Structural and functional dissection of the MLL1 histone methyltransferase complex

Vanja Avdic

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfilment of the requirements
for the MSc degree in Biochemistry

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

ABSTRACT

The mixed lineage leukemia (MLL) proteins regulate an array of developmental and differentiation processes. Similar to other members of the SET1 family, association of MLL1-4 with Ash2L, RbBP5 and WDR5, collectively termed the MLL core complex, is required for MLL mediated histone H3 Lys-4 di/tri-methylation. Each member of the core complex has a unique role in modulating the activity of MLL1. WDR5 is key in nucleating the formation of the core complex by acting as a structural scaffold, whereas Ash2L and RbBP5 are responsible for stimulating MLL methyltransferase activity. Currently, the structural and biochemical mechanisms utilized by the core complex to regulate MLL1 activity are unknown. Through structural and biochemical dissection of the core complex we have assigned specific functions to core complex subunits and have identified the minimal structural requirements for methyltransferase activity. Furthermore, through structure based drug design, we have identified a peptidomimetic inhibitor of MLL1 methyltransferase activity.

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge the support, guidance and mentorship of my supervisor, Dr. Jean-Francois Couture. All my successes are a direct result of his involvement in my training and I am very thankful for his infectious enthusiasm for teaching and science.

Several professors and students have been instrumental in the completion of my Master's and for that I would like to acknowledge my thesis advisory committee, Dr. Marjorie Brand and Dr. Natalie Goto and Dr. Alexandre Blais and Dr. Adam Rudner, who provided guidance as well. Our collaborators have provided excellent partnerships, ideas and reagents and for that I must acknowledge Dr. Marjorie Brand, Dr. Joseph S. Brunzelle, Dr. Chandra-Prakash Chaturvedi, Dr. Jeffery Dillworth, Dr. Jean-Philippe Lambert, Dr. Ilona Skerjanc and Anastassiva Voronova.

Additional thanks must be given to past members of the Couture Lab, Ehsan Hakimi, John-Paul Muggeridge and Yan-Mo Wang and particularly to Adam Groulx for the work he did that helped in the completion of our manuscripts. I am especially grateful to current members of the Couture Lab, Sylvain Lanouette, Sabina Sarvan, Veronique Tremblay and Pamela Zhang, for their friendship, support and especially for their dedication to our work.

I would especially like to thank everyone on the fourth floor of RGN for creating such a wonderful work environment, especially Dr. Corrie da Costa and everyone in the Baetz and Rudner labs.

Finally, this experience would not have been complete or possible without the support of my friends and family. Thanks to my parents, sister Maja and my "brother from another mother" Alain and friends, Andrea Lau, Jenny Liu, Leslie Mitchell, Sarmila Satunarajah, Ted Sewell, Jason Tetro, Anastassia Voronova and Darren Yip who have always provided support when needed and more importantly laughs and good times.

TABLE OF CONTENTS

List of abbreviations

Ac	Acetylated
AdoMet	Adenosyl methionine
ALR	Augmentor of liver regeneration
ALX1	ALX homeobox 1
Ap2d	Associated protein 2 delta
Ash2L	Absent, small or homeotic like protein 2
$Ash2L_C$	C-terminus of Ash2L
$Ash2L_N$	N-terminus of Ash2L
ATP	Adenosine triphosphate
bp's	Base pairs
cSET	C-terminus of SET domain enzyme
CAPN3	Calpain 3
CBP	Creb binding protein
CD	Circular dichroism
Ce	*Caenorhabditis elegans*
CHD1	Chromodomain-helicase-DNA-binding protein 1
ChIP	Chromatin immunoprecipitation
ckm	Creatine kinase
COOT	Crystallographic Object-Oriented Toolkit
COMPASS	A complex of proteins associated with trithorax related SET domain protein SET1
Cps35	Compass subunit 35
CyP33	Cyclophillin 33
DAVID	Database for Annotation, Visualization, and Integrated Discovery
DBD	DNA binding domain
DNA	Deoxyribonucleic acid
Dm	*Drosophila melanogaster*
DMSO	Dimethylsulfoxide
Dr	*Danio renio*
DTT	Dithiothreitol
EMSA	Electrophoretic mobility shift assay
Ezh1/2	Ehancer of Zeste homolog 1
FL	Full length
FOXO	Forkhead box
GM/CA-CAT	General Medicine and Cancer Institutes Collaborative Access Team
GST	glutathione S-transferase
H4K20	Histones and specific residues are denoted by one letter code
H4K20me	Methylated histone
H	Hinge
HC	Hinge and C-terminus
HDAC	Histone deactylase
HMG	High mobility group
HOX	Homeobox
HS	Hypersensitive

Hs	*Homo sapiens*
HWH	Helix wing helix
IC50	Inhibition constant
ING	Inhibitor of growth
ITC	Isothermal titration calorimetry
iSET	Inserted SET
ISWI	Imitation switch
JMJD2A	Jumonji protein D2A
LSMT	Large-Subunit MethyTransferase
LCR	Locus control region
LS-CAT	Life Science-Collaborative Access Team
MEL	Myeloid erythroleukemia
MEF2d	Myocyte enhancing factor
MLL	Mixed lineage leukemia
MyoG	Myogenin
NaH3	N-alpha acetylation histone H3 peptide
NCP	Nucleosome core particle
NF-E2	Nuclear factor – erythroid derived
NheJ1	Nonhomologous end-joining factor 1
nSET	N-terminus of SET domain enzyme
NURF	Nucleosome remodeling factor
P	Phosphorylated
p38	Protein 38
PAX-7	Paired box protein 7
PBS	Phosphate buffer saline
PBST	Phosphate buffer with Triton-X
PcG	Polycomb group
PCR	Polymerase chain reaction
PEG	Poly ethylene glycol
PH	Propeller and Hinge
PHD	Plant homeodomain
PWM	Position weight matrices
PTD	Partial tandem duplication
qPCR	Quantitative polymerase chain reaction
RbBP5	Retinoblastoma binding protein 5
r.m.s	Root mean squared
RNA	Ribonucleic acid
SAD	Single-wavelength anomalous dispersion
SAH	S-adenosylhomocysteine
SAM	S-adenosyl methionine
Sc	*Saccharomyces cerevisiae*
s.d.	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate, polyacrylamide gel electrophoresis
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SeMet	Selenomethionine
SET	Suppressor of position effect variegation, Enhancer of Zeste, Trithorax

Sp	*Schizosaccharomyces pombe*
SPRY	SPla and RYanodine Receptor
SV40	Simian virus 40
SUMO	Small ubiquitin like modifier
TAD	Trans-activation domain
TEV	Tobacco etch virus
TPP1	Tripeptidyl-peptidase I
Tre	Trithorax response element
Trx	Trithorax
SWI/SNF	SWItch/Sucrose Non-Fermentable
Ub	Ubiquitinated
ULP1	Ubiquitin like protein specific
WDR5	Tryptophan aspartate repeat 5 protein
WDR82	Tryptophan-aspartate repeat protein 82
WIN	WDR5 interaction motif

List of Figures

List of Tables

CHAPTER 1. INTRODUCTION

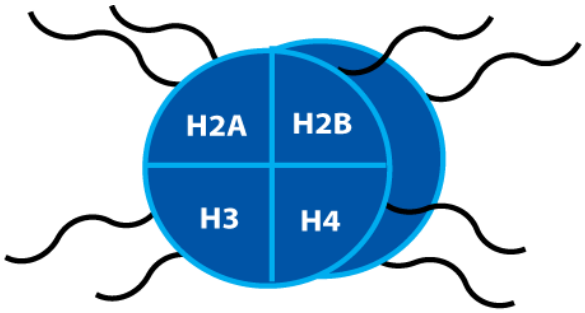
The structure and function of eukaryotic genomes is subject to multiple levels of regulation to ensure proper transcription, DNA replication, recombination and repair. Additionally, Multifaceted methods of control are required to ensure correct spatial and temporal gene expression. Furthermore, large eukaryotic genomes must be compacted in order to fit vast amounts of genetic material inside a cell. This requires the cell to carry out two opposing functions, compacting its genome but also preventing the packaging of DNA from hindering processes that require access to DNA. In order to accomplish this, the cell packages its genome into a structure termed chromatin.

1.1 Chromatin Structure and Function

1.1.1 The nucleosome core particle

A typical eukaryotic cell contains up to 2 m of linear DNA that must fit inside a cell's microscopic nucleus. In order for the DNA to fit, it is wrapped around a macromolecular complex termed the nucleosome core particle (NCP) (Davey et al, 2002; Kornberg, 1977). The NCP is composed of a histone octamer containing two copies of each core histone protein, H2A, H2B, H3 and H4 (Luger et al, 1997). Two heterodimers of H2A-H2B and an H3-H4 tetramer fold together to form a globular structure around which approximately 146 or 147 base pairs (bp) of DNA are wound (Richmond & Davey, 2003) (Fig. 1a). The first step of DNA compaction is achieved when NCP's are adjoined to the next by linker DNA of variable length to form a "beads on a string" structure (Fig. 1b).

a)



b)

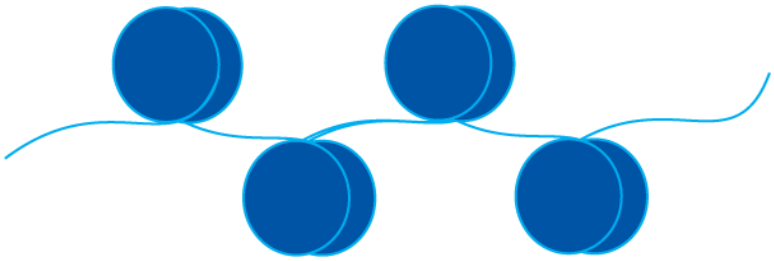


Fig. 1. Chromatin Structure

a) The building block of chromatin is the nucleosome core particle (NCP). It is composed of two copies of each core histone protein, H2A, H2B, H3 and H4. The histones fold together to form a globular structure around which 146 or 147 bp of DNA are wound. Note that the N-terminal tails of the histone proteins (shown in black) protrude out from the nucleosome core.
b) NCP's are linked together by linker DNA of variable length (blue) to form the basal chromatin structure often described as "beads on a string". This structure serves as a template for the formation of higher order chromatin structures.

1.1.2 Nucleosome positioning

The NCP serves to protect DNA that is wrapped around the nucleosome core but exposes linker DNA found between adjacent nucleosomes (Arya et al, 2010). From these observations, it was hypothesized that genes found within linker DNA are actively transcribed because they are highly accessible to transcriptional machinery. In contrast, inclusion of a DNA sequence within the histone octamer limits accessibility to the gene and thereby represses or prevents its transcription. To support this idea, transcription factor binding sites and DNA hypersensitive (HS) sites are often located in the linker rather than the nucleosomal DNA (Ercan & Simpson, 2004; Yuan et al, 2005). Attempts to map the global positions of nucleosomes in several organisms (Lee et al, 2007; Mavrich et al, 2008; Travers et al, 2009; Valouev et al, 2008) have concluded that although nucleosomes have defined positions across a genome, they are also highly dynamic and change position along the DNA fiber (Luger, 2006; Rando & Ahmad, 2007). A striking feature of nucleosome dynamics is that transcription start sites (TSS) often lack or are depleted of NCP's (Bernstein et al, 2004; Boeger et al, 2003). To further demonstrate the flexibility of NCP's, upon alteration of transcription programs in yeast via a change of carbon source, the expression of new genes was facilitated by a loss of nucleosome structures whereas newly silenced genes demonstrated an increase in nucleosome structures (Lee et al, 2004).

1.1.3 Chromatin structure and gene regulation

From the primary arrangement of "beads on a string" and with the aid of linker histone, (Allan et al, 1986; Raghuram et al, 2009) adjacent nucleosomes interact to form higher order and more condensed chromatin structures (Luger & Hansen, 2005). The basal chromatin structure achieves a 5-fold compaction of DNA and starting from this template the compaction can be increased through the formation of the 30nm fiber (Schalch et al, 2005) and further tertiary

structures, such as the mitotic chromosome (Belmont, 2006). Classically, chromatin has been divided into either euchromatin or heterochromatin (Passarge, 1979), based on its staining within the nucleus (Trojer & Reinberg, 2007). Later it was shown that the highly condensed heterochromatic regions of chromatin inhibit gene expression while actively transcribed genes are found in less condensed euchromatic regions (Hsu, 1962). Identification of regions that transition between euchromatic and heterochromatic structures gave birth to a third class of chromatin, the facultative heterochromatin.

1.1.4 Chromatin remodeling

Several mechanisms lead to changes in nucleosome positioning or chromatin structure. These mechanisms include nucleosome sliding, chromatin remodeling, and post-translational modifications of histone proteins (Campos & Reinberg, 2009). Nucleosome sliding is a change in the position of the nucleosome relative to its position on DNA. This mechanism serves to expose previously protected DNA and facilitates expression of genes that were once protected by the nucleosome. *In vitro,* nucleosome sliding occurs spontaneously and is also temperature dependent (Pennings et al, 1991). The *in vivo* mechanisms that determine nucleosome positioning are dependent on DNA sequence (Travers & Klug, 1987) and require the action of ATP dependent chromatin remodelers.

ATP dependent chromatin remodeling complexes alter chromatin structure by utilizing energy from ATP hydrolysis to disrupt histone-DNA contacts. Specific complexes have been identified that either produce a transcriptionally competent or repressive chromatin structure (Gangaraju & Bartholomew, 2007; Lusser & Kadonaga, 2003). Of the most studied complexes, the SWItch/Sucrose Non-Fermentable (SWI/SNF) proteins have been shown to disrupt nucleosome structures and aid in promoter activation (Kwon et al, 1994) whereas the Imitation

Switch (ISWI) complexes promote nucleosome positioning that represses gene expression (Corona et al, 2007). Multi-subunit ATP-dependent chromatin remodeling complexes are frequently associated with other important enzymes that play a critical role in indirectly controlling the structure of chromatin. These include enzymes that catalyze the addition of various chemical moieties to histone proteins.

Section 1.2 – Post-translational modifications of histone proteins

1.2.1 Histone acetylation, phosphorylation and ubiquitination

Histone proteins are subject to a large array of post-translational modifications. These include acetylation, methylation, phosphorylation and ubiquitination (Fig. 2) (Kouzarides, 2007). The modifications are predominately found on the N- or C-terminal tails of histone proteins (Grant, 2001). These tails protrude out form the nucleosome core (Fig. 1a) and are easily accessible to chromatin modifying enzymes. In addition, several modifications have been reported in the core of histone proteins (Mersfelder & Parthun, 2006).

Histone lysine acetylation is strongly correlated to transcriptional activation (Brownell et al, 1996). The added acetyl group introduces a negative charge to the nucleosome environment and weakens DNA-histone contacts, which thereby causes charge repulsion between neighboring nucleosomes (Bauer et al, 1994; Libertini et al, 1988). Both of these effects lead to a less compact chromatin structure and facilitate access to DNA to allow for transcriptional activation. Additionally, acetylated lysine residues provide binding sites for proteins, including transcription factors (Dhalluin et al, 1999), harboring bromo domains, which are known to specifically recognize acetylated lysine residues (Zeng & Zhou, 2002).

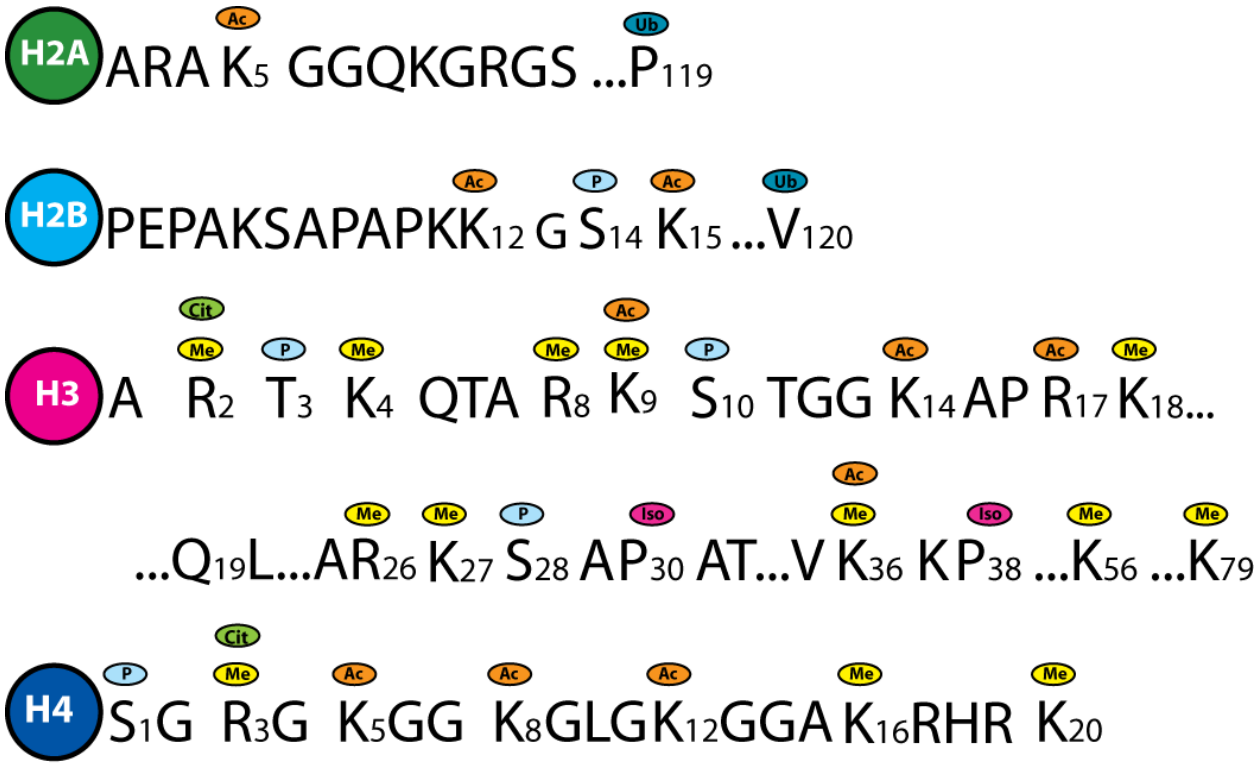


Fig. 2 – Post-translational modifications of histone proteins.

Histone proteins are subject to a wide array of site specific modifications including acetylation (ac), methylation (me), phosphorylation (p) and ubiquitination (ub).

Similar to lysine acetylation, phosphorylation of histone proteins were originally mapped to actively transcribed regions of chromatin (Nowak & Corces, 2000). However, depending on the phosphorylation site, this mark also signals chromosome condensation and segregation during the cell cycle (Pérez-Cadahía et al, 2009). Akin to phosphorylation, histone ubiquitination is involved in a plethora of biological processes (Osley, 2004; Zhou et al, 2009). H2A ubiquitination has been associated with transcriptionally inactive chromatin (Wang et al, 2004) and DNA damage response (Bergink et al, 2006; Kapetanaki et al, 2006) while H2B ubiquitination, in yeast, has been found in the promoter regions of actively transcribed genes (Xiao et al, 2005b). Interestingly, ubiquitinating protein complexes have been recently shown to be heavily regulated in trans, by another covalent modification, namely histone lysine methylation (Nakanishi et al, 2008; Shukla et al, 2008).

1.2.2 Histone methylation

Historically, methylation was one of the first covalent modifications found on histone proteins. In a series of fractionation experiments using calf thymus, Allfrey et al. found that a significant fraction of histone proteins were methylated (Allfrey et al, 1964). Further analyses revealed that arginine and lysine residues are the more prominently methylated amino acids (Paik et al, 2007). Despite the importance of this discovery, the role of these marks long stood unresolved, as the methyltransferases catalyzing these reactions remained unknown. However, since the characterization of the first histone lysine methyltransferases (Jones & Gelbart, 1993; Stassen et al, 1995; Tschiersch et al, 1994) a large number of methylation sites have been found which include Lys-4, Lys-9, Lys-27, Lys-36 and Lys-79 on histone H3, Lys-20 on histone H4 and Lys-26 on histone H1 (Qian & Zhou, 2006). Consistent with their important roles, methylation of a specific lysine residue is linked to a specific set of biological processes (Fig. 3).

In addition, contrasting with phosphorylation and acetylation, lysine methylation displays another level of specificity. In fact, the ε-amine of lysine residue can be mono, di or trimethylated with different cellular consequences for each level of methylation. Following the discovery of the first histone lysine methyltransferase in 2000, several enzymes have been functionally characterized (Xiao et al, 2003). These enzymes contain structural features that allow them to site specifically methylate specific lysine residues on histone proteins as well as transfer a specific number of methyl groups.

1.2.3 Histone lysine methyltransferase: the SET domain enzymes.

In 2000, Jenuwein and colleagues established the link between histone methylation and SET domain enzymes (Rea et al, 2000). Based on the sequence homology between various SET domain proteins and the Large-Subunit Methyltransferase (LSMT) of Rubisco from *Arabidopsis thaliana*, they found that SET domain enzymes were able to methylate specific lysine residues on histone proteins.

