doi: 10.1074/jbc.C500395200

- Song, J. J., & Kingston, R. E. (2008). WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *J Biol Chem*, 283(50), 35258-35264. doi: 10.1074/jbc.M806900200
- Southall, S. M., Wong, P. S., Odho, Z., Roe, S. M., & Wilson, J. R. (2009). Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol Cell*, 33(2), 181-191. doi: 10.1016/j.molcel.2008.12.029
- Steward, M. M., Lee, J. S., O'Donovan, A., Wyatt, M., Bernstein, B. E., & Shilatifard, A. (2006). Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat Struct Mol Biol*, 13(9), 852-854. doi: 10.1038/nsmb1131
- Stoller, J. Z., Huang, L., Tan, C. C., Huang, F., Zhou, D. D., Yang, J., . . . Epstein, J. A. (2010). Ash2l interacts with Tbx1 and is required during early embryogenesis. *Exp Biol Med (Maywood)*, 235(5), 569-576. doi: 10.1258/ebm.2010.009318
- Takahashi, Y. H., Westfield, G. H., Oleskie, A. N., Trievel, R. C., Shilatifard, A., & Skiniotis, G. (2011). Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. *Proc Natl Acad Sci U S A*, 108(51), 20526-20531. doi: 10.1073/pnas.1109360108
- Tan, C. C., Sindhu, K. V., Li, S., Nishio, H., Stoller, J. Z., Oishi, K., . . . Gelb, B. D. (2008). Transcription factor Ap2delta associates with Ash2l and ALR, a trithorax family histone methyltransferase, to activate Hoxc8 transcription. *Proc Natl Acad Sci U S A*, 105(21), 7472-7477. doi: 10.1073/pnas.0711896105
- Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., & Patel, D. J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol*, 14(11), 1025-1040. doi: 10.1038/nsmb1338
- Thiel, A. T., Blessington, P., Zou, T., Feather, D., Wu, X., Yan, J., . . . Hua, X. (2010). MLL-AF9-induced leukemogenesis requires coexpression of the wild-type Mll allele. *Cancer Cell*, 17(2), 148-159. doi: 10.1016/j.ccr.2009.12.034
- Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., . . . Bird, A. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, 464(7291), 1082-1086. doi: 10.1038/nature08924
- Tremblay, V., Zhang, P., Chaturvedi, C. P., Thornton, J., Brunzelle, J. S., Skiniotis, G., . . . Couture, J. F. (2014). Molecular basis for DPY30 association to COMPASS-like and NURF complexes. *Structure*, 22(12), 1821-1830. doi: 10.1016/j.str.2014.10.002
- Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., & Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *Embo j*, 13(16), 3822-3831.
- Tyagi, S., Chabes, A. L., Wysocka, J., & Herr, W. (2007). E2F Activation of S Phase Promoters via Association with HCF-1 and the MLL Family of Histone H3K4 Methyltransferases. *Mol Cell*, 27(1), 107-119. doi: <http://dx.doi.org/10.1016/j.molcel.2007.05.030>
- Ullius, A., Luscher-Firzlaff, J., Costa, I. G., Walsemann, G., Forst, A. H., Gusmao, E. G., . . . Luscher, B. (2014). The interaction of MYC with the trithorax protein ASH2L promotes gene transcription by regulating H3K27 modification. *Nucleic Acids Res*, 42(11), 6901-6920. doi: 10.1093/nar/gku312
- van Nuland, R., Smits, A. H., Pallaki, P., Jansen, P. W., Vermeulen, M., & Timmers, H. T. (2013). Quantitative dissection and stoichiometry determination of the human

- SET1/MLL histone methyltransferase complexes. *Mol Cell Biol*, 33(10), 2067-2077. doi: 10.1128/mcb.01742-12
- Vastenhouw, N. L., Zhang, Y., Woods, I. G., Imam, F., Regev, A., Liu, X. S., . . . Schier, A. F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. *Nature*, 464(7290), 922-926. doi: 10.1038/nature08866
- Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T., & Noel, J. P. (2000). Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat Struct Biol*, 7(8), 639-643. doi: 10.1038/77929
- Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W., van Schaik, F. M., Varier, R. A., . . . Timmers, H. T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*, 131(1), 58-69. doi: 10.1016/j.cell.2007.08.016
- Voo, K. S., Carlone, D. L., Jacobsen, B. M., Flodin, A., & Skalnik, D. G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol Cell Biol*, 20(6), 2108-2121.
- Wan, M., Liang, J., Xiong, Y., Shi, F., Zhang, Y., Lu, W., . . . Songyang, Z. (2013). The trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells. *J Biol Chem*, 288(7), 5039-5048. doi: 10.1074/jbc.M112.424515
- Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., . . . Chang, H. Y. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, 472(7341), 120-124. doi: 10.1038/nature09819
- Wang, P., Lin, C., Smith, E. R., Guo, H., Sanderson, B. W., Wu, M., . . . Shilatifard, A. (2009). Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol Cell Biol*, 29(22), 6074-6085. doi: 10.1128/mcb.00924-09
- Wang, Y. L., Faiola, F., Xu, M., Pan, S., & Martinez, E. (2008). Human ATAC Is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. *J Biol Chem*, 283(49), 33808-33815. doi: 10.1074/jbc.M806936200
- Wang, Y. Y., Liu, L. J., Zhong, B., Liu, T. T., Li, Y., Yang, Y., . . . Shu, H. B. (2010). WDR5 is essential for assembly of the VISA-associated signaling complex and virus-triggered IRF3 and NF-kappaB activation. *Proc Natl Acad Sci U S A*, 107(2), 815-820. doi: 10.1073/pnas.0908967107
- Wang, Z., Song, J., Milne, T. A., Wang, G. G., Li, H., Allis, C. D., & Patel, D. J. (2010). Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to Cyp33 and HDAC-mediated repression. *Cell*, 141(7), 1183-1194. doi: 10.1016/j.cell.2010.05.016
- Wu, L., Lee, S. Y., Zhou, B., Nguyen, U. T., Muir, T. W., Tan, S., & Dou, Y. (2013). ASH2L regulates ubiquitylation signaling to MLL: trans-regulation of H3 K4 methylation in higher eukaryotes. *Mol Cell*, 49(6), 1108-1120. doi: 10.1016/j.molcel.2013.01.033
- Wu, L., Zee, B. M., Wang, Y., Garcia, B. A., & Dou, Y. (2011). The RING finger protein MSL2 in the MOF complex is an E3 ubiquitin ligase for H2B K34 and is involved in