The SET domain, an acronym for the first three drosophila proteins in which the domain was identified, namely **S**uppressor of position effect variegation, **E**nhancer of zeste and **T**rithorax (Jones & Gelbart, 1993; Stassen et al, 1995; Tschiersch et al, 1994), are a family of enzymes that catalyze histone lysine methylation via the transfer of the methyl group from S-adenosyl methionine (SAM) to a target lysine's ε-amine group. SET domain enzymes are divided into families, depending on their substrate specificity and similarities in amino acid sequence and structure (Kouzarides, 2002). Despite differences between the families, the structure of a SET domain is divided, structurally, into four functional regions which include the N-terminus (nSET), the inserted SET (iSET), SET and the C-terminus (cSET) (Cheng et al, 2005) (Fig. 4a). Each region plays an important role in SET enzyme's function. The nSET is

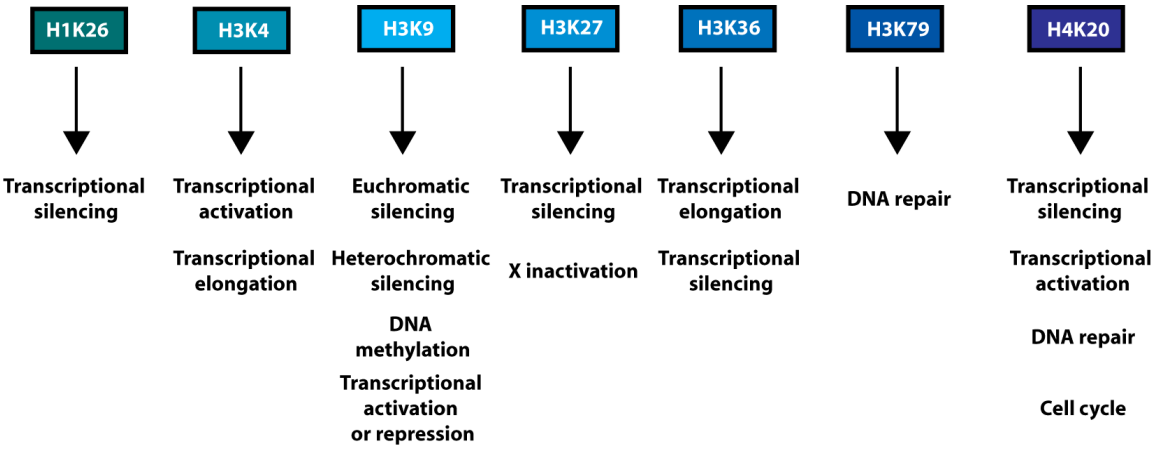


Fig. 3 – Histone lysine methylation and associated biological processes.
Figure adapted from Qian and Zhou, 2006.

important for the structural stability of the SET domain (Couture et al, 2005; Xiao et al, 2005a). The iSET and cSET participate in substrate and cofactor binding. The cSET domain contains the highly conserved NHS sequence (Fig. 4b) and provides key structural elements for the binding of the cofactor. Of all the regions, the iSET is particularly important for substrate recognition and each iSET is unique. This particularity is an important aspect conferring to a SET domain enzyme its specificity for a particular histone. The SET domain itself is highly conserved in both amino acid sequence and structure (Fig. 4) and aids in providing the structural determinants controlling the methylation of the target lysine (Dillon et al, 2005). Upon binding of the histone substrate and cofactor, a structural rearrangement occurs in the SET domain that leads to the formation of a methyltransfer pore. This pore helps to position the target lysine ε-amine and SAM sulfonium group in a 180º alignment to accommodate the methyltransfer reaction. (Zhang & Bruice, 2008). Finally, the SET domain active site architecture ensures that a specific number of methyl groups are transferred during the reaction (Couture et al, 2008; Del Rizzo et al, 2010), a feature termed the SET domain product specificity (Zhang et al, 2003). Enzymes that catalyze the transfer of a single methyl group contain a structurally conserved active site tyrosine (Couture etal, 2005;Xiao et al, 2003) that tightly coordinates a water molecule which thereby prevents rotation of the methylated lysine residue within SET domain active site. SET domain enzymes which di or trimethylate lysine residues characteristically harbor a non-tyrosine active site residue (Zhang et al, 2003). Without the tyrosine's hydroxyl group, a key hydrogen bond is lost which allow the eviction of the active site water molecule. The space created can therefore be occupied by the monomethylated product in an orientation conducive for the transfer of additional methyl groups (Couture et al, 2008; Del Rizzo et al, 2010).

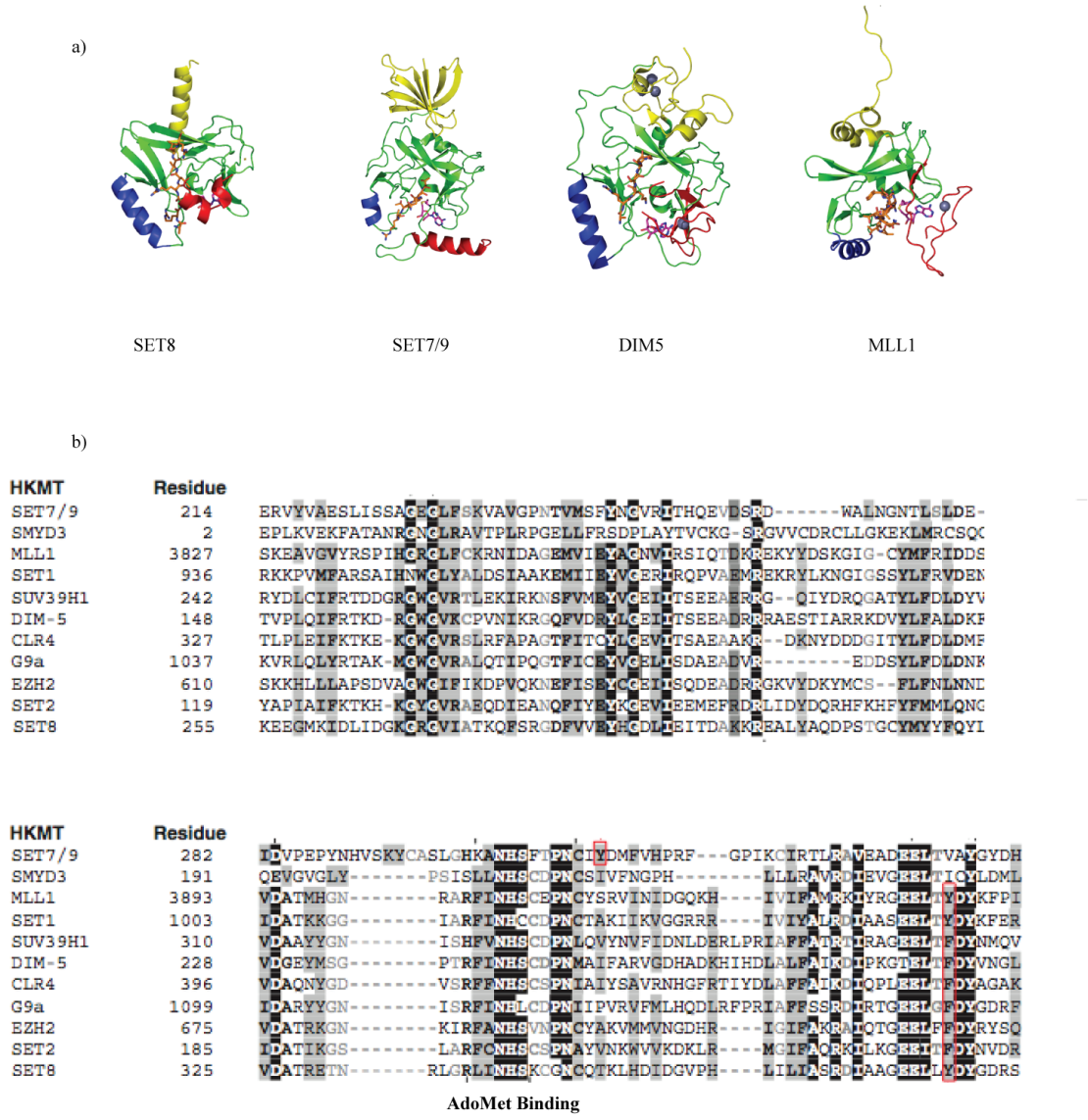


Fig. 4 – Structure and alignment of SET domain enzymes.

a) Each SET domain enzyme is divided into four functional regions: the nSET (yellow), iSET (blue), SET (green) and cSET (red).

b) Alignment of SET domain enzymes highlighting the highly conserved NHS sequence used for SAM binding, the Tyrosine/Phenylalanine switch (red box) for product specificity and high conserved active site motifs and tyrosines. Adapted from Dillon et al. 2005.

1.2.4. – H3K4 methylation

Distinct patterns of histone methylation occur across the genome (Barski et al, 2007), with modifications linked to various biological processes (Fig. 3). However, given that the focus of my thesis is on histone H3 Lys-4 methyltransferases, I will strictly focus on this mark. Methylation of H3K4 (from hereon, histones and specific amino acid residues will be denoted with one-letter code) was first reported 35 years ago (Honda et al, 1975). Despite the novelty of these findings, it's only with the development of H3K4me1/2/3 specific antibodies that the roles and the distribution of these marks were established. For example, the monomethylated form of Lys-4 was mapped at the 3' ends of genes (Pokholok et al, 2005) while in yeast its dimethylated form is broadly distributed along the body of the genes (Ng et al, 2003; Santos-Rosa et al, 2002) and co-localized with H3K4me3 in mammals (Bernstein et al, 2002; Schneider et al, 2004b). Although the role of H3K4me1/2 is unclear (Pinskaya & Morillon, 2009), punctuate enrichment of H3K4me3 was found at the 5' end of actively transcribed genes.

H3K4me3 is enriched in the 5' region of most actively transcribed genes in several organisms (Bernstein et al, 2002; Santos-Rosa et al, 2002) and linked to the expression of key developmental genes (Eissenberg & Shilatifard, 2010). In yeast, this mark is associated with transcriptional initiation, whereas in mammals it participates in the recruitment of the transcriptional machinery (Wang et al, 2009). How this mark directs transcriptional activation remains poorly understood, however indirect evidence have shown that several ATP-dependent chromatin-remodeling complexes are able to recognize the H3K4me3 mark (Ruthenburg et al, 2007; Sims & Reinberg, 2006). The Chromodomain-helicase-DNA-binding protein 1 (CHD1) and nucleosomal remodeling factor (NURF) chromatin remodelers, which function in stimulating gene expression (Mizuguchi et al, 1997), binds H3K4me3 with effector domains

including Chromo and plant homeo (PHD) domains, respectively (Flanagan et al, 2005; Taverna et al, 2007; Wysocka et al, 2006). Additionally, several other histone-modifying proteins have been shown to bind the H3K4me3 mark. These include the Inhibitor of Growth (ING) tumor suppressors (Shi & Gozani, 2005) histone demethylase JMJD2A (Huang et al, 2006). Both mammalian and yeast ING proteins associate with and modulate the activity of both histone acetylase and deacetylase complexes (Shi & Gozani, 2005). The ING subunits of histone acetylase complexes have been reported to directly bind H3K4me2/3 (Peña et al, 2006; Shi et al, 2006). This association provides an indirect link between H3K4 methylation and the concurrent hyperacetylation of histones reported at sites of H3K3me3 enrichment (Pokholok et al, 2005; Schübeler et al, 2004). The ING2 PHD domain also binds H3K4me2/3 but associates with a histone deactylase complex, mSin3a-HDAC1 (Shi et al, 2006). This study demonstrated a novel role for H3K4 methylation in histone deactylase mediated gene repression as a response to DNA damage (Shi et al, 2006). Finally, although several studies have defined the determinants of JMJD2A recognition of H3K4me3 (Huang et al, 2006; Kim et al, 2006), the biological consequence of this association is poorly understood (Sims & Reinberg, 2006). Although the mechanism through which H3K4me3 regulates transcription, the enzymes responsible for catalyzing the methylation of H3K4 have been well characterized and studies of these enzymes have provided further insight into the biological role of this histone modification.

1.3 – MLL1 Structure and function

1.3.1 – The SET1 family

The bulk of H3K4 di/tri-methylation in mammals is carried out by the SET1 family, which is composed of SET1A, SET1B and MLL1-5 (Fig. 5). Despite sharing similarities in

substrate specificity, many studies have shown that the MLL's (mixed lineage leukemia) are not functionally redundant and have a unique role during various developmental programs such as hematopoiesis, adipogenesis and neurogenesis (Ansari & Mandal, 2010). An additional shared feature of the members of the SET1 family is that each, with the exception of MLL5 (Wu et al, 2008), function in multi-protein complexes**.** Amongst the SET1 multi-protein complexes, a common and highly conserved tripartite complex consisting of the absent, small or homeotic like 2 (Ash2L) protein, retinoblastoma binding protein 5 (RbBP5) and the Tryptophan-Aspartate repeat protein 5 (WDR5) (Dou et al, 2006), also referred as the MLL core complex, is found in each. MLL1 is a positive regulator of *HOX* gene expression during embryonic development (Yu et al, 1998; Yu et al, 1995) and in adult hematopoietic stem cells (Jude et al, 2007). *Mll1* knockout mice are embryonic lethal and heterozygous *Mll1* mutants display hematopoietic abnormalities due to a decreased in *Hox* gene expression (Yu et al, 1995). One of the mechanisms controlling *Hox* gene expression is mediated by MLL1's methyltransferase activity (Milne et al, 2005). Indeed, mice expressing an MLL1 protein lacking a SET domain are viable but have skeletal muscle defects and show a sharp decrease in *Hox* gene expression and H3K4 trimethylation (Terranova et al, 2006a). More recently, MLL1 methyltransferase activity has been shown to be essential for additional development and differentiation processes such as myogenesis (McKinnell et al, 2008) and neurogenesis (Lim et al, 2009). MLL1 is associated with severe and aggressive leukemias (Hess, 2004). Two mechanisms have been identified that allow the MLL gene to become oncogenic (Fig. 6), MLL fusion partners (Slany, 2009) and MLL gene partial tandem duplications (Basecke et al, 2006). Firstly, approximately 73 MLL fusion partners have been identified (Fig. 6a). The MLL fusion partner replaces the C-terminus of the

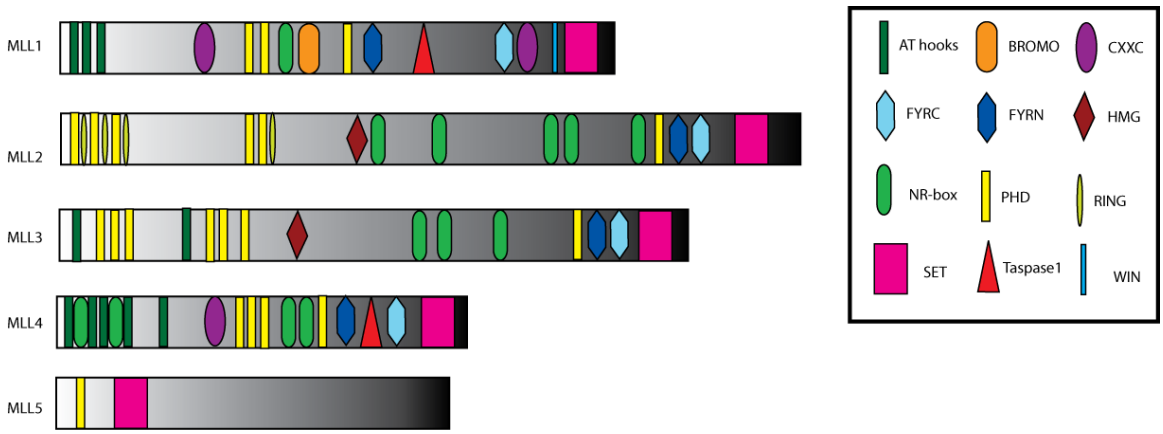


Fig. 5 - Structure of MLL1-5

Each MLL protein is composed of several functional domains. Each contains a C-terminal SET domain except for MLL5 where the SET domain is located at the N-terminus.

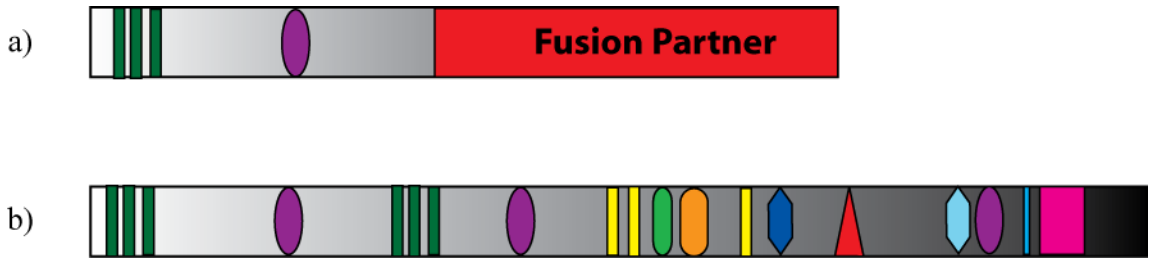


Fig. 6 – Mechanisms of MLL mediated leukemias.

Two mechanisms have been identified that lead to MLL mediated leukemias.

a) Several chromosomal translocations lead to an oncogenic fusion partner being added to the MLL C-terminus.
b) MLL gene partial tandem duplications result in an MLL protein with a duplicated N-terminus.

proteins and a hallmark of these cancers is the up-regulation of *HOX* genes (Martin et al, 2003; So et al, 2003). The most difficult cases of MLL related leukemias are due to gene partial tandem duplications (PTD's) of the MLL gene (Fig. 6b). In this chromosomal aberration, exons 3-9 of the MLL gene are duplicated and the translated product contains a duplicated amino terminus. Again, a trademark of these diseases is an increase in the expression of *HOX* genes as well as an increase in H3K4 methylation (Dorrance et al, 2006).

1.3.2 – MLL1 protein structure

MLL1 harbors several functional domains (Cosgrove & Patel, 2010) (Fig. 5). Immediately following translation, the protein is proteolytically cleaved into an N and C-terminal fragment, namely MLL_N and MLL_C (Hsieh et al, 2003). The domains regulate MLL1 recruitment to its target genes, transcriptional activation and methyltransferase activity. The MLL1 CXXC domain participates in recognition of active genes through the binding of non-methyl CpG DNA (Birke et al, 2002). The MLL trans-activation domain (TAD) enhances transcriptional activation through its interaction with transcriptional co-activators such as the Creb binding protein (CPB) (Goto et al, 2002). Furthermore, initiation and regulation of MLL mediated *HOX* gene expression is dependent on its PHD3 domain. Recognition of H3K4me3 by the MLL1 PHD3 is essential for MLL1 mediated gene activation (Chang et al, 2010). But, recognition of the H3K4me3 mark by MLL's PHD domain also serves to down regulate MLL mediated gene expression. This is achieved by recruitment of CyP33, a proline isomerase whose activity leads to a structural change in MLL1 that allows for binding of histone deacetylases (Wang et al, 2010). Finally, a C-terminal SET domain is responsible for its methyltransferase activity.

1.3.3 – The MLL1 SET domain

A key aspect of MLL mediated cellular processes is its H3K4 methyltransferase activity (Milne et al, 2005). Structural and sequence alignments between the MLL SET domain with other SET domains predict that MLL is a mono-methyltransferase. Consistent with these *in silico* predictions, the holoenzyme catalyzes the monomethylation of H3K4. However, these observations contrast with the *in vivo* results showing that MLL is a H3K4 di/tri-methyltransferase. Additionally, despite having all the necessary sequence and structural requirements to catalyze a methyltransferase reaction, the MLL SET domain is catalytically inefficient. On its own, the MLL1 SET domain catalyzes a monomethylation reaction with a k_1 value of 0.003 ± 0.0003 h^{-1} whereas this value increases to 0.96±0.07 h^{-1} if the core complex subunits are present (Patel et al, 2009). A comparison between the structure of MLL SET domain and other SET domain proteins has revealed several key structural differences underlying the poor catalytic efficiency of MLL1. Overall, the MLL SET domain is more "open" due to a shift in the position of its iSET (Southall et al, 2009a). Because of this shift, the MLL SET domain fails to align the H3 peptide with the cofactor. It was speculated thereafter that the MLL1 accessory proteins, namely the MLL1 core complex, induces a structural re-arrangement of the SET domain to promote the formation of a competent active site.

1.4 – The MLL Core Complex

1.4.1 – Stimulation of MLL1 methyltransferase activity by the core complex subunits.

The presence of the core complex subunits increases MLL1's kinetic activity 300-fold (Patel et al, 2009) *in vitro* and allows the MLL1 SET domain to catalyze a di/tri methylation reaction *in vivo* (Dou et al, 2006) (Fig. 7). Each component of the core complex has a key role in

promoting MLL1 methyltransferase activity and product specificity of all members of the SET1 family. Each core complex member is composed of several functional subunits (Fig. 8) and each has the potential to regulate a different aspect of MLL activity. The mechanism through which the core complex member regulates MLL1 methyltransferase activity is unknown although it is hypothesized that each subunit has a unique role in the formation, kinetics and product specificity of the enzyme.

1.4.2 – WDR5 Structure and Function

WDR5 is essential for drosophila (Hollmann et al, 2002) and vertebrate development (Wysocka et al, 2005). It is a key component of the MLL core complex and knockdown of WDR5 leads to a complete loss of detectable H3K4 methylation. Further underscoring its importance in histone methylation, morpholino mediated knock outs in *X. Laevis* leads to developmental defects and a global loss of H3K4 methylation. WDR5 belongs to the tryptophan-aspartate repeat β-propeller family. These proteins possess a unique structural fold that facilitates their role as a protein scaffold (Valeyev et al, 2008). The top, bottom and sides or a β-propeller are known to mediate and stabilize protein-protein interactions within multi-protein complexes. In the context of the MLL1 complex, WDR5 promotes H3K4 methyltransferase activity within the core complex through a direct interaction with MLL1 (Patel et al, 2008a; Song & Kingston, 2008a). A crystal structure of WDR5 in complex with the **W**DR5 **In**teraction (WIN) motif from MLL1 has demonstrated that binding the MLL peptide is achieved through a peptidyl arginine binding cleft found in the top groove of the beta propeller (Fig. 9) Mutations within this pocket disrupt the formation of the MLL1-WDR5-RbBP5-Ash2L complex and prevent the stimulation of MLL1 methyltransferase activity (Patel et al, 2008b). WDR5 has also been

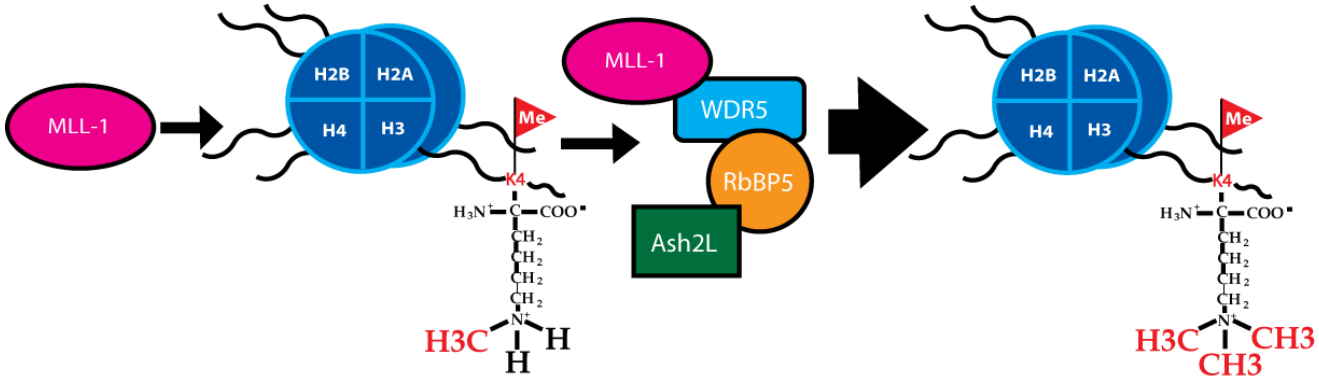


Fig. 7 – Stimulation of MLL1 methyltransferase activity by the core complex subunits.
On its own, MLL1 is only capable of catalyzing a monomethylation of H3K4. Presence of the core complex subunits, namely Ash2L, RbBP5 and WDR5, increase MLL's kinetic activity and change its product specificity to a trimethyltransferase.