- crosstalk with H3 K4 and K79 methylation. *Mol Cell*, 43(1), 132-144. doi: 10.1016/j.molcel.2011.05.015
- Wu, M., Wang, P. F., Lee, J. S., Martin-Brown, S., Florens, L., Washburn, M., & Shilatifard, A. (2008). Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol Cell Biol*, 28(24), 7337-7344. doi: 10.1128/mcb.00976-08
- Wu, M. Z., Tsai, Y. P., Yang, M. H., Huang, C. H., Chang, S. Y., Chang, C. C., . . . Wu, K. J. (2011). Interplay between HDAC3 and WDR5 is essential for hypoxia-induced epithelial-mesenchymal transition. *Mol Cell*, 43(5), 811-822. doi: 10.1016/j.molcel.2011.07.012
- Wysocka, J., Swigut, T., Milne, T. A., Dou, Y., Zhang, X., Burlingame, A. L., . . . Allis, C. D. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*, 121(6), 859-872. doi: 10.1016/j.cell.2005.03.036
- Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., . . . Allis, C. D. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, 442(7098), 86-90. doi: 10.1038/nature04815
- Xu, C., & Min, J. (2011). Structure and function of WD40 domain proteins. *Protein & Cell*, 2(3), 202-214. doi: 10.1007/s13238-011-1018-1
- Yang, Y. W., Flynn, R. A., Chen, Y., Qu, K., Wan, B., Wang, K. C., . . . Chang, H. Y. (2014). Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *Elife*, 3, e02046. doi: 10.7554/eLife.02046
- Yokoyama, A., Somervaille, T. C., Smith, K. S., Rozenblatt-Rosen, O., Meyerson, M., & Cleary, M. L. (2005). The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell*, 123(2), 207-218. doi: 10.1016/j.cell.2005.09.025
- Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D. J., Kitabayashi, I., . . . Cleary, M. L. (2004). Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol Cell Biol*, 24(13), 5639-5649. doi: 10.1128/mcb.24.13.5639-5649.2004
- Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A., & Korsmeyer, S. J. (1995). Altered Hox expression and segmental identity in Mll-mutant mice. *Nature*, 378(6556), 505-508. doi: 10.1038/378505a0
- Zeisig, B. B., Milne, T., Garcia-Cuellar, M. P., Schreiner, S., Martin, M. E., Fuchs, U., . . . Slany, R. K. (2004). Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization. *Mol Cell Biol*, 24(2), 617-628.
- Zelevnik-Le, N. J., Harden, A. M., & Rowley, J. D. (1994). 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. *Proc Natl Acad Sci U S A*, 91(22), 10610-10614.
- Zhang, J., Yang, P. L., & Gray, N. S. (2009). Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer*, 9(1), 28-39. doi: 10.1038/nrc2559
- Zhang, P., Lee, H., Brunzelle, J. S., & Couture, J. F. (2012). The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Res*, 40(9), 4237-4246. doi: 10.1093/nar/gkr1235

- Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., & Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. *Mol Cell*, 12(1), 177-185.
- Zhao, X., Su, J., Wang, F., Liu, D., Ding, J., Yang, Y., . . . Jin, J. (2013). Crosstalk between NSL histone acetyltransferase and MLL/SET complexes: NSL complex functions in promoting histone H3K4 di-methylation activity by MLL/SET complexes. *PLoS Genet*, 9(11), e1003940. doi: 10.1371/journal.pgen.1003940

CONTRIBUTIONS OF COLLABORATORS

List of Affiliations

1. Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, Ontario K1H 8M5, Canada
2. Feinberg School of Medicine, Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, Chicago, Illinois 60611, USA
3. The Sprott Center for Stem Cell Research, Regenerative Medicine Program, Ottawa Hospital Research Institute, Ottawa, Ontario K1H 8L6, Canada
4. Department of Cellular and Molecular Medicine, University of Ottawa, Ontario K1H 8L6, Canada
5. Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109, USA
6. Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA
7. Department of Biochemistry and Molecular Genetics, Northwestern University, Chicago, Illinois 60611, USA

List of Contributors

1. Jean-François Couture¹

Dr. Couture is my PhD thesis supervisor that immensely contributed to solving the structures of all interactions and finalized all structural figures (Figures 2.3, 2.4A, 3.2, 3.3, 4.1, 4.3CE and 4.4A) in addition to Tables 2.1 and 3.1. He also contributed to the writing, editing and finalizing of all manuscripts. Lastly, Dr. Couture contributed to the critical reading and editing of my thesis.

All Manuscripts

2. Joseph Brunzelle²

Dr. Brunzelle is a professor at Northwestern University that collected full diffraction data sets for solving the structure of all interactions.

Manuscripts One (MLL1-WDR5-RbBP5) and Three (RbBP5-ASH2L)

3. Véronique Tremblay¹

Véronique is our lab technician that performed the GST pull down assays for Figure 2.2ABC in manuscript one and performed additional immunoprecipitation (IP) experiments investigating the role of RbBP5 residues in ASH2L binding in HEK293 cells (Figure S4.3B).

Manuscript One (MLL1-WDR5-RbBP5)

4. Vanja Avdic¹

Vanja was a Master's student in our lab that performed all preliminary experiments and immensely contributed to the writing of manuscript one.