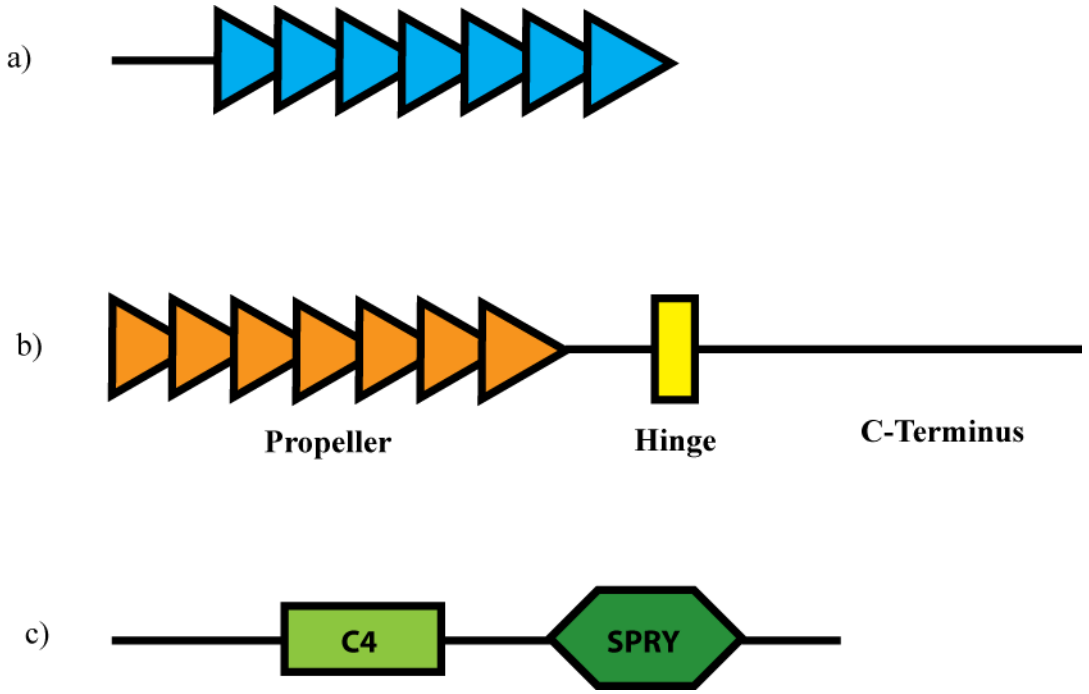


Fig. 8 – Structure of core complex subunits.

a) WDR5 is composed of an seven bladed WD repeat beta propeller
b) Ash2L is composed of a C-terminal SPRY domain and an N-terminal C4 Zinc finger.
c) RbBP5 is composed of a predicted N-terminal beta propeller, a hinge region, and C-terminus.

reported to interact with RbBP5, although this has yet to be characterized. Because of these observations it was hypothesized that WDR5 plays a key role in nucleating the core complex and that regulation of MLL1 kinetics and product specificity is carried out by the remaining subunits of the core complex, namely Ash2L and RbBP5.

1.4.3 – RbBP5 Structure and Function

RbBP5 is composed of three defined regions: an N-terminal domain composed of seven bladed WD repeat that likely folds in a structure resembling WDR5, a 40 amino acids region referred as the hinge domain and a C-terminal region rich in negatively charged amino acids (Fig. 8b). Within the core complex, RbBP5 interacts with both Ash2L and WDR5 and is hypothesized to act as key scaffold that mediates the formation of the core complex (Dou et al, 2006). Similar to Ash2L, RbBP5 knockdown leads to a decrease in the levels of H3K4 di/tri methylation. Recently, studies have suggested that, together, Ash2L and RbBP5 synergize to stimulate MLL1 methyltransferase activity (Odho et al, 2010).

1.4.4 – Ash2L Structure and Function

Ash2L was first identified as a homeotic regulator of drosophila development and body patterning (Ikegawa et al., 1999). More recently it was shown to be required for early embryogenesis in mammals (Stoller et al, 2010). It is composed of a C-terminal SPRY and a N-terminal C4 Zinc finger (Fig. 8c). Although the SPRY and C4 Zinc finger domains have been shown to be mediators of protein-protein interactions and a DNA binding domain respectively (Rohs et al, 2010), the exact function of each domain in the MLL core complex has not been defined. However the mechanism through which Ash2L promotes trimethylation is unknown.

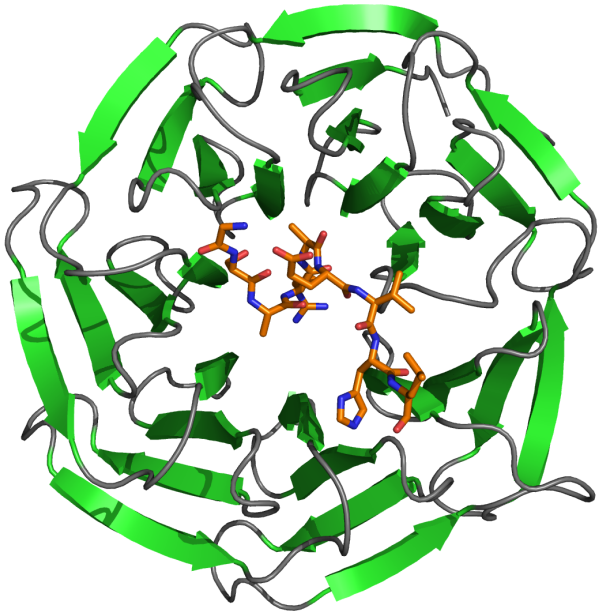


Fig. 9 – Crystal structure of WDR5 in complex with MLL1 WIN peptide.

WDR5 is a WD-repeat beta propeller (shown in green) that interacts with the MLL1 WIN peptide (orange) through a peptidyl arginine binding cleft found on the top face of the propeller.

1.5 - Aims and hypotheses

To date, several questions remain about the role of the MLL1 core complex subunits in modulating MLL1 methyltransferase activity. Firstly, the structural determinants for key interactions, such as that of RbBP5-WDR5 and Ash2L-RbBP5 remain unknown. Secondly, each member of the core complex is composed of multiple functional domains and it is unclear if a single domain or combinations of functional domains participate in complex formation, catalysis or product specificity. Finally, the structural and biochemical details of how the overall tripartite complex WDR5-RbBP5-Ash2L stimulates MLL1 methyltransferase activity remain unknown. We aim to answer these questions utilizing several biochemical, biophysical and structural methods. We hypothesize that each subunit will have a unique function for modulating kinetics and product specificity of MLL SET domain.

CHAPTER 2 – THE MANUSCRIPTS

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

Vanja Avdic[1]*, Pamela Zhang[1]*, Sylvain Lanouette[1], Adam Groulx[1], Veronique Tremblay[1], Joseph Brunzelle[2], and Jean-Francois Couture[1]

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

2) Feinberg School of Medicine, Department of Molecular Pharmacology and Biological Chemistry, Northwestern University

Running head: Dissection of the MLL core complex assembly

* These authors equally contributed to this article.

2.1.1 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 Ash2L, RbBP5 and WDR5. We herein report that RbBP5 mediates the formation of the core complex using a region of 50 amino acids. This minimal domain links Ash2L to WDR5 and stimulates, to the extent obtained with the full-length reconstituted protein complex, MLL1 methyltransferase activity. The crystal structure of WDR5 in ternary complex with RbBP5 and MLL1 reveals that both proteins are maintained in peptide binding clefts on opposite sides of WDR5's β-propeller domain. Binding of RbBP5 is mediated by a pair of evolutionary conserved hydrophobic residues nestled in a V-shaped pocket formed by the junction of two blades. The RbBP5/WDR5/MLL1 structure offers an unprecedented view on the organization of the core complex subunits with MLL1 SET domain and further reemphasize on the role of WDR5 as a primary scaffold of the MLL1 core complex.

2.1.2 INTRODUCTION

Histone lysine methylation is a prominent post-translational modifications linked to diverse biological pathways including gene transcription, maintenance of heterochromatin (Lachner et al, 2001; Schotta et al, 2002) and double stranded DNA break repair (Greeson et al, 2008). Notably, due to its association with actively transcribed genes, methylation of Lys-4 on histone H3 (H3K4me) 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 et al, 2004a)

The SET1 family of enzymes, which includes MLL1 to 5 and SET1A/B, carries out the body of H3K4 di/tri-methylation *in vivo* (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 Ash2L, RbBP5 or WDR5 leads to a global loss of H3K4 tri-methylation (Dou et al, 2006; Steward et al, 2006b; Wysocka et al, 2005). These results have been further confirmed by recent functional and biochemical studies showing that Ash2L, RbBP5 and WDR5 are all

essential in the assembly of the core complex and the allosteric regulation of MLL1 methyltransferase activity (Patel et al, 2008a; Patel et al, 2009; Patel et al, 2008b; Southall et al, 2009b). Nevertheless, the mechanisms underlying the formation of the Ash2L-WDR5-RbBP5 complex and its binding with 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.

In this study, we report the minimal structural determinants underlying the formation of the Ash2L- RbBP5-WDR5 complex. Notably, we demonstrate that WDR5-Ash2L interaction is supported by a segment of 50 amino acids of RbBP5. The interaction between RbBP5 and Ash2L is mediated through residues 330-370 of RbBP5 and the C-terminal domain of Ash2L while residues 371-380 of RbBP5 are responsible for its association with WDR5. Importantly, the minimal and the full-length core complexes restore MLL1 methyltransferase activity to the same extent. 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 side of WDR5's β-propeller domain. Notably, RbBP5 binds WDR5 using a pair of hydrophobic residues pinioned in two hydrophobic pockets formed by the junction of two blades. Overall, our results provide the fundamental basis underlying the formation of the core complex and suggest that Ash2L N-terminus and RbBP5 β-propeller domain might have alternative functions related to MLL1 biological roles.

2.1.3 RESULTS AND DISCUSSION

I - The C-terminal domain of Ash2L binds the hinge region of RbBP5.

The core complex controls the allosteric regulation of the SET1 family of methyltransferases. Initial work performed on the core complex revealed that Ash2L directly

interacts with RbBP5 (Patel et al, 2009). The latter is composed of an N-terminal seven-bladed β-propeller domain, a short hinge region previously shown to scaffold the core complex (Dou et al, 2006) and a C-terminal domain of unknown function (Fig. 10A). To define the role of 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 330-370 (Hinge: H) (Fig. 10A) were expressed as GST-fusion proteins and used as baits for *in vitro* pull-down experiments. Bound Ash2L was eluted from the column, resolved by SDS-PAGE and visualized by coomassie staining and western blot using an Ash2L specific antibody (Rampalli et al, 2007). As shown in Fig. 11A, 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. These results concur with previous studies (Dou et al, 2006) and further underscore the essential role of RbBP5 hinge region in binding Ash2L and forming the core complex.

Ash2L comprises a C4 zinc-finger motif and a SPla and RYanodine Receptor (SPRY) domain located on its N- and C-terminus, respectively (Fig. 10B). 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 Fig. 11B, we observed a consistent binding of $Ash2L_C$ with all GST-RbBP5 fragments while no binding was observed for $Ash2L_N$. These results reveal that a scaffolding domain, located between Ser-330 and the Ash2L C-terminus is involved in the binding of RbBP5 hinge region and suggest that Ash2L SPRY domain may contribute to the formation of the core complex. SPRY domains are scaffolding modules which employ a set of solvent-exposed loops to mediate protein-protein interactions (Filippakopoulos et al, 2010; Woo et al, 2006). Homology modeling of Ash2L SPRY domain shows that it harbors a

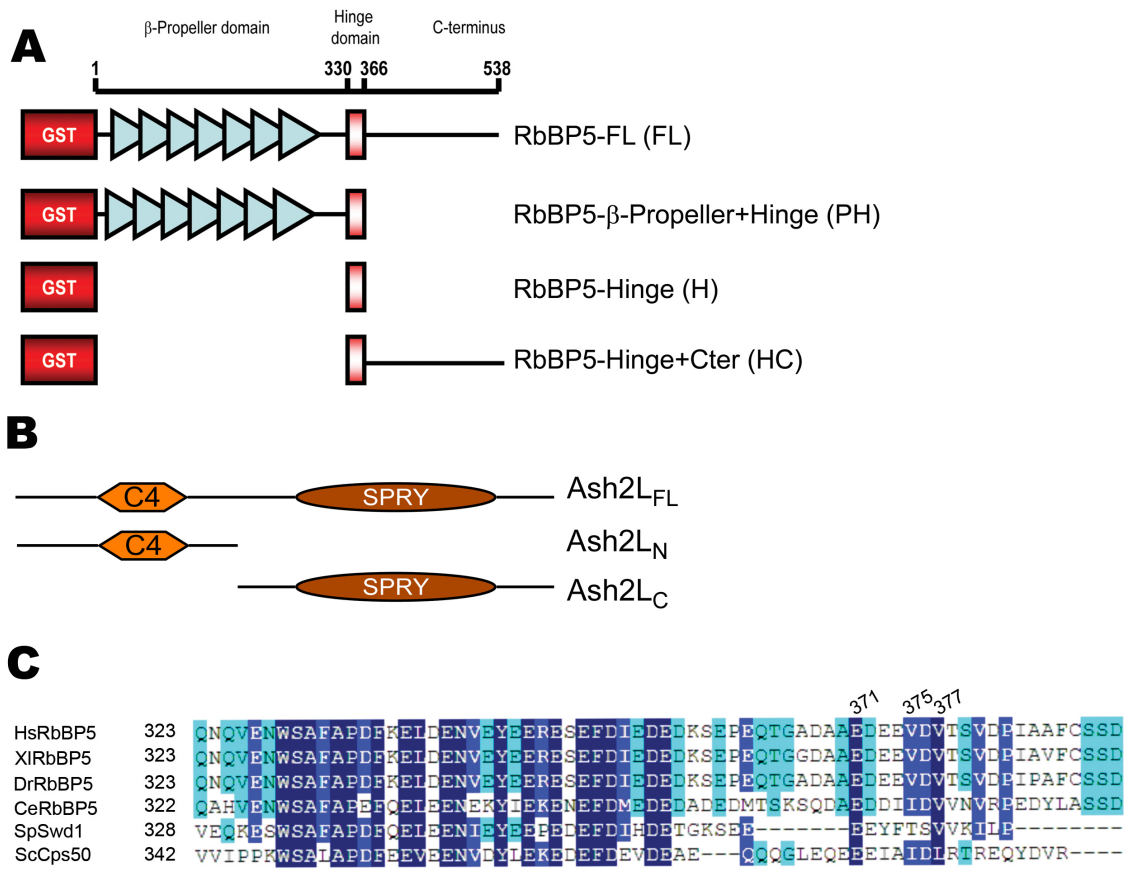


Fig. 10. Ash2L and RbBP5 harbor several evolutionary conserved domains.
Schematic representation of RbBP5 (A) and Ash2L (B) constructs used in the GST-pull down assays. C) A protein sequence alignment of H*omo sapiens* (Hs), *Danio rerio* (Dr), *Drosophila melanogaster* (Dm), *Caenorhabditis elegans* (Ce), *Schizosaccharomyces pombe* (Sp) and *Saccharomyces cerevisiae* 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.

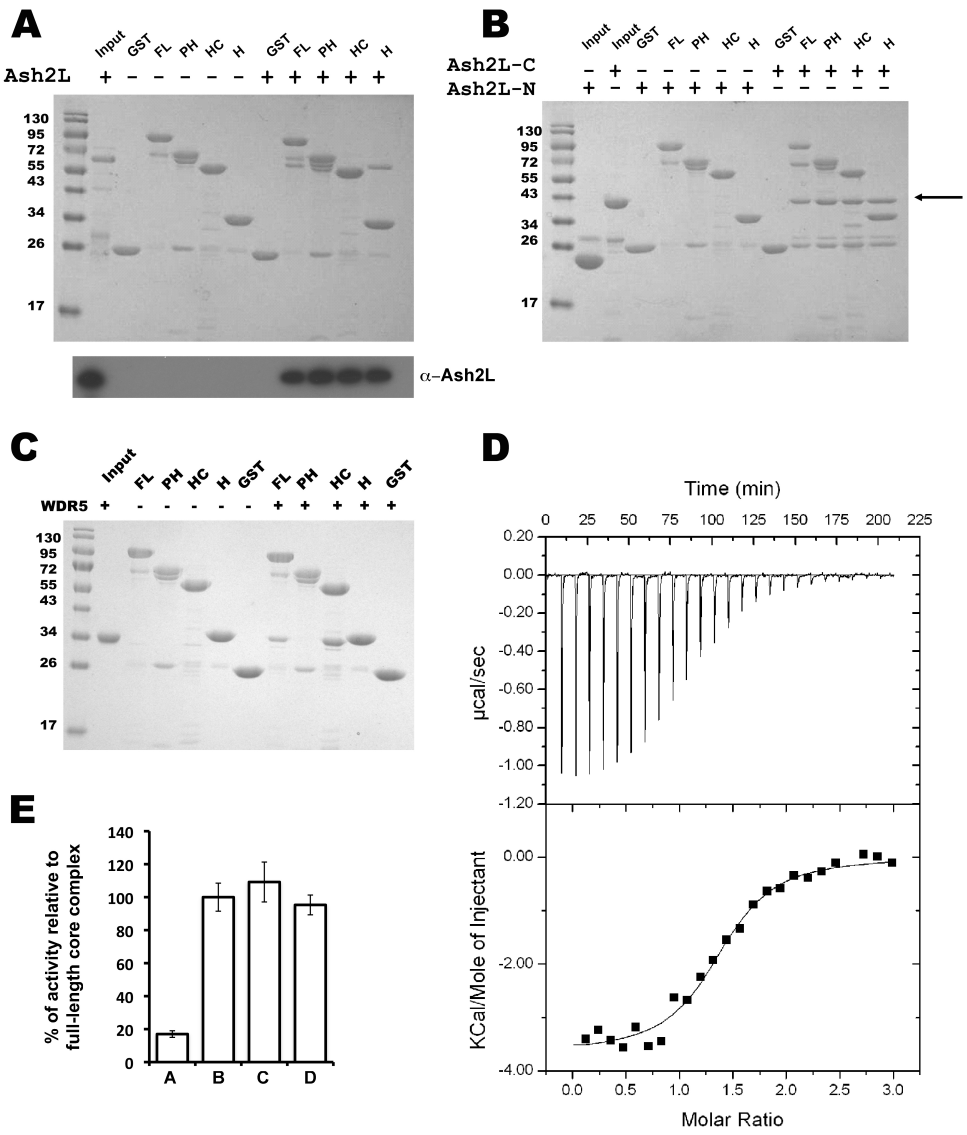


Fig.11. Dissection of the core complex.

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) Isothermal titration calorimetry data for binding of RbBP5 peptides to WDR5. The upper and lower panels show the heat of binding for peptide binding and integrated data respectively. E) The mini-core complex fully stimulates the methyltransferase activity of MLL1 SET domain. Radiometric methyltransferase assays performed with MLL1 SET domain alone (A), or the core complex reconstituted with $Ash2L_{FL}$-$RbBP5_{FL}$-$WDR5_{FL}$ (B), $Ash2L_C$-$RbBP5_{FL}$-$WDR5_{FL}$ (C) or $Ash2L_C$-$RbBP5_{330-380}$-$WDR5_{FL}$ (D) complexes.

similar set of loops with, however, a noticeable increase in surface basic residues. In addition, 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).More work is required to determine whether this insertion, or the punctuate 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; 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 $Ash2L_C$-RbBP5-WDR5-MLL1 to other regulatory subunits of MLL1.

II - 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 Fig. 10A with purified WDR5 and subsequently determined that WDR5 binds to full-length RbBP5. However, contrasting with Ash2L, an enrichment of WDR5 is only observed in the binding assays performed with RbBP5-HC fragment (Fig. 11C). These results indicate that WDR5 binds a region located between RbBP5 hinge region and 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 (Fig. 10C). Given the functional overlap between RbBP5-

WDR5 and Cps50-Cps35 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 evolutionarily conserved domain outside of the β-propeller domain and the hinge region of RbBP5 is located between residues 371-410 (Fig. 10C). To confirm our observations and to further define the elements of RbBP5 which 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 (Table 1). Using isothermal titration calorimetry (ITC) (Fig. 11D), we found that WDR5 binds, with similar binding affinities, 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 (Table 1). Given that initial binding studies established that full-length RbBP5 binds to WDR5 with a K_d of 5mM (Patel et al, 2009), our data suggest that $RbBP5_{371-380}$ comprises the minimal structural determinants underlying the interaction between WDR5 and RbBP5.

III - The "mini-core" complex is sufficient to allosterically regulate MLL1 methyltransferase activity.

After defining the minimal structural elements involved in the formation of the core complex, we sought to determine whether the associating elements of Ash2L and RbBP5 were sufficient to stimulate MLL1 activity. As previously reported, incubation of full-length Ash2L, RbBP5 and WDR5 resulted in significantely increased MLL1 methyltransferase activity. Similarly, incubation of $Ash2L_C$ with full length WDR5 and RbBP5 induced maximal

Table 1. Equilibrium dissociation constants between WDR5 and RbBP5 peptides	
Peptides	**K_d (mM)**
RbBP5 371-410	2.0 ± 0.3
RbBP5 371-400	1.3 ± 0.1
RbBP5 371-390	2.0 ± 0.2
RbBP5 371-380	2.2 ± 0.2
RbBP5 381-390	N.B[a]

a : no heat of binding is detected

stimulation of MLL1 activity (Fig. 11E). Moreover, incubation of the catalytic domain of MLL1 with $RbBP5_{330-380}$, $Ash2L_C$ and WDR5 recapitulated full activity of the core complex, strongly suggesting that the structural determinants underlying the formation of the core complex are sufficient for the stimulation of MLL1 methyltransferase activity when performed with a protein fragment corresponding to its catalytic domain. Protein lysine methyltransferases are often found in multiprotein 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, 2006b; Takahashi et al, 2009) while Ash2L is essential for stimulating MLL1 methyltransferase activity in mammals (Dou et al, 2006; Patel et al, 2008b; Southall et al, 2009b). To explain the role of these subunits, a structural reorganization of the iSET and cSET region has been proposed. Binding of the Ash2L-RbBP5-WDR5 complex is believed to reorganize MLL1 iSET and contribute to the formation of MLL1 peptide binding cleft (Southall et al, 2009a). Correlatively to these hypothesis and contrasting with well-characterized lysine methyltransferases SET8 (Couture et al, 2005) and DIM-5 (Zhang et al, 2003), mutation of acidic residues located in the iSET region of MLL1 does not impair its methyltransferase activity when bound to its core complex subunits (Southall et al, 2009a). We

postulate that the high content in glutamic and aspartic acid residues of RbBP5 hinge region and/or Ash2L 19 amino acid insertion might participate in the binding of MLL1 cognate substrate. Alternatively, binding of Ash2L and RbBP5 with positively charged residues found in the iSET and cSET domains may allow a structural reorganization of the MLL1 SET domain and result in an enlargement of the MLL1 active site, thereby rendering the MLL1 lysine binding channel conducive to di-methylation.

IV - Structure of the WDR5/MLL1/RbBP5 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 $RbBP5_{371-380}$ and MLL1 WIN motif (Fig. 12A, Table 2). Inspection of the ternary complex reveals that MLL1 WIN motif is maintained in a 3_{10} helix conformation with its C-terminus exiting the WDR5 peptidyl arginine-binding cleft toward blade 6. As previously described, MLL1 WIN motif is maintained by several hydrogen bonds and hydrophobic contacts (Patel et al, 2008a; Song & Kingston, 2008a). Upon closer inspection of WDR5 surface, we found a continuous electron density corresponding to residues 374-380 of RbBP5. The first three residues of RbBP5 peptide were not modeled as no discernable electronic density was observed. Analysis of $RbBP5_{371-380}$ binding mode reveals that it adopts an elongated conformation with a small hairpin centered on V377. The RbBP5 peptide is found in proximity of a V-shaped cleft formed by the junction of blades 6 and 7 of WDR5 (**Fig. 12A**) and relative to the WDR5 peptidyl arginine binding cleft, is maintained on the opposite side of the β-propeller. Overall, the structure of WDR5/WIN/ $RbBP5_{371-380}$ is similar to apo-WDR5 with an overall r.m.s. of 0.58Å for all protein atoms. Close inspection of the V-shaped cleft reveals that $RbBP5_{371-380}$ is maintained by several hydrogen

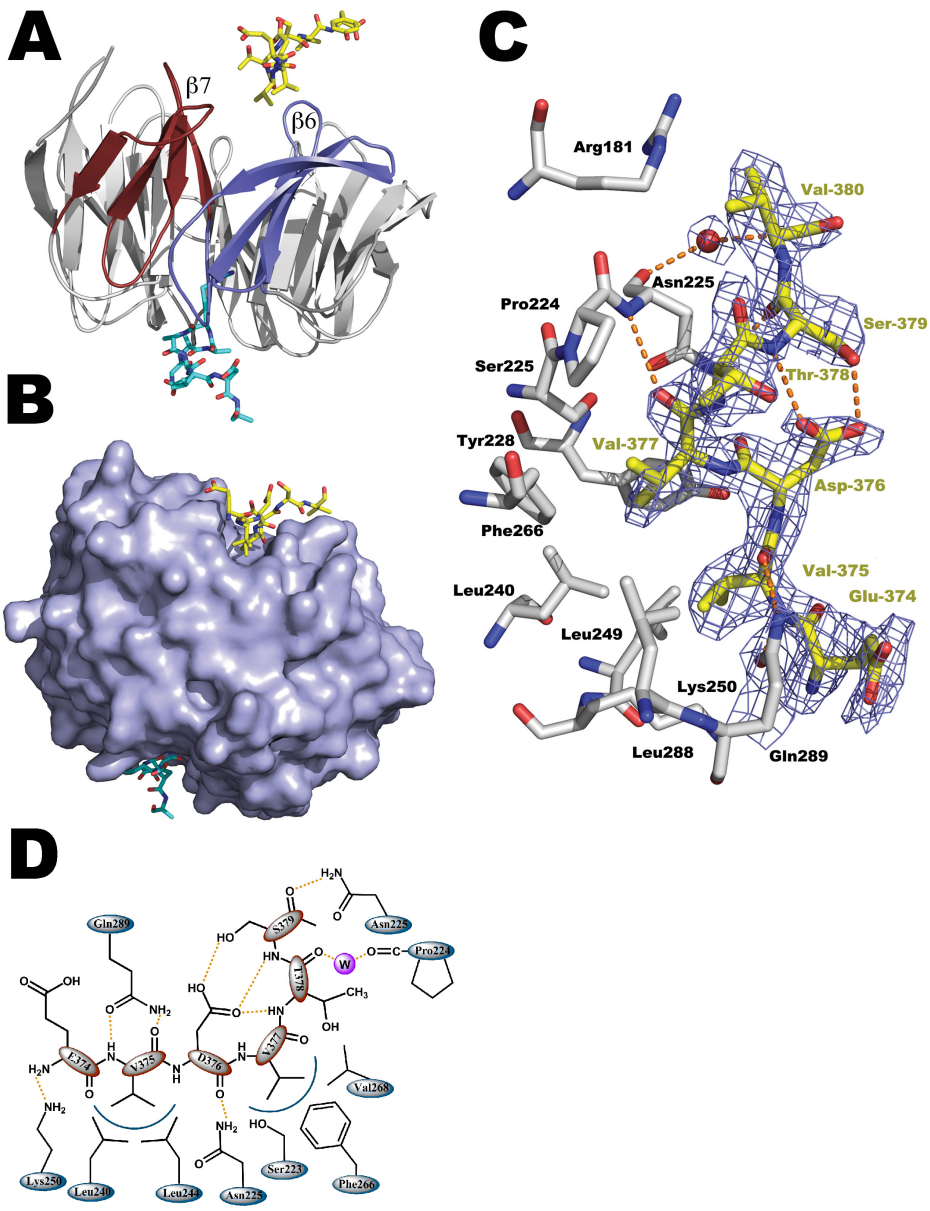


Fig. 12. WDR5 binds RbBP5 and MLL1 using two independent binding clefts.

A) Overall structure of WDR5 in complex with MLL1 WIN motif and RbBP5$_{371\text{-}380}$. Blades 6 and 7 are highlighted in blue and red respectively. MLL1 WIN motif and RbBP5$_{371\text{-}380}$ 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.

D) Schematic representation of WDR5/RbBP5$_{371\text{-}380}$ interaction. WDR5 and RbBP5$_{371\text{-}380}$ residues are represented with green and red ovals, respectively. Hydrogen bonds and hydrophobic contacts are highlighted in black and blue 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.

Table 2: Data collection and refinement statistics for WDR5 structure in complex with RbBP5.	
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	53.9, 81.4, 81.8
Resolution	32.6 – 2.35 (2.43 – 2.35)!
R_{meas}	8.9 (44.2)
$I / \sigma I$	8.3 (2.1)
Completeness (%)	99.2 (98.9)
Redundancy	3.3 (3.3)
Refinement	
Resolution (Å)	32.6 – 2.35
No. reflections	15414
R_{work} / R_{free}	23.0 / 27.3
No. Atoms	
Protein	2252
Peptides	123
Water	58
B-factors (Å^2)	
Protein	35.1
Ligands	40.6
Water	32.9
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.085
Molprobity scores	1.42
Ramachandran favored (%)	96.0
Ramachandran allowed (%)	4.0

! Highest resolution shell is shown in parentheses.

bonds and van der Waals contacts. Notably, V375 side chain participates in hydrophobic interactions within a small pocket composed of Leu249, Tyr228, Leu240 and Lys250. In addition, the carbonyl oxygen and backbone amide of V375 engages in two hydrogen bonds with Gln289 amide side chain (Fig. 12C). Following V375 in RbBP5, D376 interacts with WDR5 through a hydrogen bond between its backbone carbonyl and Asn225 amide side chain. Similar to MLL1 WIN binding mode (Patel et al, 2008a; Song & Kingston, 2008a), RbBP5 engages in a trio of intra-molecular hydrogen bonds between D376 carboxylate group and T378 backbone

amide, S379 hydroxyl group and backbone amide (Fig. 12C). This network of hydrogen bonds likely stabilizes RbBP5's hairpin conformation and orients V377 side chain in a hydrophobic pocket composed of Phe266, Tyr228, Ser223 and Leu288. The residues T378 and S379 participate in water-mediated and direct hydrogen bonds with Pro224 carbonyl group and Asn225 amide side chain respectively. Overall, the crystal structure of WDR5 in ternary complex with MLL1 WIN motif and $RbBP5_{371-380}$ and sequence homology of RbBP5 homologs (Fig. 10C) 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.

Similar to HsAsh2L and Cps60 SPRY domains, the central region of $RbBP5_{371-380}$, including the triad V375-D376-V377, is conserved in RbBP5 yeast homolog, Cps50. Inspection of Cps50 reveals a similar triad composed of I291-D292-L293 (Fig. 10A), suggesting that the interaction of this pair of hydrophobic residues onto Cps35's β-propeller domain is important for the binding of Cps50 and complement 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, pull-down and binding assays indicate that the $RbBP5_{330-380}$ region is of importance in the formation of the core complex, and suggest that this mechanism is shared between the human SET1 family of methyltransferases and its yeast homolog COMPASS.

V - Insights into the organization of a catalytically competent Ash2L/RbBP5/WDR5/MLL1 complex.

Recent studies have posited that the binding of the Ash2L/RbBP5/WDR5 complex would reorganize MLL1 iSET region and concomitantly stimulate H3K4 mono/di-methylation

(Southall et al, 2009a). The results identify a minimal core complex, composed of $Ash2L_C$-$RbBP5_{330-380}$-WDR5, necessary to the allosteric regulation of MLL1 methyltransferase activity. In combination with the WDR5/WIN/$RbBP5_{371-380}$ ternary complex, these results yield insights into the spatial organization of the core complex subunits in a catalytically competent MLL1 complex. The MLL-WIN peptide is bound to the peptidyl arginine-binding cleft with its C-terminus exiting the pocket toward blade 6 of WDR5. In contrast, $RbBP5_{371-380}$ is maintained in proximity of WDR5's blade 6 and 7 in an orientation anti-parallel to MLL1 WIN motif. We postulate that this anti-parallelism of MLL1 and RbBP5 binding modes can constitute a crucial element of allosteric regulation of MLL1 methyltransferase activity. It can be expected that, in a fully reconstituted complex, WIN and $RbBP5_{371-380}$ interaction within their respective binding cleft would orient MLL1 and RbBP5 so as to allow their bridging by $Ash2L_C$. The concomitant recruitment of $Ash2L_C$ would therefore provide a structural platform to interact with MLL1 iSET region or histone H3 substrate and thereby stimulate MLL1 methyltransferase activity.

Overall, our results strongly suggest 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. Moreover, this association is sufficient to stimulate enzymatic activity of a protein fragment corresponding to the catalytic domain of MLL1. They also suggest that $Ash2L_N$ domain as well as RbBP5 β-propeller may interact with other subunits or co-factors of the MLL complexes and thereby serving alternative regulatory purposes related to MLL complexes. Finally, the organization of RbBP5 and MLL peptides on WDR5 strongly suggest that Ash2L SPRY domain is of critical importance in the allosteric regulation of MLL1 and point to a shared mechanism of allosteric regulation between COMPASS and MLL1.

2.1.4 - MATERIAL AND METHODS

I - 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 et al, 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 et al, 1999). Proteins were overexpressed with 0.1mM IPTG in Rosetta cells (Novagen) for 16 hours at 18°C. Cells were harvested in 50mM sodium phosphate, 500mM NaCl and 5mM β-mercaptoethanol, lysed by sonication, clarified by centrifugation and purified by Talon Co^{2+} affinity chromatography. TEV-cleaved Ash2L-FL and $Ash2L_N$ were further purified by metal affinity and by size exclusion chromatography (Superdex 200). ULP1-cleaved $Ash2L_C$ was purified by Q-sepharose ion exchange and size exclusion chromatography (Superdex 200). GST-RbBP5 constructs were expressed in BL-21 using the same experimental conditions as for the N-terminally hexahistidine tagged $Ash2L_N$. Following GST-RbBP5 overexpression, cells were centrifuged and harvested in phosphate buffer saline (PBS). The catalytic domain of MLL-1 was produced as previously described (Southall et al, 2009a).

II - GST-pull down experiment - GST and GST fusion RbBP5 proteins were applied onto glutathione-sepharose beads for 1 hour and washed extensively with PBS. Binding assays were performed in PBS supplemented with 0.1% Triton-X 100 (PBST) for two hours with 10mg of recombinant Ash2L or WDR5 proteins. Binding reactions were extensively washed with PBST and bound proteins were eluted with 50mM Tris-HCl pH 8.0 and 20mM reduced glutathione for

1 hour at 4ºC. 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).

III - In vitro methyltransferase assay - Methyltransferase assays were conducted using 5mM of recombinantly purified MLL1 SET domain (3785-3969) with equimolar amounts of Ash2L, RbBP5 and WDR5. Reactions were performed in 50mM Tris 8.5, 200mM NaCl, 3mM DTT, 5mM $MgCl_2$, 250mM of peptide (residue 1-20 of histone H3) and 5% glycerol. Each assay was initiated by the addition of 1mCi of radiolabeled AdoMet and incubated for 2 hours at 22ºC. Reactions were stopped by spotting the reactions onto Whatman P-81 filter papers and free AdoMet was removed by washing the filter papers in 250mL of 50mM $NaHCO_3$ at pH 9.0. Activity was quantified by liquid scintillation counts.

IV - Isothermal titration calorimetry - Isothermal titration calorimetry (ITC) experiments were performed using a VP-ITC calorimeter (MicroCal). RbBP5 peptides (0.65mM) were injected into a solution of WDR5 (0.06mM) as previously described (Couture et al, 2006). Experiments were performed at 19°C in 20 mM sodium phosphate (pH 7.0), 100 mM sodium chloride, and 5 mM β-mercaptoethanol. The titration data were analyzed using Origin 7.0 (OriginLab Corp.) and found that the RbBP5 peptides bind WDR5 with binding stoichiometries (N-values) between 0.9-1.1.

V - Crystallization, data collection and structure determination - Purified WDR5 was mixed in equimolar ratio with MLL1 WIN motif and $RbBP5_{371-380}$ and incubated on ice for 30 minutes.

Diffraction quality crystals were grown in 0.2M 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 an Raxis IV image plate detector. Diffraction data were indexed and scaled using d*TREK® (Pflugrath, 1999). A molecular replacement solution was found using MOLREP (Vagin & Teplyakov, 2000) and apo-WDR5 as a starting model (Couture et al, 2006a). The structure was further refined using iterative cycles of energy minimization and modeling using Refmac (Vagin et al, 2004) and COOT (Emsley & Cowtan, 2004), respectively. Quality of the structure was assessed using Molprobity (Chen et al, 2009) (Table 2).

2.1.5 - 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.

2.2 - An atypical helix-wing-helix domain mediates Ash2L binding to the *beta-globin* gene.

Vanja Avdic[1*], Sabina Sarvan[1*], Véronique Tremblay[1*], Chandra-Prakash Chaturvedi[2,3], Sylvain Lanouette[1], Alexandre Blais[1], Joseph S. Brunzelle[4], Marjorie Brand[2,3] Jean-François Couture[1]

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

2) The Sprott Center for Stem cell Research, Regenerative medicine Program, Ottawa Hospital Research Institute

3) Department of Cellular and Molecular Medicine, University of Ottawa

4) Feinberg School of Medicine, Department of Molecular Pharmacology and Biological Chemistry, Northwestern University

* These authors equally contributed to this article

Short title: Ash2L physically interacts with *b-globin* locus control region.

Corresponding author: Jean-Francois Couture

2.2.1 - ABSTRACT

Ash2L functions as a regulatory subunit of the Myeloid Lymphoma Leukemia protein complex, and is key in regulating global histone H3 Lys-4 tri-methylation and β-globin expression in erythroid cells. We have found that the N-terminal domain of Ash2L binds the hypersensitive site 2 (HS2) of the β-*globin* locus control region (LCR) using an atypical helix-wing-helix domain. Mutational analysis identified two residues, Lys-225 and Lys-229 as essential to DNA binding by Ash2L, recruitment to HS2 and transcriptional activation of a reporter gene. Our data provide the first evidence suggesting that Ash2L binds DNA and unravel a novel mechanism involved in the binding of this trithorax protein to its target genes.

2.2.2 - INTRODUCTION

The trithorax protein Ash2L is a subunit of the SET1 family of histone H3 Lys-4 methyltransferases. When found in a tri-partite complex with WDR5 and RbBP5, it stimulates the methyltransferase activity of MLL1-4 and SET1A/B (Dou et al, 2006). This activity is essential for regulating gene expression as decreased levels of Ash2L results in a loss of histone H3 Lys-4 trimethylation and a concomitant loss of expression of specific MLL target genes (Steward et al, 2006a).

Several lines of evidence have recently demonstrated that Ash2L links basal transcription factors to histone H3 Lys-4 methylation. In C2C12 muscle cells, the p38a dependent phosphorylation of Mef2d recruits the Ash2L complex to the *MyoG* and *Ckm* gene promoters. The binding of Ash2L to these loci enhances histone H3 Lys-4 methylation and maintains *MyoG* and *Ckm* expression (Rampalli et al, 2007). In Neuro2a neuroblastoma cells, Ap2δ directly binds Ash2L to tether ALR and MLL2 to the HoxC8 promoters, thereby maintaining its expression

(Tan et al, 2008). In line with these findings, Ash2L is required for the mitotic inheritance of active transcriptional states in *Dictyostelium* cells (Muramoto et al, 2010). Additionally, in differentiating erythroid cells, the NF-E2 transcription factor complex tethers MLL2, as well as Ash2L, to hypersensitive site 2 (HS2) of the *β-globin gene* promoter (Demers et al, 2007). This recruitment of Ash2L is important as decreased levels of Ash2L diminish β-globin transcription. Overall, these studies suggest that Ash2L plays a role in linking transcription factors to histone H3 Lys-4 methylation; however, the underlying mechanisms have remained unanswered.

Here we present structural and functional evidence that Ash2L N-terminal domain binds DNA. In combining SELEX and electrophoretic mobility shift assays (EMSA), we have demonstrated that Ash2L preferably binds torsionally stressed DNA. To investigate the structural determinants underlying Ash2L DNA binding activity, we have solved the crystal structure of Ash2L N-terminal domain and found that it adopts an atypical helix-wing-helix fold. To complement our biochemical analysis, we have performed chromatin immunoprecipitation experiments and confirmed that the atypical helix-wing-helix domain is essential for the binding of Ash2L to β-globin LCR.

2.2.3 - RESULTS

I - Ash2L binds DNA.

Initial characterization of the binding of Ash2L to the *β-globin* gene (Demers et al, 2007) prompted us to investigate whether Ash2L could be recruited to this locus by directly binding to DNA. To verify this hypothesis, we performed EMSA using a probe corresponding to the hypersensitive site 2 (HS2) of *β-globin* LCR and purified recombinant full length Ash2L (**Fig. 13a**). The results show that Ash2L binds to HS2 as a clear retardation is observed upon

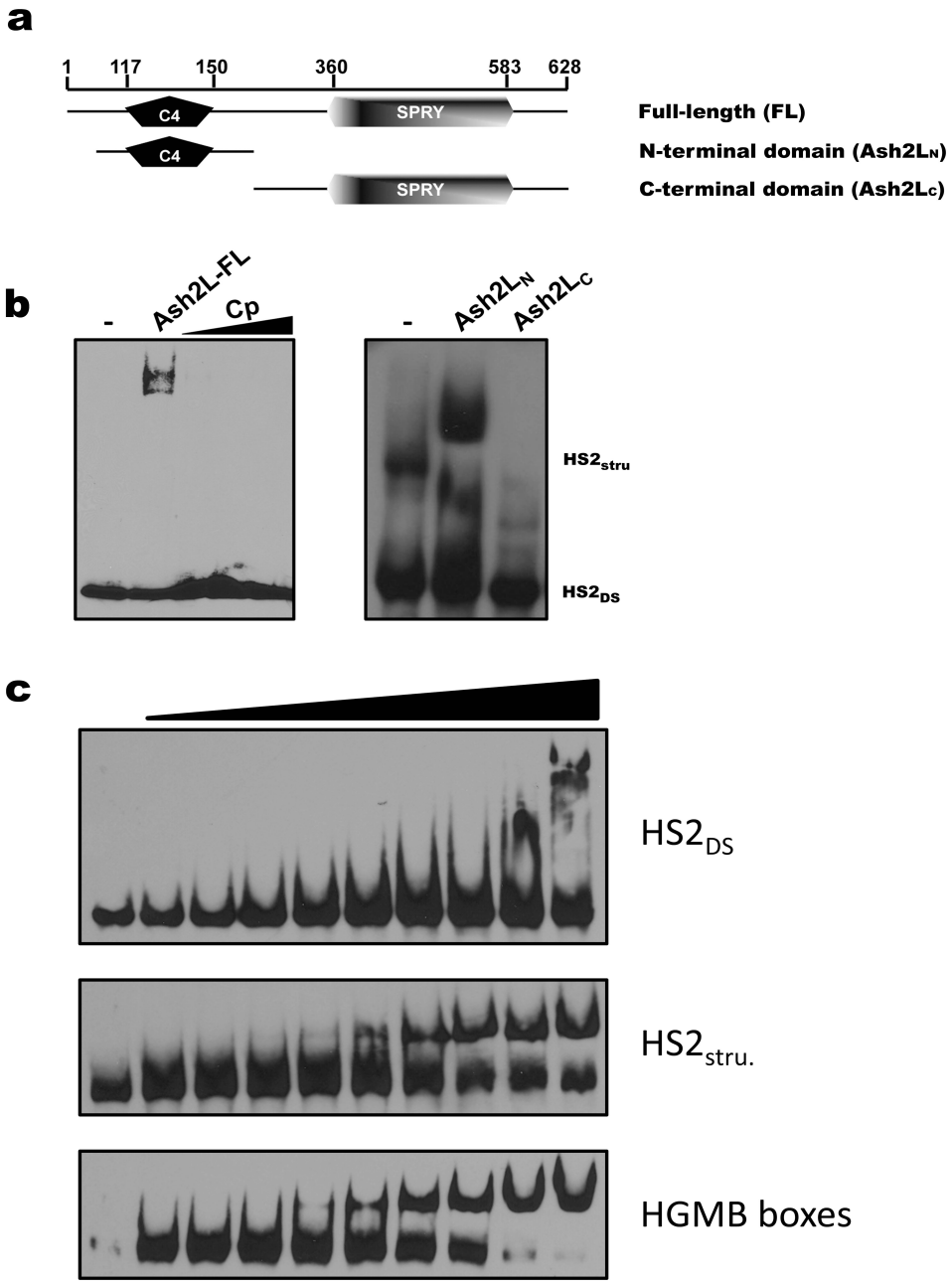


Figure 13. Ash2L preferentially binds torsionally constrained DNA.