5. Sylvain Lanouette¹

Sylvain is a PhD student in our lab that performed the RbBP5 sequence alignment for Figure 2.1C.

6. Adam Groulx¹

Adam was a 2010 summer research student that crystallized WDR5 in complex with the MLL1 and RbBP5 peptides.

Manuscript Two (WDR5-KMT2)

7. Hwabin Lee¹

Hwabin was my 2011-2012 fourth year Honours student that helped me purify various protein constructs for investigating the role of WDR5 in scaffolding the mammalian KMT2 enzymes.

Manuscript Three (RbBP5-ASH2L)

8 and 9. Chandra-Prakash Chaturvedi^{3,4} and Marjorie Brand^{3,4}

Dr. Chaturvedi was a postdoctoral fellow in Dr. Brand's lab at the Ottawa Hospital Research Institute that performed all experiments investigating the role of RbBP5-ASH2L interaction in erythroid cell terminal differentiation (Figure 4.2CD).

10. Myriam Cramet¹

Myriam was our lab technician that performed the IP experiments investigating the role of RbBP5 S350 in ASH2L binding in HEK293 cells and finalized Figure 4.3B.

11. Georgios Skiniotis^{5,6}

Dr. Skiniotis is a professor at the University of Michigan that collaborates with our lab on determining the interactions within the core complex using cryo-EM. He also made Figure 1.3.

12. Ali Shilatifard⁷

Dr. Shilatifard is currently a professor at Northwestern University who invited me to his lab then at the Stowers Institute for Medical Research to investigate the role of RbBP5-ASH2L interaction in the budding yeast using ChIP-seq (September to December 2013).

Pamela Peng Zhang

EDUCATION

Doctorate in Philosophy (PhD) in Biochemistry (CGPA 9.5/10, defended on June 2, 2015)
University of Ottawa, Ottawa, ON

Thesis topic: To determine the interactions underlying formation of the Mixed Lineage Leukemia (MLL) core complex, a histone H3 lysine 4 (H3K4) methyltransferase complex that is required for hematopoiesis and involved in acute forms of leukemia.

- Published in 7 peer-reviewed scientific journals with 5 as first author including *Genes & Development* and *Nucleic Acids Research*
- Determined the crystal structure of 8 protein sub-complexes within the MLL core complex
- Presented research findings at 17 institutional and international conferences
- Developed and optimized purification protocols; purified novel protein constructs to crystallization quality
- Trained and supervised 4 undergraduate students; teaching assistant for a graduate class
- Received scholarship of \$105,000 over 3 years (Canadian Institutes of Health Research (CIHR) Frederick Banting and Charles Best Canada Graduate Scholarships Doctoral Award 2012-2015)
- Received scholarship of \$34,600 over 2 years (Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Training Experience Program Scholarship 2010-2012)
- University of Ottawa Admission Scholarship for graduate studies (2010-2015)
- Department of Biochemistry, Microbiology and Immunology Syed Sattar PhD Student Award (2015)
- Faculty of Medicine Fisher Scientific Biochemistry Program Award of Excellence (2011)

Bachelor of Science, Honors in Biopharmaceutical Sciences, Genomics Option (Class of 2010)

University of Ottawa, Ottawa, ON

- Graduated cum laude
- Faculty of Science Dean's Honor List (2010)
- University of Ottawa Admission Scholarship for undergraduate studies (2005)

WORK EXPERIENCE

Visiting Research Student

September – December 2013

Stowers Institute for Medical Research, Kansas City, MO

- Determined the effect of mutants disrupting MLL complex assembly and H3K4 methylation in the budding yeast using ChIP-seq and western blot

- Independently designed and conducted experiments including the generation of yeast mutant strains
- Received \$4,000 to support visiting research (University of Ottawa Faculty of Graduate and Postdoctoral Studies Student Mobility Scholarship)

Summer Research Student

May – August 2009, 2010

University of Ottawa, Ottawa, ON

- Purified MLL-associated proteins for crystallization using affinity chromatography and ÄKTA purification systems
- Characterized proteins and protein complexes using structural, biochemical and biophysical techniques