(a) Organization of Ash2L domains.

(b) Left panel: EMSA of biotinylated HS2 using 1mM of full-length Ash2L. Specificity of the binding is confirmed using 10X, 25X and 50X molar excess of non-biotinylated probe (Cp). Right panel: Binding reactions performed with 2mM of $Ash2L_N$ or $Ash2L_C$.

(c) EMSA of double-stranded (top panel), structured (middle panel) and cruciform DNA (lower panel) using increasing amount of $Ash2L_N$ (0.5-15mM). DNA was detected according to DNA detection kit (Pierce).

incubation of the probe with the protein (**Fig. 13b**). The specificity of the interaction between Ash2L and HS2 was confirmed as an excess of a non-biotinylated probe competed the shift. Ash2L harbors a zinc finger motif and a SPla and RYanodine receptor (SPRY) domain that are located on its N- and C-terminus, respectively (**Fig. 13a**). To identify which domain(s) of Ash2L is interacting with HS2, each domain was homogeneously purified and incubated with HS2. As shown in **figure 13b**, the N-terminal domain of Ash2L can directly bind to DNA while no binding was detected with $Ash2L_C$. Upon inspection of our EMSA, we noticed the presence of a slower migrating band (referred therein as $HS2_{Stru}$) suggesting that HS2 DNA may adopt different conformations in solution. To determine which band Ash2L specifically binds; each DNA species was purified and tested in EMSA. **Figure 13c** shows that Ash2L binds with more affinity $HS2_{Stru}$, while weak binding could only be detected for HS2 DNA duplex. Based on initial results showing that HS2 confers negative supercoil to a pBluescript vector (data not shown), we posited that HS2 harbors a motif that induce torsional stresses to a vector backbone and that such DNA element is key for Ash2L binding. To confirm this hypothesis, we performed EMSA with recombinant Ash2L and HMG box (HMGB) DNA probes (Stefanovsky and Moss 2010), a canonical cruciform forming DNA. Consistent with the binding reactions performed with $HS2_{Stru}$, we found that Ash2L binds cruciform DNA with similar binding affinities (**Fig. 13c**).

To further confirm the ability Ash2L to bind non-duplex forms of DNA, we performed systematic evolution of ligands by exponential enrichment (SELEX) using GST-$Ash2L_N$ and a pool of random DNA aptamers. Following seven rounds of binding and PCR amplification, a total of 190 amplicons were selected and sequenced. Using a *de novo* motif finding algorithm,

we found a motif specifically enriched in the SELEX clones as compared to random sequences **(Fig. 14a)**. In an attempt to align all sequences containing this motif, we made the serendipitous finding that the majority of clones containing a match to the CAGCCT motif consisting of two palindromic hemi-sites with the sequence motif intervening between them **(Fig. 14b)**. Upon closer inspection of the other SELEX clones, several additional palindromic sequences were discovered. Interestingly, out of 30 palindromic sequences, 24 of them (80%) have a G/C nucleotide at the position closest to the central intervening sequence. Human genes with a match to this motif within 1000 bp of their transcription start sites were identified and clustered according to their biological function using Gene Ontology annotation. Consistent with previous genome-wide studies performed on drosophila Ash2 (Beltran et al, 2003), we found a significant enrichment for genes involved in system development, cell-cell signaling and DNA dependent transcription **(Fig. 14c)**. Using chromatin immunoprecipitation (ChIP) experiments, we confirmed the enrichment of Ash2L in loci harboring the SELEX motif neighbored by a G/C rich palindromic sequence including *TPP1, CAPN3, ALX1* and *NheJ1* genes **(Fig. 14d)**. Interestingly, upon closer inspection of the HS2 element, we found the AGGCT motif, a sequence corresponding to the reverse complement of the SELEX motif. To evaluate whether this region is important in the formation of $HS2_{Stru}$, an AGGCT – CTGGG substituted probe were annealed and separated on PAGE. As shown in **figure 14e**, we confirmed that the AGGCT motif is important in the formation of $HS2_{Stru}$ as a significant loss of the slower migrating DNA species was observed. Correlatively, the newly formed duplex could not compete the Ash2L (**Fig. 14f**) shifted wild-type probe, further underscoring the specificity of $Ash2L_N$ for torsionally constrained DNA.

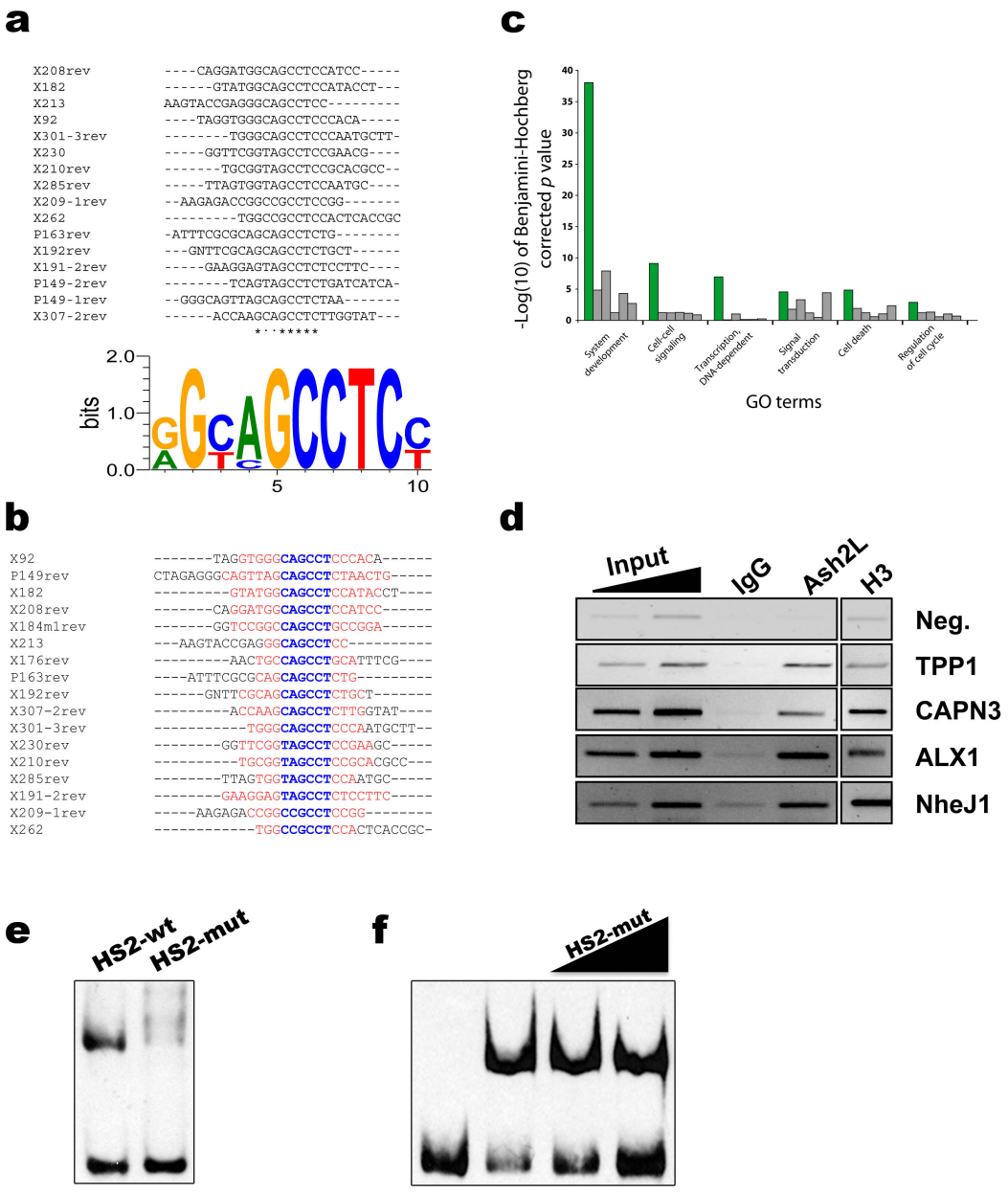


Figure 14. Identification of putative Ash2L target genes.

- **(a)** Alignment of the SELEX clones that yielded the motif.
- **(b)** Identification of the G/C rich inverted repeat sequences neighboring the SELEX motif.
- **(c)** Gene ontology term enrichment analysis was performed on human genes containing a good match to the motif in their proximal promoters (green histogram bar). The six non-redundant terms shown in this histogram were the most strongly enriched. As a control, 5 control sets of genes (without consideration of their proximal promoter sequence, grey bars) were created randomly and analyzed in the same way.
- **(d)** Validation of Ash2L target genes using chromatin immunoprecipitation in Hela cells.
- **(e)** The SELEX motif is key for the formation of $HS2_{stru}$. 2nM of each annealed DNA were loaded on a 7.5 % acrylamide gel and detected as in figure 1.
- **(f)** HS2-mut is unable to compete with the $HS2_{stru}$-wt. $Ash2L_N$ was incubated with 2nM of $HS2_{stru}$ alone or with 10X and 25X fold excess of HS2-mut.

II - Ash2L harbors an evolutionary conserved helix-wing-helix domain.

To further characterize Ash2L DNA binding domain, we used single wavelength anomalous dispersion (SAD) to solve the structure of $Ash2L_N$. As defined by the electron density, $Ash2L_N$ comprises residues 106-269 with one molecule in the asymmetric unit (**Table 3**). The crystallized protein consists of two anti-parallel β-sheets intersected by one α-helix (α1) (**Fig. 15a**). The β-sheets are orthogonal to each other and strands β3 and β4 are connected by a 12-amino acid loop. One zinc atom is maintained by a quartet of cysteine residues localized in β1/β2 and β3/β4 β-hairpins. The other half of $Ash2L_N$ is composed of four consecutive α-helices (α2-α5) and a C-terminal anti-parallel β-sheet in which strands β5 and β6 are linked by a 16-residue inter-connecting loop (**Fig. 15a**). The structure ends by a 4 amino acid α-helix which folds back onto β1. To locate the Ash2L DNA binding domain, we queried the Protein Data Bank using the structural database **DALI** (Holm et al, 2008). The C-terminal half of $Ash2L_N$, which comprises α2-α5 and β5-β6 (**Fig. 15a**), reveals similarities with the helix-wing-helix DNA binding domain (DBD) of the forkhead transcription factors FOXO1 and FOXO4 (z-score of 7.8 and 7.5). Structural alignment of $Ash2L_N$ C-terminal half with FOXO1 and FOXO4 DBD reveals root mean square deviations of 2.9 Å and 2.3 Å for Cα-atoms over 75 residues respectively (data not shown). Using ClustalW (Thompson et al, 2002) a multiple sequence alignment of $Ash2L_N$ was performed on a wide range of model eukaryotic species including representatives from yeast, worm, fly, fish and mammals. Mapping of the evolutionary conserved residues identified several highly conserved regions including those involved in the formation of α-helices and β-strands as well as in β4-α2 and β6-α6 inter-connecting loops (**Fig. 15b**). Interestingly, the helix-wing-helix domain of Ash2L shares high sequence similarity with

Table 3. Data collection, phasing and refinement statistics of Ash2L$_N$	
Data collection	
Space group	$P3_121$
Cell dimensions	
A, b, c (Å)	49.8, 49.8, 167.5
a, b, g (º)	90, 90, 120
Resolution (Å)	30.0 – 2.45 (2.49-2.45)!
I/σI	55.3(5.0)
Completeness (%)	99.6 (100)
Redundancy	18.9 (19.7)
Phasing	
Phasing power	2.7
Sites found/total	8/8
FOM*	83.5
Refinement	
Resolution (Å)	24.7 – 2.45
No. reflections	8970
R_{work}/R_{free}	24.5 / 29.1
No Atoms	
Protein	1299
Ligands	1
water	13
B-factors (Å^2)	
Protein	30.9
Ligands	4.4
water	52.3
R.m.s. deviations	
Bond lengths (Å)	0.012
Bond angles (°)	1.532

! Highest resolution shell is shown in parentheses.

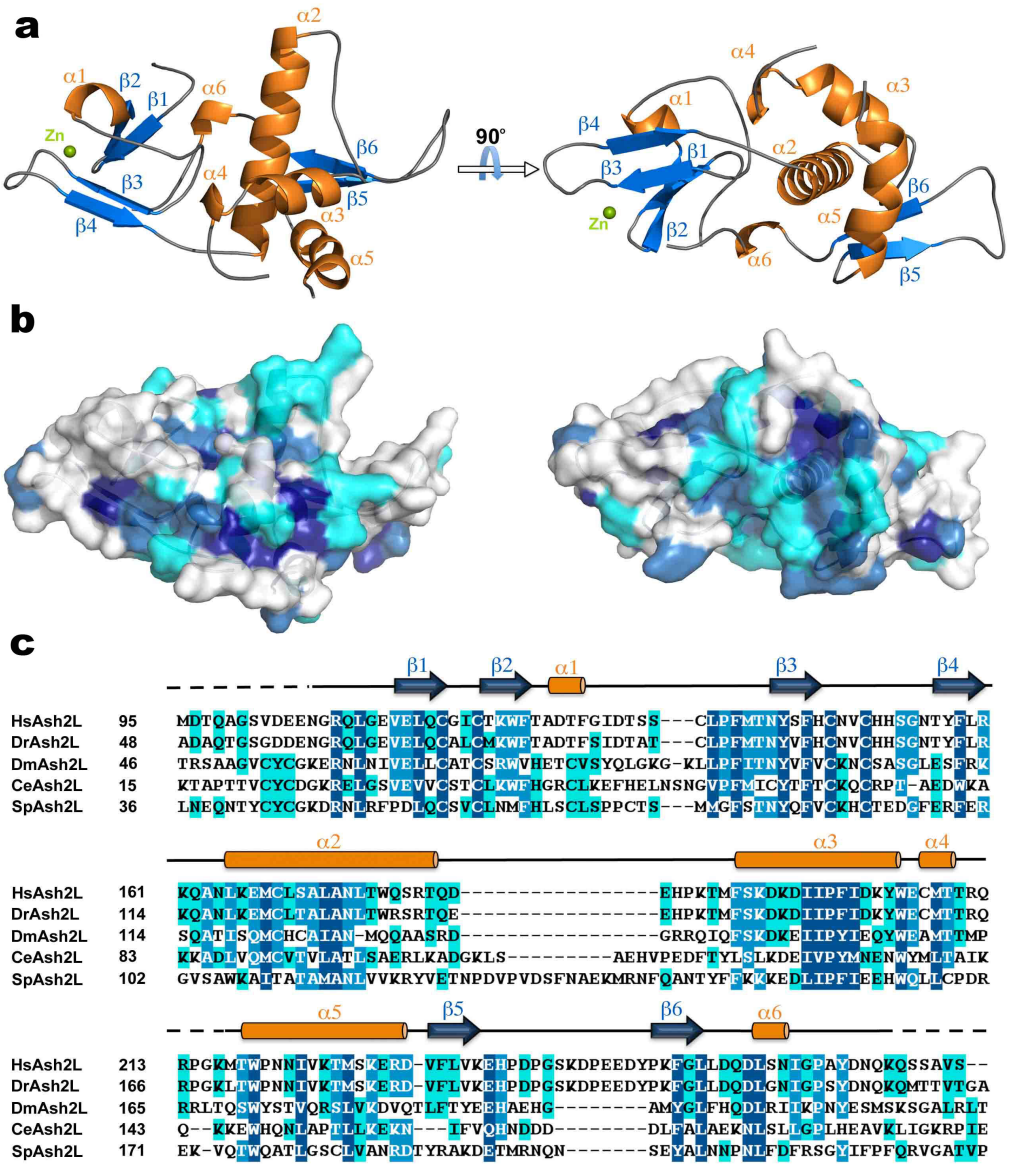


Figure 15. Ash2L harbors a C4 zinc finger and an evolutionary conserved extended helix-wing-helix domain.

(a) Crystal structure of Ash2L$_N$ secondary structure. β-strands and α-helices are rendered in blue and orange and the zinc atom is depicted in green.

(b) Evolutionary conserved surface residues colored accordingly to the sequence alignment.

(c) A protein sequence alignment of the *Homo sapiens* (Hs) of Ash2L$_N$ (numbered accordingly to the isoform 1 of the Ash2L (uniprot.org) with the corresponding N-terminal domains of *Danio rerio* (Dr), *Drosophila melanogaster* (Dm), *Caenorhabditis elegans* (Ce) and *Schizosaccharomyces pombe* (Sp) Ash2L proteins. Ash2L$_N$ β-strands and α-helices are depicted as cylinders and arrows, respectively. Positions with 100-80%, 80-60% and less than 60% of amino acid conservation are represented respectively in dark, medium and pale blue.

other Ash2L homologs as several residues lining Ash2L$_N$ α5-helix, which include Asn-221, Lys-225 and Lys-229, are conserved (**Fig. 15C**). respectively (data not shown). Using ClustalW (Thompson et al, 2002) a multiple sequence alignment of Ash2L$_N$ was performed on a wide range of model eukaryotic species including representatives from yeast, worm, fly, fish and mammals. Mapping of the evolutionary conserved residues identified several highly conserved regions including those involved in the formation of α-helices and β-strands as well as in β4-α2 and β6-α6 inter-connecting loops (**Fig. 15b**). Interestingly, the helix-wing-helix domain of Ash2L shares high sequence similarity with other Ash2L homologs as several residues lining Ash2L$_N$ α5-helix, which include Asn-221, Lys-225 and Lys-229, are conserved (**Fig. 15C**).

In FOXO proteins, hydrophobic and positively charged residues lining this α-helix are key in binding DNA (Brent et al, 2008; Obsil & Obsilova, 2008) (**Fig. 16a**). To seek whether this region of Ash2L played a similar role, we combined site-directed mutagenesis to EMSA and determined that substitution of Lys-225 or Lys-229 by glutamic acid residues completely impaired DNA binding activity as no binding could be detected (**Fig. 16b**). Loss of binding was not attributable to loss of folding as each mutant elicits similar CD spectra (S**upplementary Figure 1**). Overall, our analysis suggests that Ash2L helix-wing-helix domain encompasses an evolutionary conserved surface that comprises residues that confer DNA binding activity.

III - The helix-wing-helix domain is required for Ash2L binding to β-globin LCR and transcriptional activity.

Based on previous reports showing that Ash2L participates in the transcriptional regulation of the *β-globin* gene (Demers et al, 2007;McKinnell et al, 2008;Rampalli et al, 2007), we posited that the Ash2L DNA binding domain would play an important role in the binding of

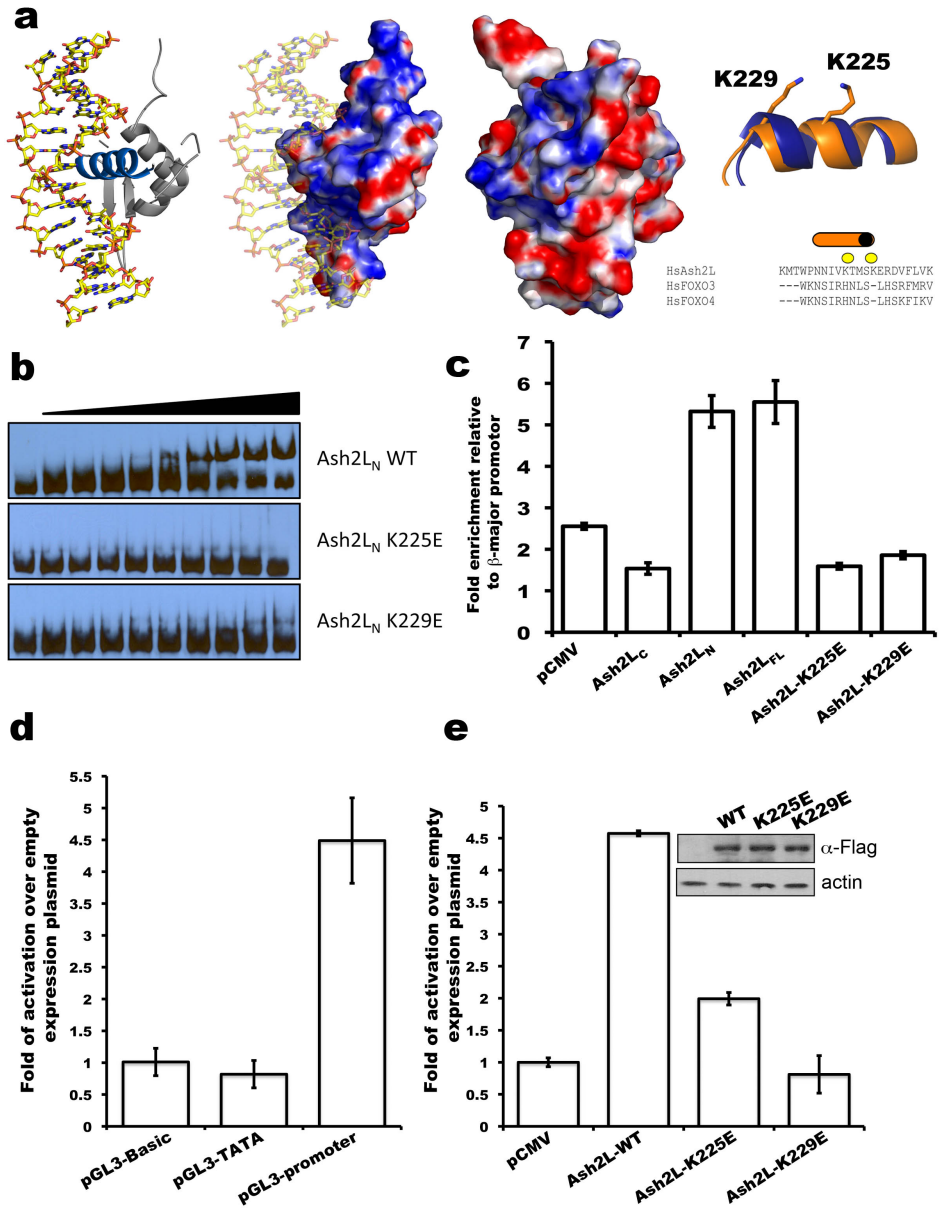


Figure 16. Ash2L helix-wing-helix domain is essential for its binding to b-globin LCR.