Research Analyst

September – December 2008

Health Canada, Ottawa, ON

- Researched, developed and presented a detailed report on the environmental exposure of nanopharmaceuticals tailored for government publication
- Assisted in the development of a Best Management Practices (BMPs) guideline for nanopharmaceuticals

Research Student

January – August 2008

ViroChem Pharma Inc. (acquired by Vertex Pharmaceuticals Inc.), Laval, QC

- Synthesized and characterized compounds that inhibit specific activities of the Hepatitis C virus
- Received a NSERC Undergraduate Student Research Award to support research

Medical Assistant

August 2006 – December 2007

Appletree Medical Group Inc., Ottawa, ON

- Reduced the wait time of walk-in patients by triaging patients efficiently
- Trained and qualified to administer flu and allergy injections

PEER-REVIEWED PUBLICATIONS

1. **Zhang, P.**, Chaturvedi, C. P., Tremblay, V., Cramet, M., Brunzelle, J. S., Skiniotis, G., Brand, M., Shilatifard, A., & Couture, J. F. (2015). A phosphorylation switch on RbBP5 regulates histone H3 Lys4 methylation. *Genes Dev*, 29(2), 123-128. (Impact factor: 12.6)
2. Tremblay, V., **Zhang, P.**, Chaturvedi, C. P., Thornton, J., Brunzelle, J. S., Skiniotis, G., Shilatifard, A., Brand, M., & Couture, J. F. (2014). Molecular Basis for DPY-30 Association to COMPASS-like and NURF Complexes. *Structure*, 22(12), 1821-1830. (Impact factor: 6.8)
3. **Zhang, P.**, Bergamin, E., & Couture, J. F. (2013). The many facets of MLL1 regulation. *Biopolymers*, 99(2), 136-145. (Impact factor: 2.3)
4. **Zhang, P.**, Lee, H., Brunzelle, J. S., & Couture, J. F. (2012). The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Res*, 40(9), 4237-4246. (Impact factor: 8.8)

5. Sarvan, S., Avdic, V., Tremblay, V., Chaturvedi, C. P., **Zhang, P.**, Lanouette, S., Blais, A., Brunzelle, J. S., Brand, M., & Couture, J. F. (2011). Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain. *Nat Struct Mol Biol*, 18(7), 857-859. (Impact factor: 11.6)
6. Avdic, V.*, **Zhang, P.***, Lanouette, S., Voronova, A., Skerjanc, I., & Couture, J. F. (2011). Fine-tuning the stimulation of MLL1 methyltransferase activity by a histone H3-based peptide mimetic. *FASEB J*, 25(3), 960-967. (Impact factor: 5.5) *Co-first author
7. Avdic, V.*, **Zhang, P.***, Lanouette, S., Groulx, A., Tremblay, V., Brunzelle, J., & Couture, J. F. (2011). Structural and biochemical insights into MLL1 core complex assembly. *Structure*, 19(1), 101-108. (Impact factor: 6.8) *Co-first author

CONFERENCES AND PRESENTATIONS AS GUEST SPEAKERS

Title: A phosphorylation switch on RbBP5 regulates histone H3 Lys4 methylation.

1. **Faculty of Medicine Group for Research in Inflammation and Pathogenesis Poster Symposium**, University of Ottawa, Ottawa, Canada, August 2014 (faculty)
2. **3rd China-Canada Symposium on Systems Biology and Personalized Medicine**, Shanghai, China, May 2014 (international) *Travel grant from the Ottawa Institute of Systems Biology (OISB)*
3. **Department of Biochemistry, Microbiology and Immunology (BMI) Graduate Seminar Day**, University of Ottawa, Ottawa, Canada, March 2014 (departmental)
4. **Invited student speaker at the OISB Research Symposium**, Mont-Tremblant, Canada, June 2013 (institutional) *CIHR Training Program in Neurodegenerative Lipidomics (CTPNL) Conference Travel Award*

Title: Probing the plasticity of WDR5 peptide binding clefts.