(a) Ash2L harbors an atypical helix-wing helix domain. Structural comparison of FOXO3 and Ash2L helix-wing-helix domain in which the DNA binding α-helix is rendered in blue and DNA carbon atoms are highlighted in yellow. Electrostatic potentials are contoured from $+10k_bTe^{-1}$ (blue) to $-10\ k_bTe^{-1}$ (red). Zoomed view and sequence alignment of the Ash2L (orange) and FOXO3/4 (blue) DNA binding α-helices.

(b) Two residues are key for Ash2L DNA binding activity. EMSA using $HS2_{Stru}$ with increasing amount of Ash2L (0.5-15mM) wild type or glutamate substituted Lys-225 or Lys-229 residues.

(c) Ash2L$_N$ is important for the recruitment of Ash2L to HS2. Relative recruitment of Ash2L was measured by ChIP at the β-globin LCR HS2 site. Immunopurified DNA was quantified as previously described (Demers et al, 2007). Average values of duplicate qPCR reactions are shown; error bars represent s.d. Each experiment was performed twice with independent chromatin samples.

(d) Ash2L stimulates luciferase expression. Hela cells were transfected with the indicated firefly luciferase reporter constructs and pCMV-Ash2L. Normalized firefly luciferase activities were divided by the protein concentration and reported as fold of induction over empty plasmid.

(e) Ash2L wild-type or mutants are detected by western blot with indicated antibodies (inset). Luciferase activity in cells transfected with pGL3-promoter and various constructs of Ash2L. Activity is reported as in d. Each experiment was performed a minimum of three times and error bars represent s.d.

Ash2L to *β-globin* LCR. To test this hypothesis, we performed ChIP experiments using differentiating myeloid erythroleukemia (MEL) cells transfected with DNA constructs corresponding to $Ash2L_{FL}$, $Ash2L_N$ or $Ash2L_C$. As shown in **Figure 16C** and as expected from previous studies (Demers et al, 2007), we detected a significant enrichment of $Ash2L_{FL}$ and $Ash2L_N$ at the HS2 site (**Fig. 16c**). Importantly, and consistent with the EMSA results shown above, ChIP assays revealed that $Ash2L_C$ and the Ash2L K225E and K229E mutants do not associate with the HS2 site. These results further highlight the important role of Ash2L DNA binding domain in its recruitment to a target locus.

We next set out to determine whether Ash2L can stimulate transcription by binding to DNA. Due to the enrichment of GC rich elements in our SELEX data, we took advantage of the presence of the pentanucleotide sequences 5′-GAGGC-3′/5′-GCCTC-3′ in the SV40 promoter of the pGL3-promoter plasmid. Co-transfection experiments using pGL3-basic, -TATA and – promoter vectors revealed that Ash2L stimulates luciferase expression by approximately 4.5-fold when co-transfected with pGL3-promoter specifically **(Fig. 16d)**. Luciferase reporter assays were then performed using Ash2L K225E and K229E mutants and, consistent with the EMSA and ChiP results, we found that K225E and K229E mutants were greatly impaired in their ability to activate transcription above control levels. Immunoblot analysis using an Anti-Flag (Sigma) antibody revealed that the loss of transcriptional activity of Ash2L mutants was not related to their instability, as a similar amount of protein was detected for all the mutants **(Fig. 16e)**.

2.2.4 - DISCUSSION

The structural, biochemical and functional analysis of Ash2L N-terminal domain presented herein have provided insights into the mechanism by which this trithorax protein

selectively binds DNA. Furthermore, our analysis showing that Ash2L preferentially binds torsionally stressed DNA has important ramifications for the regulation of *β-globin* gene. Indeed, initial structure-function studies performed on the *β-globin* LCR have revealed that this region, owing to its poor nucleosome contents (Kim et al, 2007) and its propensity to form secondary structures(Kukreti et al, 2010), displays sensitivity to DNAse 1 and nuclease S1. Other studies have posited that such DNA architecture could provide a docking site for proteins (McGhee et al, 1981). Our results showing that Ash2L preferentially binds bent DNA appears to confirm these hypotheses and suggest that the Ash2L helix-wing-helix domain is important for its binding to such specific loci.

The fine balance between the Polycomb group (PcG) and the Trithorax group (TrxG) complexes regulate *Hox* gene expression and, in contrast to Polycomb response element, identification of trithorax response elements (TRE) has been challenging. So far, only few DNA binding trithorax proteins have been identified which include GAGA (Matharu et al, 2103; Mulholland et al, 2003) and MLL1 (Broeker et al, 1996; Cierpicki et al, 2010). In addition to these studies, other reports have shown that DNA methylation (Cierpicki et al, 2010) or post-translational modifications of histone tails (Wang et al, 2010) provide the molecular signals underlying the positioning of specific trithorax proteins. Our results showing that Ash2L recognizes torsionally constrained DNA support an additional model in which DNA topology could play a role in the localization of trithorax proteins.

The recent identification of linkages between histone H3 Lys-4 methylation and several classic transcription factors, including Mef2D (Rampalli et al, 2007), Ap2δ (Tan et al, 2008), PAX-7 (McKinnell et al, 2008) and the NF-E2 complex (Demers et al, 2007), have suggested a unified mechanism in which the recruitment of an histone H3 Lys-4 methyltransferase, and

therefore recruitment of RNA polymerase II, is required to achieve maximal expression of a target gene. Our results presented herein support these observations and further suggest that $Ash2L_N$ may serve as a linking module between transcription factors and a member of the SET1 family of histone H3 Lys-4 methyltransferases. Binding of classic transcription factors to their response element could trigger the bending of the neighboring DNA, and thereby recruit Ash2L. In line with this model, a recent investigation revealed that HMG-domain protein iBRAF is essential for the recruitment MLL to NeuroD2 and synapsin promoters (Jonathan and Shilatifard 2005). Interestingly, it is known that cruciform DNA binding proteins, such as HMG proteins, induce the structural transition of DNA to a non-B structure (Stefanovsky and Moss 2009). In light of our results, Ash2L could sense this structural change and bridge the changes in DNA topology and recruitment of MLL. Alternatively, DNA bending through long-range structural reorganization may serve as a signaling event leading to the recruitment of Ash2L. Accordingly, long range interaction through DNA looping between β-globin LCR 5'region of hypersensitive sites and gene-proximal promoters (Carter et al, 2002; Tolhuis et al, 2002) are key in regulating β-globin genes. Correlatively, other genuine Ash2L targets, the *Hox* genes, are known to undergo structural rearrangements which lead to large extruding loops during differentiation (Ferraiuolo et al, 2010).

While DNA binding by Ash2L and other helix-wing-helix proteins appears mechanistically similar, especially with regards to the involvement of a positively charged α-helix, many differences also exist. For example, several helix-wing-helix proteins use their wing portion of the domain to bind DNA (Littlefield & Nelson, 1999; Littler et al, 2010). Considering that Ash2L β5-β6 wing is negatively charged (**Fig. 18a**), it is unlikely that this region contributes to the DNA binding activity. In addition, FOXO's ability to bind DNA is conferred by a

combination of hydrophobic and positively charged residues (Obsil & Obsilova, 2008). Ash2L DNA binding domain, and in particular α5-helix, contains only one hydrophobic residue, suggesting that it will bind DNA differently. In that sense, Ash2L helix-wing-helix domain is unique and will likely bind DNA atypically, employing a yet to determine DNA binding mode. The unique aspects of Ash2L N-terminal domain are further supported by its ability to bind torsionally stressed DNA which is in clear contrast to all the FOXO's ability to bind DNA duplexes.

2.2.5 - METHODS

I - Electrophoretic Mobility Shift Assays. DNA binding assays were performed using biotinylated DNA duplexes (Europhin MWP Operon) encompassing the hyper sensitive 2 (HS2) site of the β-globin gene (Demers et al, 2007) (5'-GCAATGCTGAGTCATGATGAGTCATGCT ACTGGGTAGGGTGTGTGC-3'). Binding reactions were carried out in 20mM Hepes pH 8.0, 50mM KCl, 5% glycerol, 10mM ZnSO4, 2mM β-mercaptoethanol, 2mM of freshly purified $Ash2L_N$ and 1nM of biotinylated DNA. Binding assays were initiated by the addition of the protein and incubated for 90 minutes on ice. The complexes were separated on a 7.5% acrylamide gel during 1 hr at 100V in 0.5x TGE at 4ºC. The polyacrylamide gel was blotted onto a Biodyne B (VWR) membrane at 100V for 30minutes in 0.5X TBE. After UV cross-linking (Stratagene), biotinylated DNA was detected according to the nucleic acid detection protocol from Pierce.

II - Protein purification and crystallization. PCR-amplified from the IMAGE clone #3921999 (Open Biosystem), a fragment encoding residues 95-280 of Ash2L was subcloned in a modified

version of pET3d (Novagen) vector (Couture et al, 2007). The fragment was overexpressed as a TEV-cleavable N-terminal Hexahistidine-tag fusion protein using Bl-21 E. coli cells (Stratagene). The protein was produced by inducing with 0.1mM IPTG for 16h at 18ºC and purified using TALON-Co^{2+} (Clontech). Following TEV-cleavage, the protein was reapplied onto TALON-Co^{2+} and the protein was further purified by size-exclusion chromatography using a Superdex 75 (GE Healthcare). The protein was concentrated to 80mg ml^{-1} and stored at -80ºC. Selenomethionyl-derivatized $Ash2L_N$ protein was obtained according to the manufacturer (Molecular Dimensions) using the methionine auxotroph B834 cells. Protein was purified as for the wild type and concentrated to 10mg ml^{-1}. Native and SeMet crystals were grown in 0.2M MgAcetate and 17-20% PEG 3350 at 4ºC. Crystals were harvested, soaked in 0.2M MgAcetate and 35% PEG3350 and flash frozen in liquid nitrogen. Mutants were produced using the Site-directed mutagenesis kit (Stratagene) and purified as previously described for the wild-type protein.

III - Data collection and structure determination. A single-wavelength anomalous dispersion (SAD) data set was collected at the 21-ID-G beamline of Life Science-Collaborative Access Team at the Advanced Photon Source Synchrotron. Data were collected on a single crystal of the selenomethionyl-derivatized $Ash2L_N$ protein at 0.9785Å with a MarMosaic300 CCD detector (Rayonix) and subsequently processed and scaled using Denzo and Scalepack (Otwinowski & Minor, 1997) (Table 3). The heavy atom sites were determined using the program HYSS from Phenix (Zwart et al., 2008) and subsequent phases were determined and refined with SHARP (de La Fortelle & Bricogne, 1997), DM/Solomon(Abrahams & Leslie, 1996; Cowtan & Main, 1993), and an initial model was auto built with ARP/wARP (Perrakis et al, 2001). The Ash2L model

was further refined using iterative cycles of model building and structure refinement using COOT (Emsley & Cowtan, 2004) and REFMAC (Vagin et al, 2004), respectively.

I*V - Systematic Evolution of Ligands by Exponential Enrichment (SELEX). In vitro* selection of aptamer was performed using 70-mer oligonucleotides (GCCGCAGATAGTGGATTCTAGAN_{20}CTCGAGGGGTCCTAGAAGGCAT) carrying a 20-mer long randomized sequence (Europhin MWP Operon). Each wobble nucleotide has been manually incorporated during synthesis to ensure equimolar ratios between purine and pyrimidine bases (25:25:25:25). Seven iterative rounds of SELEX were performed with GST-bound Ash2L_N and 10mM of DNA duplex. Enriched DNA fragments were subsequently cloned in pBluescript and a total of 190 clones (20-mers) were sequenced. Exact duplicates (comparing both strands) were filtered out, leaving 148 clones that were further analyzed. The program Amadeus was used to identify sequence motifs specifically enriched within this group of sequences (Linhart et al, 2008), and assumed to represent the DNA binding site of Ash2L in this assay. Both DNA strands were analyzed. As a background set, a group of 12,000 random 20 bp sequences was generated. Motifs of varying lengths were searched (from 6 to 12 base pairs long). One 10 bp motif emerged as specifically enriched, as evidenced by fixed p-values (after 20 bootstrap cycles) of 9.4×10^{-38}.

V - Prediction of Ash2L target genes. In order to predict Ash2L genomic targets, the position weight matrices (PWMs) of each motif were used to scan the entire human genome for target loci. The Cisgenome program was used for this search (Ji et al., 2008). The parameters of the search imposed a likelihood ratio of at most 1 in 10,000 and a phylogenetic conservation among

the top 10% of all sequences. This resulted in the identification of 9761 loci corresponding to Motif 1, and 9335 loci corresponding to Motif 2. PCR primers were designed to flank some of these loci, in order to verify by ChIP assays whether the endogenous Ash2L protein is recruited to them.

The genes closest to these loci (relative to their transcriptional start sites) were identified, and functional annotation was performed using Gene Ontology categories. Analysis of category enrichment was performed using the NIH DAVID tool (Dennis Jr et al, 2003). Five sets of randomly selected genes (equal in numbers to the Ash2L predicted target genes) were created as controls, using this specific function within Cisgenome.

VI - Chromatin Immunoprecipitation experiments using Hela cells. Chromatin immunoprecipitation experiments were performed as previously described (Blais et al, 2007). Briefly, chromatin extraction, protein cross-linking and immunoprecipitation were performed with an anti-Ash2L antibody and protein A agarose (Zymed). Reverse crosslinked DNA was PCR-amplified using gene-specific primers listed in supplementary Table 1.

VII - Transient Transfection & Chromatin Immunoprecipitation using mouse erythroleukemia cells. Mouse Erythroleukemia (MEL) cells were transfected by electroporation using 16µg of the following expression constructs: FLAG-empty (vector name), FLAG-Ash2L$_C$ (WDR5-RbBP5 interacting domain), FLAG-Ash2L$_N$ (DNA binding domain), FLAG-Ash2L-WT, FLAG-Ash2L-K225E or FLAG-Ash2L-K229E. DMSO-induced erythroid differentiation was initiated 6h later and allowed to proceed for 3 days as previously described (Demers et al, 2007). Crosslink of cells, chromatin extraction and immunoprecipitation were performed as previously described

except that the anti-Flag-M2 magnetic beads (SIGMA) were used. Real-time qPCR analysis was done on a Rotorgene instrument 6000 (Corbett research) using Taqman probes and primers (Demers et al, 2007). Binding of FLAG tagged protein to the HS2 site of the ß-globin LCR (measured using the standard curve method) is expressed 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).

VII - Luciferase reporter assays. HeLa cells were grown at 37ºC in Dulbecco's Modified Eagle medium supplemented with 10% fetal bovine serum, 2mM of glutamine and Penn/Strep mixture. HeLa cells were seeded at $1.5x10^4$ cells per well in a 24-well plate (Corning). After 24 hours, cells were transiently co-transfected with 1mg of pCMV-3X FLAG expressing either wild type or mutants of Ash2L and 0.5mg of the pGL3 promoter using Lipofectamine 2000 (Invitrogen). After 24 hours, cells were harvested according to the manufacturer (Promega) and luciferase activity was measured on a Glomax 96 microplate luminometer. Light intensity was normalized with protein concentration and each assay represents the mean of three independent experiments performed in triplicate.

2.2.6 - ACKNOWLEDGEMENTS

We wish to thank Dr. Jeffrey Dilworth for the Ash2L antibody and other reagents. We are highly indebted to Dr. Raymond C. Trievel laboratory, where part of this work has been initiated. This work was supported by a Canadian Institutes of Health Research grant to J.-F. C. J.-F. C. holds a Canadian Research Chair in Structural Biology and Epigenetics.

2.3 - Fine-tuning the stimulation of MLL1 methyltransferase activity by a histone H3 based peptide mimetic.

Vanja Avdic[1,2*], Pamela Zhang[1,2*], Sylvain Lanouette[1,2], Anastassia Voronova[2], Ilona Skerjanc[2] and Jean-Francois Couture[1,2]

1) Ottawa Institute of Systems Biology

2) Department of Biochemistry, Microbiology and Immunology, University of Ottawa

Corresponding author: Jean-Francois Couture

FOOTNOTES

*Authors equally contributed to this work.

Running title: Inhibition of MLL1 methyltransferase activity.

2.3.1 SUMMARY

The SET1 family of methyltransferases carries out the bulk of histone H3 Lys-4 methylation *in vivo*. One of the common features of this family is the regulation of their methyltransferase activity by a tri-partite complex composed of Ash2L, RbBP5 and WDR5. To selectively probe the role of the SET1 family of methyltransferases, we have developed a library of histone H3 peptide mimetics and report herein the characterization of an N alpha acetylated form of histone H3 peptide (NαH3). Binding and inhibition studies reveal that addition of an acetyl moiety to the N-terminus of histone H3 significantly enhances its binding to WDR5 and prevents the allosteric regulation of MLL1 by the Ash2L-RbBP5-WDR5 complex. The crystal structure of NαH3 in complex with WDR5 reveals that a unique high affinity hydrophobic pocket accommodates the binding of the acetyl moiety. These results provide the structural basis to control MLL-WDR5-Ash2L-RbBP5 activity and a tool to manipulate stem cell differentiation programs.

2.3.2 INTRODUCTION

Histone methylation plays an important role in a large variety of biological processes. Depending on its position on a substrate and the number of methyl groups transferred to its ε-amine, lysine methylation has been associated with DNA damage response (Yang & Mizzen, 2009), DNA replication (Probst et al, 2009) and transcription (Martin & Zhang, 2005). Of these positions, lysine 4 of histone H3 (H3K4) is a key methylation site and commonly linked to actively transcribed genes. In humans, the SET1

family of methyltransferases catalyzes the bulk of H3K4 di/tri-methylation and includes Ash1, SET1A/B and five Myeloid Lymphoma Leukemia proteins (MLL1-5) (Cosgrove & Patel, 2010).

Of these five MLL proteins, MLL1 has remained the most extensively characterized member. MLL1 is essential for proper development during embryogenesis (Yu et al, 1995). Homozygous disruption of the mouse *MLL1* gene is embryonic lethal while the heterozygotes display retarded growth and haematopoietic abnormalities (Terranova et al, 2006b; Yu et al, 1995). Correlative to these observations, MLL proteins regulate several differentiation programs, including hematopoiesis (Argiropoulos & Humphries, 2007) myogenesis (McKinnell et al, 2007; Rampalli et al, 2007) and neurogenesis (Lim et al, 2009).

MLL1 and its yeast homolog SET1 (ySET1) harbor several evolutionarily conserved domains and are virtually always found in multi-subunit complexes. Interactions with other protein subunits regulate their cellular localization and enzymatic activity (Popovic & Zeleznik-Le, 2005). The catalytic domain of MLL1, also referred to as the Suppressor of variegation 3-9, Enhancer of zeste and Trithorax (SET) domain (Couture & Trievel, 2006), is located on its C-terminus and is pivotal in controlling skeletal development and *hox* gene expression (Terranova et al, 2006a). Similar to other members of the SET1 family, MLL1 has minimal methyltransferase activity toward its substrate in absence of its protein cofactors. Upon interaction with a three-protein complex composed of Ash2L, RbBP5 and WDR5, MLL1 methyltransferase activity is enhanced (Dou et al, 2006; Patel et al, 2009; Southall et al, 2009b).

The biochemical determinants underlying the formation of the MLL complex have recently been detailed. Initial work carried out by Dou et al. showed that each subunit contributes in modulating the enzymatic activity of MLL1 and the level of histone H3 Lys-4 methylation *in vivo* (Dou et al, 2006). Notably, the same authors also demonstrated the importance of WDR5 within the core complex as its depletion impaired MLL1 methyltransferase activity. Similarly, Patel and coworkers found that WDR5 is central in binding a region on MLL1 termed the WDR5 INteracting motif (WIN) and bridging the methyltransferase to the other subunits of the core complex (Patel et al, 2009; Patel et al, 2008b). Binding of the WIN peptide is achieved through a peptidyl argininine-binding pocket located in one of the central clefts of the β-propeller. Site directed mutagenesis studies of the MLL WIN region and WDR5 peptidyl arginine binding cleft revealed that loss of binding between MLL1 and WDR5 caused severe impairment in MLL-1 methyltransferase activity, further underscoring the importance of the role of WDR5-MLL1 interaction (Dou et al, 2006; Martens & Stunnenberg, 2010; Patel et al, 2008b).

The detrimental effects of misregulating histone H3 Lys-4 methylation have been documented. The *MLL* gene is prone to chromosomal aberrations such as translocations and duplications which are associated with aggressive leukemic disorders (Martens & Stunnenberg, 2165). Alterations of H3K4me3 levels are observed in peripheral blood mononuclear cells from systemic erythematosus patients (Dai et al, 2010;Greer et al, 2010). In *C. elegans*, H3K4 methyl marks leads to a reduction in life span (Greer et al, 2010). The involvement of the SET1 family of methyltransferases in various diseases

underlines the importance of better understanding their functions and eventually modulates their allosteric regulation.

In this study, we report the characterization of a histone H3-based peptide mimetic on MLL1 methyltransferase activity. We found that the addition of an N-alpha acetyl moiety to histone H3 N-terminus (NαH3) increases its affinity for WDR5 by 84-fold compared to the unacetylated form of the histone H3 peptide. Determination of the WDR5-NαH3 peptide mimetic complex structure shows that the NαH3 peptide undergoes structural reorganization to fit in a shallow pocket that stabilizes the acetyl moiety using novel hydrophobic contacts and hydrogen bonds. We also show that incubation of the NαH3 peptide mimetic with WDR5 impairs the allosteric regulation of MLL1 methyltransferase activity. These results provide a basis for the use of histone H3 peptide mimetics in regulating MLL1 activity and present a new tool for the study of biological processes regulated by MLL1 di-/tri-methyltransferase activity.

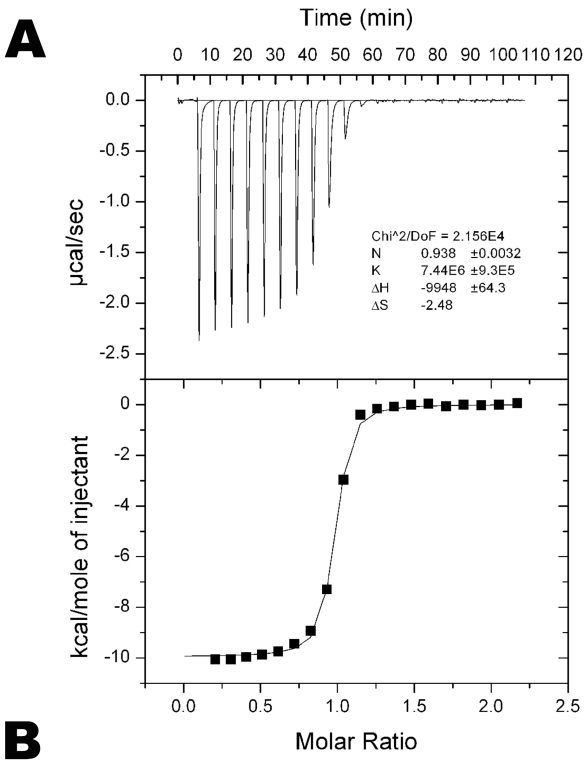
2.3.3 RESULTS

I - An Nα-acetyl moiety and C-terminal truncation allow high affinity association to WDR5

Basing ourselves on the results showing that MLL1-WDR5 interaction is essential for proper histone H3 Lys-4 methylation (Patel et al, 2008b), we posited that the development of a high affinity peptide mimetic targeting WDR5 would prevent its interaction with MLL1 and impair methyltransferase activity. To verify this hypothesis, we determined the affinity and inhibitory effect on MLL methyltransferase activity of a small library of peptides (data not shown). All peptides designed included an arginine

residue in position 2, which is considered to be essential for binding to WDR5 (Couture et al, 2006b; Patel et al, 2008a; Ruthenburg et al, 2006; Song & Kingston, 2008b). Using isothermal titration calorimetry (ITC) (**Fig. 17A**), we found that an N-alpha acetylated form of an histone H3 (NαH3) peptide binds WDR5 with a dissociation constant (K_d) of 0.13μM. This is a marked decreased in comparison with the K_d of the unacetylated form of a histone H3 peptide (K_d = 11μM) and MLL1-WIN (K_d = 1.6μM) peptides (Couture et al, 2006b; Patel et al, 2008b) (**Fig. 17C**). We then sought to verify whether this peptide could prevent the allosteric regulation of MLL1 enzymatic activity. Using a radiometric methyltransferase, we determined that NαH3 peptide inhibits MLL1 methyltransferase activity with an IC50 of 4.06μM. These results confirm that blocking the WDR5 peptidyl arginine binding cleft provide a way to control, *in vitro*, the allosteric regulation of MLL1 methyltransferase activity by its core complex subunits. The residual activity observed above the background level also suggests that MLL1 mono-methyltransferase activity, which does not require the association of MLL1 with the core complex subunits (Patel et al, 2009), is not affected by the NαH3 peptide.

These initial results indicate that adding an acetyl moiety to the N-terminus of histone H3 improves the binding capabilities of a peptide mimetic to WDR5. However, an initial modeling of an NαAc moiety on the available WDR5-H3 crystal structure (rcsb code 2H13.pdb) (Couture et al, 2006a) revealed numerous steric clashes between the WDR5 side chains and the NαAc moiety (data not shown), suggesting that the NαAc H3 peptide must undergo significant structural rearrangement to bind the peptidyl arginine binding cleft.



Binding and inhibitory analyses of NαH3.	
α-$A_1R_2T_3K_4Q_5T_6A_7R_8K_9S_{10}Y$	Binding affinities
K_D(µM)	0.13 ± 0.18
IC50 (µM)	4.06 ± 0.9

Figure 17: Binding profiles of NαH3.

A) ITC titration experiment with NaH3 peptide and WDR5 (*upper* trace) and the fitted binding curve (*lower* trace).

B) Inhibition assays of MLL1 methyltransferase activity with increasing concentration of NaH3 peptide and 5mM of MLL1 and its core complex subunits.

C) The equilibrium dissociation constant and IC50 values for the binding of NaH3 to WDR5 and inhibition of MLL1 methyltransferase, respectively.

II - A novel peptide conformation is observed in the crystal structure of WDR5-NαH3 complex.

To understand the structural basis underlying the recognition of NαH3, we have solved its crystal structure in complex with WDR5 (Fig. 18, Table 4). Overall, the WDR5 structure is similar to the previously published apo-structure with root-mean-square deviations of 0.42 Å for all protein atoms. The density of the peptide is unambiguous (**Fig. 18A**) for the first three amino acids including the Nα-acetyl moiety while there is no visible electron density for Lys-4 ε-amine and Gln-5 residue. The peptide is maintained in the central peptidyl arginine-binding cleft of WDR5 in which it engages in several hydrogen bonds and hydrophobic contacts. Briefly, NαH3 Thr-3 makes van der Waals contacts with residues delineating the cleft including Tyr-260, Leu-321, Ala-47 and Ala-65 side chains (**Fig. 18B & D**). As observed for the WDR5-H3 and WDR5-WIN complexes, Arg-2 of NαH3 is maintained in the central cavity of the β-propeller by a pair of phenylalanine residues.

Substitutions in position 2 for a leucine, a glutamic acid and an alanine residue were assayed to determine whether Arg-2 is still required for association of the Nα-acetylated peptide (Table 5). As shown in Table 5, no binding could be detected for the R(2)L and appears to be essential for the association of NαH3 peptide mimetic to WDR5 and concur with the previous characterization of WDR5-WIN and WDR5-H3 complexes (Couture et al, 2006b; Han et al, 2006b; Patel et al, 2008a; Ruthenburg et al, 2006; Song & Kingston, 2008b).

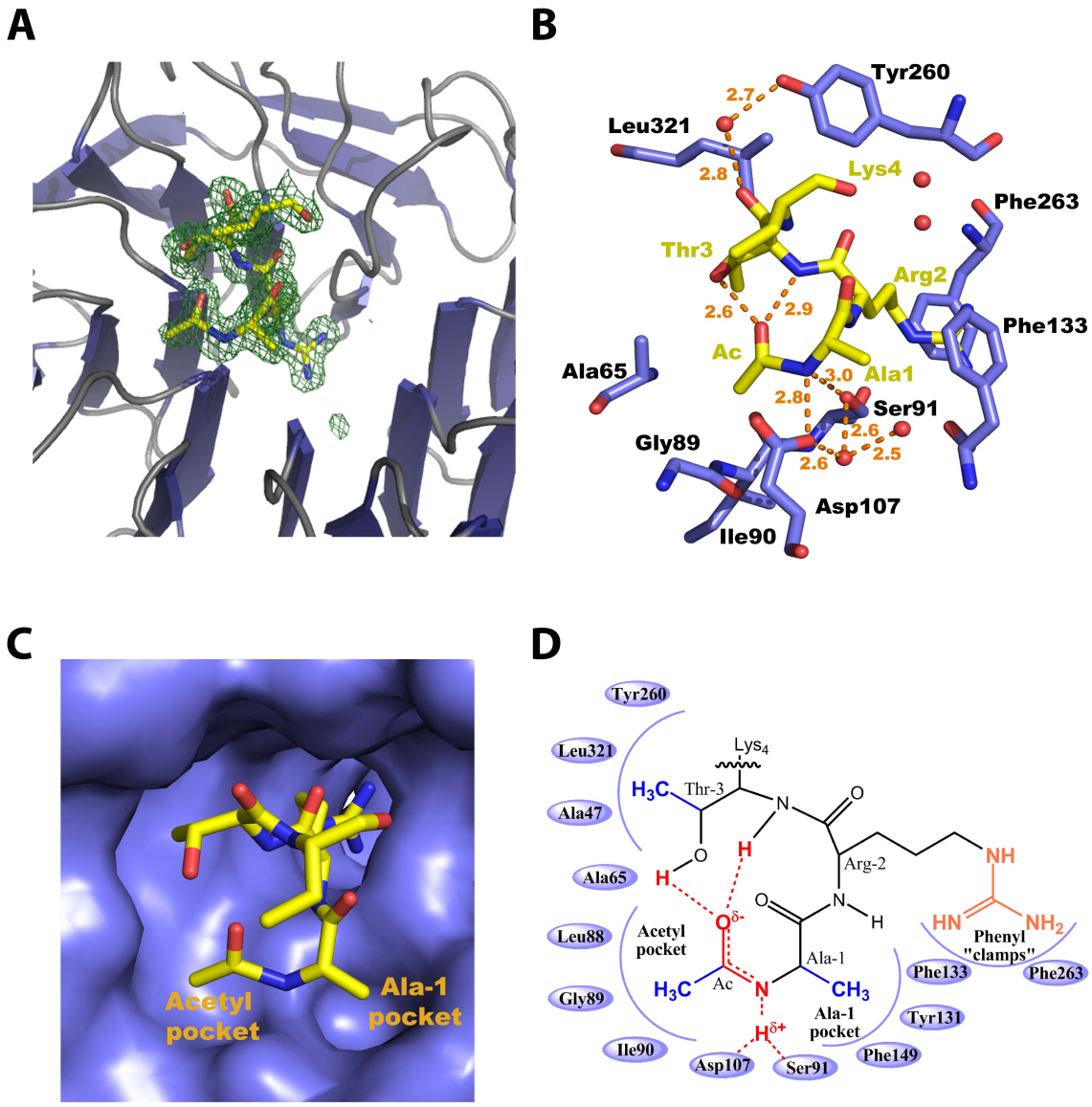


Figure 18: WDR5 binds NαH3 through the peptidyl arginine-binding cleft.

A) Zoom view of a simulated annealing $F_o - F_c$ omit map (green) contoured at 2σ.

B) View of the Nα-acetylated histone H3 binding pocket in which WDR5 carbon atoms are colored in blue. Hydrogen bonds are rendered as orange dash lines.

C) Surface representation of the WDR5 peptidyl arginine-binding pocket showing the three pockets.

D) Shown is a 2-D schematic representation of important hydrogen bonds and resonance of the new amide bond.

Table 4: Data collection and refinement statistics	
	WDR5 – NaH3*
Data Collection	
Space group	P1
Cell dimensions	
a, b, c (Å)	45.9, 48.6, 63.3
a, b, g (◦)	98.8, 90.9, 117.5
Resolution	40.0 -1.7 (1.76-1.70) !
R_{merge}	6.5 (23.2)
I / σI	20.6 (2.4)
Completeness (%)	97.0 (93.8)
Redundancy	3.1 (2.4)
Refinement	
Resolution (Å)	35.65 – 1.70
No. Reflections	47864
R_{work}/ R_{free}	17.1 / 21.2
No. Atoms	
Protein	4601
Peptides	67
Water	351
B-factors (Å^2)	
Protein	17.7
Ligands	15.6
Water	31.7
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (◦)	1.108
Molprobity scores	
Ramachandran favored (%)	96.6
Ramachandran allowed (%)	3.4
*Data have been collected at GM/CA-CAT (Argonne National Laboratory)	
! Highest resolution shell is shown in parentheses.	

Table 5: Mutational analysis of the Nα-acetylated peptides.

α-$A_1R_2T_3K_4Q_5A_6R_7K_8S_9Y$	**Kd (μM)**
WT	0.13 ± 0.02
A1S	0.09 ± 0.01
A1G	1.30 ± 0.18
R2L	N.B.[a]
R2A	N.B.
R2E	>500[b]
T3V	0.10 ± 0.01
A1G/R2S	N.B.

(a) : No heat of binding could be detected

(b) : Saturation could not be achieved.

pocket (**Fig. 18D**), composed of Phe-133, Phe-149 and Tyr-131 phenyl rings. In addition, Ala-1 amide group is maintained by two 2.8Å and 3.0Å hydrogen bonds with Asp-107 carboxylate and Ser-91 hydroxyl groups respectively. To probe the role of Ala-1 in binding WDR5, the equilibration dissociation constant was determined for a NαH3 A(1)G mutant. Using ITC, we show that removal of the methyl group results in a 10-fold decrease in affinity (Table 5). In contrast, substitution for a serine residue does not appear to affect the dissociation constant of the peptide, thus underscoring the importance of a small residue in position 1 for high affinity binding to WDR5.

Unambiguous electron density for this region of the peptide allowed us to accurately model the Nα acetyl moiety. Initial analysis of the acetyl group reveals that it has comparable geometry to an amide bond, as its carbonyl group is planar with the nitrogen atom of Ala-1. This geometry allows the methyl group of the Nα acetyl moiety to snuggly fit into a hydrophobic cleft, henceforth referred as the Nα acetyl pocket, composed of Ala-65, Leu-88, Gly-89 and Ile-90 (**Fig. 18D**). The carbonyl group of the acetyl moiety is further stabilized by two intra-molecular hydrogen bonds of 2.6Å and 2.9Å with Thr-3 hydroxyl and amide groups, respectively. This hydrogen bonding pattern concurs with a recent modeling analysis and molecular dynamics study showing that two intra-molecular hydrogen bonds are important to maintain high affinity binding of peptide mimetics to WDR5 (Karatas et al, 2010). To further probe the role of the hydrogen bond between the carbonyl group of the acetyl moiety and Thr-3 hydroxyl group, binding studies were performed with a T(3)V substituted peptide. To our surprise, no loss of binding was observed as identical binding constants were calculated for the mutant NαH3 peptide (Table 5). These results suggest that a single hydrogen bond is

sufficient to maintain high affinity binding capability of the NαH3 peptide. Overall, the mapping and structural analyses suggest that for high-affinity binding, a peptide mimetic requires an Nα-acetyl moiety, a residue with a small side chain other than a glycine in position 1 and an arginine residue in position 2.

III - The NαH3 binding mode is unique.

Structural alignment and analysis of the WDR5-NαH3 complex with the previously described WDR5-WIN and WDR5-H3 complexes allowed us to further understand the structural basis underlying the differences in equilibrium dissociation constants between MLL1, histone H3 and NαH3 peptides. Overall, the structure of WDR5-NαH3 is similar to the previously published structure of WDR5-WIN and WDR5-H3 with root-mean-square deviations of 0.33Å and 0.34Å for all atoms, respectively. However, close inspection of the peptides revealed several conformational differences. The additional van der Waals contacts of Ala-1 side chain in its aromatic pocket are permitted by a 180° rotation along its Cα-C bond comparatively to H3 peptide (**Fig. 19A)**. This rotation places Ala-1 amide group within hydrogen bond distances with Ser-91 and Asp-107 hydroxyl and carboxylate groups, respectively. Moreover, the addition of the N-terminal acetyl moiety also permits new van der Waals contacts in the WDR5 peptide binding cleft. Furthermore, the carbonyl group of the acetyl moiety allows the formation of intra-molecular hydrogen bonds with Thr-3, inducing a tighter packing of the peptide. However, close inspection of the peptides revealed several conformational differences. The additional van der Waals contacts of Ala-1 side chain in its aromatic pocket are permitted by a 180° rotation along its Cα-C bond comparatively to H3 peptide (**Fig. 19A)**. This rotation places Ala-1 amide group within hydrogen bond distances with

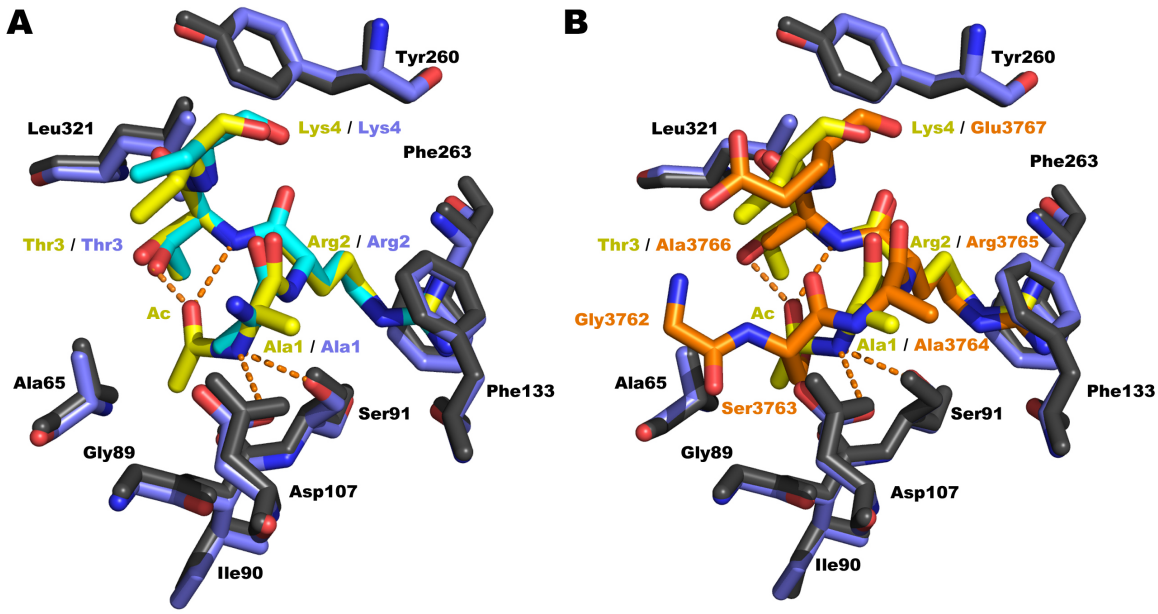


Figure 19: NαH3 binding mode is unique.

A) A comparison of WDR5-H3 and WDR5-NαH3 complexes in which carbon atoms are rendered in yellow and cyan for NαH3 and H3 complexes, respectively.

B) Comparison of WDR5-NαH3 and WDR5-WIN complexes in which WIN carbon atoms are colored in orange. Hydrogen bonds are rendered as in Figure 2.

Ser-91 and Asp-107 hydroxyl and carboxylate groups, respectively. Moreover, the addition of the N-terminal acetyl moiety also permits new van der Waals contacts in the WDR5 peptide binding cleft. Furthermore, the carbonyl group of the acetyl moiety allows the formation of intra-molecular hydrogen bonds with Thr-3, inducing a tighter packing of the peptide.

The equilibrium dissociation constant calculated for the NαH3 peptide is approximately 13-fold lower than the MLL-WIN peptide. Although the number of contacts with WDR5 is greater for the MLL-WIN peptide (Patel et al, 2008a; Song & Kingston, 2008a) the differences in binding affinities between these peptides can be rationalized by the peptide-binding mode employed by NαH3. Close inspection of the overlay of WDR5-WIN and WDR5-NαH3 reveals a clear shift of 1Å inward the propeller cleft for NαH3 Ala-1 in comparison to MLL Ala-3764 (**Fig. 19B**). This shift places the Ala-1 amide group closer to Ser-91 and Asp-106 hydroxyl and carboxylate groups, respectively, which consequently leads to shorter hydrogen bonds. In the WDR5-MLL1-WIN complex, residues preceding Ala-3764 stack onto Ala-65. This additional interaction could force the peptide to adopt a more relaxed conformation in comparison to NαH3. Collectively, our observations point to a model in which the compaction, achieved by a shift of Ala-1 and an intramolecular hydrogen bond between the carbonyl group of the acetyl moiety and Thr-3 amide proton, of NαH3 in the peptidyl-arginine cleft, plays an important role in high affinity binding to WDR5.

2.3.4. DISCUSSION

The MLL1 protein catalyzes methylation of histone H3 on Lys-4, an enzymatic activity required for proper control of skeletal development (Terranova et al, 2006a). As MLL1 methyltransferase activity is dependent on its association with the core complex, the disruption of this association is an attractive means to modulate lysine methylation *in vivo*. Since this assembly relies on WDR5 peptide-binding ability, a peptide mimetic with a high affinity for its peptidyl-arginine cleft would be expected to impair MLL1 activity. As demonstrated for many other protein targets (Chonghaile & Letai, 2008; Cragg et al, 2009; Mendez, 2010), an understanding of the structural determinants of peptide binding is key to the design of such mimetics. In order to achieve this, we sought to characterize the interaction of WDR5 interaction with specifically designed peptides.

Our results show that the addition of an acetyl moiety to the N-terminus of a histone H3 peptide increases its binding affinity to WDR5 by 84-fold. The same peptide has a 13-fold higher affinity comparatively to the previously characterized "WIN" peptide. These results are supported by a recent study by Karatas et al., who also observed a higher affinity for short Nα-acetylated peptides mimetics of histone H3 and MLL WIN motifs (Karatas et al, 2010).

I - Three binding pockets within WDR5 control its high affinity binding to the peptide mimetic.

Comparison of the WDR5-peptide complexes and the probing analyses provided several insights on the structural determinants required for the high affinity binding of NαH3 to WDR5. First, the arginine in position 2 is an obligatory element for binding

WDR5 as its substitution to leucine, glutamic acid and glycine residues completely abrogated binding to WDR5. These observations are supported by a comparison of WDR5-peptide structures which reveals that an identical network of direct and water-mediated hydrogen bonds as well as hydrophobic interactions maintain the arginine side chain in the central aperture of WDR5 (Couture et al, 2006a; Han et al, 2006a; Ruthenburg et al, 2006; Schuetz et al, 2006). Second, the addition of an acetyl group to the histone H3 N-terminus allows intra-molecular hydrogen bonds in the peptide structure. Consistent with the predictions made by Karatas et al., two intra-molecular hydrogen bonds are observed in the crystal structure. However, our mutational analysis suggests that the hydrogen bond between the carbonyl group of the Nα-acetyl moiety and Thr-3 hydroxyl group is dispensable for high affinity binding to WDR5. This discrepancy may be attributed to the change in polarity of the peptidyl arginine-binding pocket as substitution of Thr-3 to a valine residue decreases the polarity of the microenvironment whereby reinforcing the hydrogen bond between Thr-3 amide proton and NαAc carbonyl group (Gao et al, 2009). Collectively, our studies suggest that modulating intra-molecular contacts will influence the packing of the peptide in the WDR5 peptidyl-arginine binding cleft, allowing stronger hydrogen bonds between Ala-1 amide group and WDR5 residues.

Third, these bonds are likely to be strengthened further by the delocalization of Ala-1 nitrogen atom electrons along the amide bond created by the addition of the acetyl moiety. This conformation is further stabilized by a delocalization of the charge through Thr-3 hydrogen bond with the carbonyl oxygen of NαAc and Asp-107 carboxylate, hydrogen bond with the amide proton (**Fig. 18D**). The network of water-mediated

hydrogen bonds with the Ser-91 hydroxyl group likely stabilizes the free electron doublet of Asp-107 carboxylate group.