5. **Abcam Conference on Chromatin: Structure and Function**, Seven Mile Beach, Grand Cayman, November 2013 (international) *Travel grants from the University of Ottawa*
6. **Centre for Catalysis Research and Innovation Annual Meeting**, University of Ottawa, Ottawa, Canada, September 2013 (institutional)
7. **BMI Graduate Poster Day**, University of Ottawa, Ottawa, Canada, May 2013 (departmental) *2nd place*
8. **5th Annual Groupe de Recherche Axé sur la Structure des Protéines (GRASP) Symposium**, McGill University, Montréal, Canada, November 2012 (national)

Title: The plasticity of WDR5 peptide binding cleft enables the binding of the SET1 family of histone methyltransferases.

9. **Invited student speaker at the 19th Methods in Protein Structure Analysis and the 2nd China Canada Systems Biology Conference**, Ottawa, Canada, June 2012 (international) *CTPNL Conference Travel Award*
10. **BMI Graduate Seminar Day**, University of Ottawa, Ottawa, Canada, February 2012 (departmental)
11. **Abcam Conference on Chromatin: Structure and Function**, Palm Beach, Aruba, December 2011 (international) *Travel grants from the University of Ottawa*

12. **4th Annual GRASP Symposium**, McGill University, Montréal, Canada, November 2011 (national)

Title: Structural and biochemical insights into MLL1 core complex assembly.

13. ***Invited student speaker at the OISB Science Retreat***, Montebello, Canada, June 2011 (institutional)
14. **BMI Graduate Poster Day**, University of Ottawa, Ottawa, Canada, May 2011 (departmental) *2nd place*
15. **1st Canadian Conference on Epigenetics: Epigenetics, Eh!**, London, Canada, May 2011 (international)
16. ***Invited student speaker at the Biochemistry RNA Club***, University of Ottawa, Ottawa, Canada, April 2011 (departmental)
17. **3rd Annual GRASP Symposium**, McGill University, Montréal, Canada, November 2010 (national)

COPYRIGHT PERMISSIONS

Chapter 2

Supplier	Elsevier Limited
Registered Company Number	1982084
Customer name	Pamela Zhang
License number	3646821243396
License date	Jun 12, 2015
Licensed content publisher	Elsevier
Licensed content publication	Structure
Licensed content title	Structural and Biochemical Insights into MLL1 Core Co-assembly
Licensed content author	Vanja Avdic,Pamela Zhang,Sylvain Lanouette,Adam Groulx,Véronique Tremblay,Joseph Brunzelle,Jean-François Couture
Licensed content date	12 January 2011
Licensed content volume number	19
Licensed content issue number	1
Number of pages	8
Start Page	101
End Page	108
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Actual number of figures/tables/illustrations	11
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Original figure numbers	Figures 1, 2, 3, 4 and table 1
Title of your thesis/dissertation	Structural and Biochemical Dissection of the KMT2 Co-repressor
Expected completion date	Jun 2015
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD

VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD

Chapter 3

License Number	3646820547832
License date	Jun 12, 2015
Licensed content publisher	Oxford University Press
Licensed content publication	Nucleic Acids Research
Licensed content title	The plasticity of WDR5 peptide-binding cleft enables t the SET1 family of histone methyltransferases:
Licensed content author	Pamela Zhang, Hwabin Lee, Joseph S. Brunzelle, Jean Couture
Licensed content date	05/01/2012
Type of Use	Thesis/Dissertation
Institution name	None
Title of your work	Structural and Biochemical Dissection of the KMT2 Co
Publisher of your work	n/a
Expected publication date	Jun 2015
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Total	0.00 USD

Chapter 4

We retain copyright of the *Genes & Development* article because at the time this thesis is submitted, the article is still within six months after the full-issue publication date (http://genesdev.cshlp.org/site/misc/permissions_landing.xhtml).