Fourth, the formation of an amide bond orients the Ala-1 side chain, Thr-3 side chain and the methyl group of the Nα acetyl moiety toward three new hydrophobic pockets where they engage in additional van der Waals contacts. The importance of these pockets is exemplified by the 10-fold decrease in affinity after mutating Ala-1 to a Gly residue, thus impairing the hydrophobic contacts with Phe-133, Phe-149 and Tyr-131 side chains. The additional interactions made by NαH3 with Ala-65, Leu-88, Gly-89 and Ile-90 side chains are unique and are likely the underlying mechanisms controlling the high affinity binding to WDR5. The addition of an acetyl group in the N-terminus is also likely to lower the energy of desolvation comparatively to a free N-terminus group.

II - NαH3 as a lead molecule?

The combination of *in vitro* methyltransferase and binding assays as well as structural studies reveal that NαH3 potently inhibits the allosteric regulation of MLL1 methyltransferase activity. The crystal structure of NαH3 bound to WDR5 gives us a rationale to design new peptide mimetics with high binding capabilities. The proximity of Thr-3 side chain and the acetyl moiety suggests that cyclisation of NαH3, through the formation of an oxaloacetate or malonate bridges, would increase the binding of the peptide to WDR5 by inducing a tighter packing in the binding cleft. Notwithstanding that NαH3 does not penetrate cells efficiently (data not shown), it is likely that inducing a planar geometry to the peptide would likely increase its ability to penetrate the cells.

In conclusion, our data offer the first structural basis for controlling the assembly of WDR5-mediated protein scaffolding with an unnatural peptide mimetic and hold

promise for future studies of biological processes related to methyltransferase activity of the SET1 family.

2.3.5. EXPERIMENTAL PROCEDURES

I. Peptides - Peptides were purchased either from New England Peptides or Peptide2.0. Peptides were ordered with 95-99% purity with an additional tyrosine on their C-terminus for UV quantification. Peptides were suspended in water at 100mM concentration and stored at -20ºC.

II - WDR5 expression, purification, crystallization and structure determination.

A fragment of WDR5 corresponding to residues 22-334 was overexpressed and purified as previously described (Couture et al, 2006a). WDR5 and the peptides were mixed in equal molar ratio and incubated on ice for 15 minutes. Crystals were grown at 22ºC in a buffer containing 50 mM NaAcetate, 100 mM NH_4SO_4 and 20% PEG4000, pH 4.6. Crystals were harvested and soaked in a cryoprotecting solution composed of the mother liquor supplemented with 20% glycerol. Complete data sets were collected at the General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) at the Advance Photon Source. Reflections were integrated and scaled using HKL2000 (Otwinowski & Minor, 1997). Using Molrep (Vagin & Teplyakov, 2000), a molecular replacement solution has been found using the apo-WDR5 structure (rcsb code 2H14.pdb) as a search model. After two rounds of refinement with REFMAC5 (Vagin et al, 2004), the structure and electron density map were inspected using COOT (Emsley & Cowtan, 2004). The structure has been solved at 1.7Å with final R_{factor}/R_{free} of 17.1/21.1

and two molecules in the asymmetric unit. Geometric parameters were calculated using Molprobity (Chen et al, 2009)(**Table 4**).

III - Isothermal titration calorimetry and inhibition studies **-** Equilibrium dissociation constants were obtained as previously described (Couture et al, 2006a). Briefly, binding of NaH3 peptide was measured using a VP-ITC microcalorimeter from MicroCal with 0.5-1mM of peptide and 30-50 mM of WDR5. Equilibrium dissociation constants were calculated with the Origin software.

IV - Overexpression and purification of MLL1 SET domain, Ash2L and RbBP5.

Fragments corresponding to full-length Ash2L and RbBP5 were PCR amplified using the OpenBiosystem clone #3921999 and #5266066, respectively. PCR fragments were cloned in a modified version of pET3d (Couture et al, 2008) and the proteins were overexpressed with 0.1mM IPTG in Rosetta cells (Novagen) during 16 hours at 18°C. Cells were harvested in 50mM sodium phosphate, 500mM NaCl and 5mM β-mercaptoethanol, lysed by sonication, clarified by centrifugation and purified by Talon Co^{2+} affinity chromatography. TEV-cleaved Ash2L and RbBP5 were further purified by metal affinity and by size exclusion chromatography steps (Superdex 200).

A DNA fragment encoding the WIN and SET domain (3753-3969) of MLL has been synthesized (Genescript) and cloned in fusion with a TEV cleavable glutathione sulfotransferase. GST-MLL1 has been overexpressed similarly to Ash2L. Cells were harvested in PBS buffer, lysed by sonication and centrifuged. The supernatant was applied to glutathione sepharose 4B for 1 hour and unbound proteins were washed with

50-column volume of PBS. The slurry was resuspended in 25mM Tris pH 8.0, 100mM NaCl, 5mM β-mercaptethanol and 250mg of TEV protease during 12 hours. Cleaved MLL1 was further purified by size exclusion chromatography (Superdex 75) pre-equilibrated with the slurry buffer.

V - In vitro methyltransferase assay - Methyltransferase assays were conducted using 5mM of recombinantly purified MLL SET domain with equimolar amounts of Ash2L, RbBP5 and WDR5. Reactions were initiated by the addition or 1mCi of radiolabeled S-Adenosyl-L-methionine and incubated during 2 hours at 22ºC. Methyltransferase assays were carried out in 50mM Tris pH 8.5, 200mM NaCl, 3mM DTT, 5mM $MgCl_2$ and 5% glycerol and stopped by spotting the reactions onto Whatman P-81 filter papers. Free AdoMet was removed by washing the filter papers in 250mL of 50mM $NaHCO_3$ at pH 9.0. Activity was quantified by liquid scintillation counts. Inhibition analyses were performed similarly with the exception that the peptide inhibitor was added to the assay prior adding to AdoMet and the substrate peptide.

2.3.6. ACKNOWLEDGMENTS

We wish to thank Dr. Alexandre Blais for reviewing this manuscript. This work was supported by a Canadian Institutes of Health Research grant to J.-F. C. J.-F. C. holds a Canadian Research Chair in Structural Biology and Epigenetics.

CHAPTER 3 – DISCUSSION

In this study, we have employed various biochemical, biophysical and structural methods to study the stimulatory role of the MLL1 core complex subunits in promoting MLL1 methyltransferase activity. Each member of the MLL core complex, namely Ash2L, RbBP5 and WDR5, play a critical role in regulating MLL methyltransferase activity and activating the expression of several developmental genes. Our work suggests that specific functional domains of each core complex members play a key role in regulating MLL1 complex formation, catalysis and gene targeting. Furthermore by analyzing WDR5's peptidyl arginine binding cleft, we have developed a peptidomimetic inhibitor of MLL1 methyltransferase activity that will allow us to further study MLL kinetics as well as its role *in vivo*.

3.1 - Identification of the minimal structural determinants underlying the stimulation of MLL1 methyltransferase activity

WDR5 plays an important role in mediating the formation of a fully competent MLL1 core complex through its interaction with MLL1 (Patel et al, 2008b) and RbBP5 (Dou et al, 2006) while the Ash2L-RbBP5 heterodimer is pivotal in the stimulation of MLL1 di/tri-methyltransferase activity (Dou et al, 2006; Steward et al, 2006). However, the details controlling the formation of the MLL1 core complex are unknown. Additionally, given that each core complex member is composed of multiple functional domains (Fig 11a and b), it is unknown which one participates in complex formation and regulation of MLL1 methyltransferase activity. We therefore undertook a biochemical

characterization of the individual components of the core complex to map out key interactions required for the stimulation of MLL1 methyltransferase activity.

We first verified the interaction between RbBP5-WDR5 and through a GST pull-down approach we identified that the C-terminal domain of RbBP5 is required for its association with WDR5 (Fig. 11c). Using a combination of sequence alignment, GST-pull down experiments and isothermal titration calorimetry, we have defined that only a small portion of RbBP5 C-terminal domain, more precisely residues 370-380, encompassed the key residues underlying the formation of the WDR5-RbBP5 complex (Fig 11d).

After establishing the minimal structural determinants controlling the formation of the WDR5-RbBP5 complex, we determined the region of Ash2L required for its binding to RbBP5. Similar to the mapping of WDR5-RbBP5 interaction, we employed a GST-pull down approach and established that only RbBP5 residues 330-370 (also referred as the Hinge region) conferred binding to Ash2L (Fig. 11a). Given that Ash2L is composed of a N-terminal C4 zinc finger domain and a C-terminal SPRY domain, we then sought to map where the RbBP5 hinge region bound Ash2L. Through further domain mapping using the above-mentioned experimental approaches, we identified the Ash2L SPRY domain as the key RbBP5 binding platform (Fig. 11b).

After examining the structural determinants underlying the formation of the complex, we sought to determine whether RbBP5 330-380 region (referred therein as $RbBP5_{330-380}$), Ash2L SPRY domain (referred therein as $Ash2L_C$) and WDR5 could stimulate MLL1 methyltransferase activity. Surprisingly, using a radiometric filter paper assay, we determined that a MLL1 complex consisting of WDR5-$RbBP5_{330-380}$-$Ash2L_C$ is

capable of stimulating MLL1 methyltransferase activity as potently as the complex reconstituted with the full-length subunits (Fig. 11e). This suggests that the scaffolding elements of the WDR5-RbBP5-Ash2L complex are also the structural determinants stimulating MLL1 methyltransferase activity.

3.1.1 How is MLL1 methyltransferase activity regulated by WDR5-RbBP5-Ash2L?

What remains to be done now is a) determining the binding modes utilized by the three proteins to form a functional methyltransferase complex b) finding how the interaction between these three proteins allows for full MLL1 methyltransferase activity and c) elucidating what the roles are of the remaining domains of each core complex member? In order to determine the binding modes, a crystal structure of the minimal core complex would offer unlimited information. From this we would be able to ascertain the type of binding occurring and the specific residues required for the interaction from which we can employ mutagenesis studies to identify residues key in complex formation, catalysis or product specificity. Furthermore a comparison of the MLL1 SET domain solved in the absence of the core complex with the one bound by the core complex would give insights into the conformational changes required for the stimulation of MLL1 methyltransferase activity. As the MLL1 SET domain cannot catalyze H3K4 di/tri methylation, it is likely that its interactions with the core complex allow for a structural re-arrangement of the SET domain or provide the necessary chemical environment for the methyltransfer to occur (Southall et al, 2009). This model is further supported by a recent study on COMPASS, the yeast homolog of the MLL1 core complex (Takahashi et al, 2009; Takahashi and Shilatifard 2010). In this complex,

SET1, the yeast homologue of MLL1, harbors a highly conserved active site tyrosine residue controlling the product specificity of SET1. Within the catalytic domain, Tyr1052 prevents the formation of a di and trimethylated product. In the same way that Ash2L, RbBP5 and WDR5 stimulate MLL1 activity, their homologues in yeast, namely Cps60, Cps50 and Cps35 respectively, also allow SET1 to catalyze an efficient methylation reaction (Miller et al, 2001). Furthermore, when in complex, Cps60 interacts with SET1 Tyr1052 and relieves the steric hindrance of Tyr1052 hydroxyl group and enables di and tri-methylation of histone H3 on Lys-4 (Takahashi et al, 2009). Future structural studies will be instrumental in confirming this model and providing insights into MLL1 mediated histone H3K4 methylation.

To date, we have been able to assign structural functions to the WDR5 β-propeller, the RbBP5 Hinge and C-terminus and finally the Ash2L SPRY domain (Fig. 20). There now remains the task of assigning functions to the remaining functional domains of the core complex, namely the RbBP5 β-propeller and C-terminus, and the Ash2L N-terminus. Future biochemical, kinetic and structural studies of the individual core complex members and in complex is required to answer the question of how the core complex promotes MLL methyltransferase activity and to determine how they control product specificity.

Section 3.2. An atypical helix-wing-helix domain mediates Ash2L binding to the *beta-globin* gene.

The MLL1 core complex and its methyltransferase activity are key in activating the expression of several genes linked to developmental processes (Ansari and Mandal

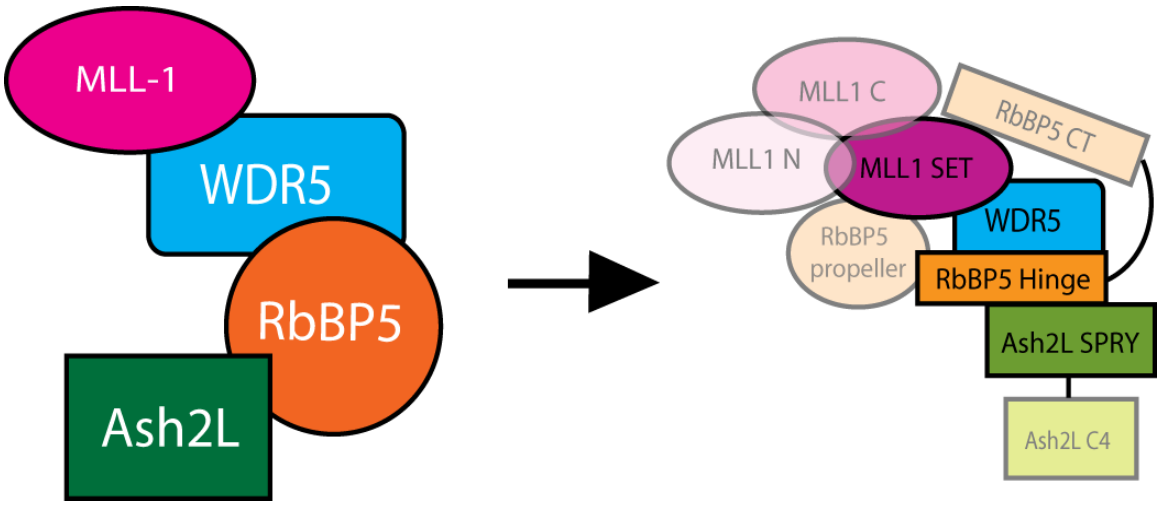


Fig. 20 – Past and current understanding of the MLL1 core complex

It was previously known that MLL1 interacts directly with WDR5, and the remaining interactions within the core complex were between WDR5-RbBP5 and Ash2L-RbBP5. After our structural and biochemical dissection we have identified a minimal core complex, consisting of the MLL1 WIN motif, the WDR5 β-propeller, residues 330-380 of the RbBP5 Hinge and C-terminus and the Ash2L SPRY domain.

2010;Eissenberg and Shilatifard 2010). Our dissection of the core complex has shown that Ash2L SPRY domain interacts with RbBP5 and participates in the formation of the MLL1 core complex. We have also noted that Ash2L N-terminal domain ($Ash2L_N$) does not bind any subunit of the core complex and has no effect on the methyltransferase activity of MLL1.

In order to better understand the role of the Ash2L N-terminal domain, we have solved its crystal structure (Fig. 15a and Table 3). Using single-wavelength anomalous diffraction experiments, we identified that Ash2L N-terminal domain harbors a C4 zinc finger motif followed by an atypical helix-wing-helix (HWH) domain (Fig. 21a). Based on previous studies showing that HWH domains are key in binding DNA (Obsil and Obsilova 2008), we hypothesized that $Ash2L_N$ played an important role in Ash2L recruitment to key target genes. To identify a putative DNA binding site, we used Systematic Evolution of Ligands by Exponential Enrichment (SELEX) and electrophoretic mobility shift assays (EMSA). Using this combinatorial approach, we identified a highly enriched motif: AGGCT (Fig. 14). Genome wide localization of the SELEX motif revealed a significant enrichment of the AGGCT sequence in the promoter region of genes involved in development. To explore the biochemical determinants controlling the binding of DNA, we performed EMSA with a sequence corresponding to the hypersensitive site 2 (HS2) of β-globin' locus control region: a region occupied by the MLL2 complex during murine erythroleukemia (MEL) cell differentiation (Demers et al, 2007) and encompassing the AGGCT motif. Consistent with our findings, we observed a notable shift when binding reactions were performed with full length Ash2L ($Ash2L_{FL}$) and $Ash2L_N$ while no binding could be detected with $Ash2L_C$ (Fig. 13b). To

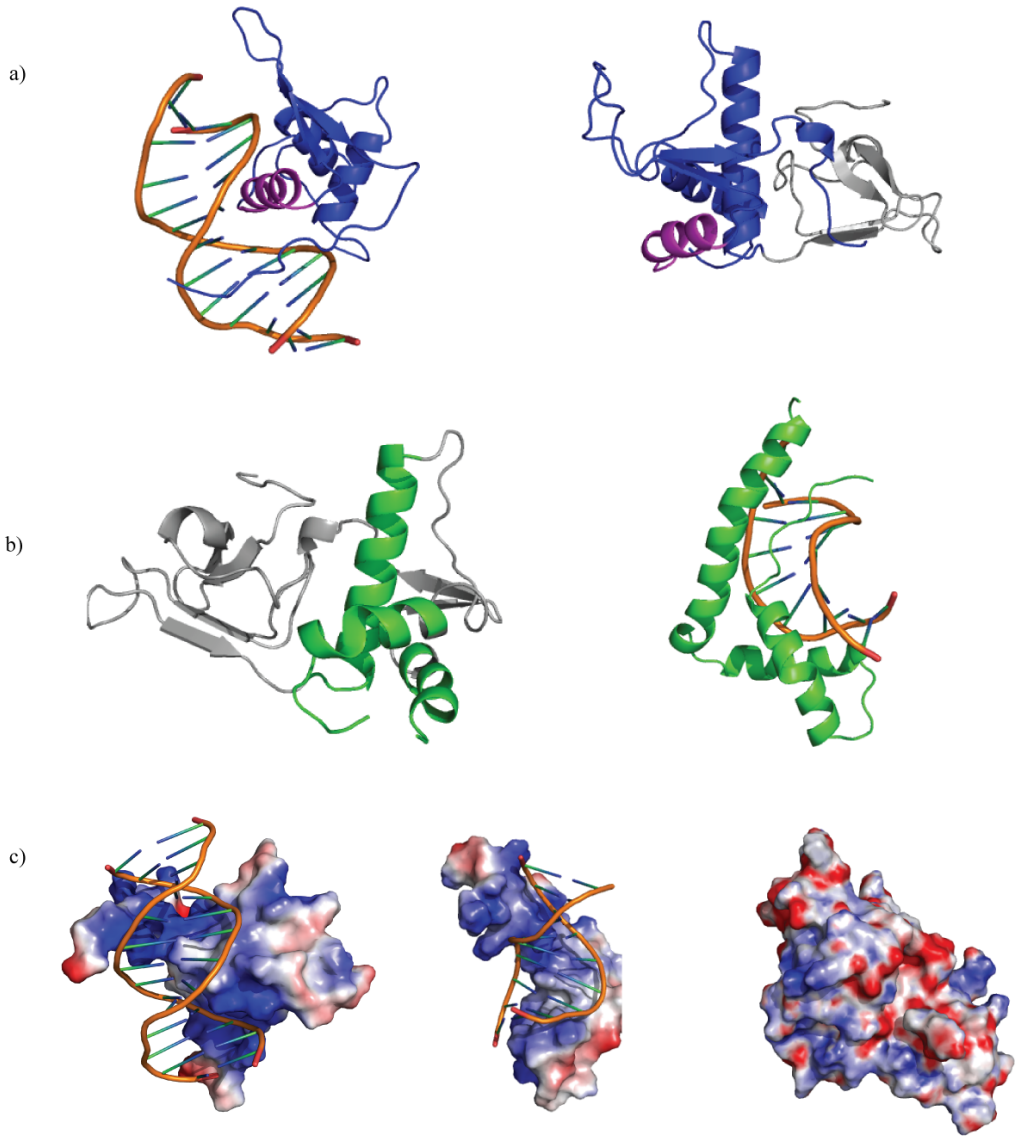


Fig. 21 – Structural comparison of Ash2L to FOXO and HMG DNA binding

a)Comparison of the Ash2L HWH (left) and FOXOA3 HWH (right) domains (blue and purple). Ash2L HWH contains a highly conserved alpha helix hypothesized to be involved in DNA binding that is similar in structure and sequence to helix 3 (magenta) of FOXOA3 that makes key contacts with DNA.

b) HMG proteins have a characteristic "L" shaped (green) structure required for DNA binding and a similar arrangement of alpha helices is seen in the structure of Ash2LN.

c) Both the FOXO and HMG proteins utilize highly positive surfaces (blue) for DNA binding, a feature missing from the surface of $Ash2L_N$.

probe for potential DNA binding residues, we generated a model of Ash2L in complex with DNA. This model was generated based on the interaction between FOXO3A HWH domain and their target DNA (Tsai et al, 2007). From this model, we tested the DNA binding activity of a library of $Ash2L_N$ mutants using EMSA. As shown in Figure 18b, substitution of Lys-225 and Lys-229 to glutamic acid residues completely abrogated DNA binding (Fig. 18b), providing some insights into the mechanisms underlying the binding of HS2 by Ash2L. However, despite these similarities, our mutational analysis revealed that other key DNA binding residues in FOXO proteins found in Ash2L (such as Asn-221) did not participate in binding DNA. These results were surprising given the fact that this asparagine residue, in FOXO proteins, makes several hydrogen bonds with DNA (Fig. 21a) (Weigelt et al, 2001). Since Ash2L also contains a structurally conserved asparagine residue and because the Ash2L-DNA model (Fig. 16a) was generated based on structures of FOXO-DNA complexes, we postulated that it would also be necessary for its DNA binding. However, since our gel shift assay showed otherwise, we hypothesized that the DNA binding mode of Ash2L HWH may differ significantly from that of the FOXO's. Further structural comparison between the FOXO proteins and the Ash2L HWH revealed that key structural features required for DNA binding by the FOXO proteins are not present in $Ash2L_N$. Indeed, Ash2L lacks the hydrophobic residues or extended positively charged patch of residues encompassing the DNA binding α-helix (Brent et al, 2008;Obsil and Obsilova 2008)**,** clearly indicating that Ash2L DNA binding mode differs from other HWH domain containing protein (Fig. 22c).

An additional factor to consider when speculating on the DNA binding mode of $Ash2L_N$ is the DNA structure itself. Based on our findings that HS2 adopts two

electrophoretically migrating DNA species (Fig. 14e), we surmised that the HS2 species with the higher apparent molecular weight (that we named $HS2_{stru}$) would harbor a bent structure resembling to a cruciform DNA (Elborough & West, 1988) and that $Ash2L_N$ recognizes such DNA structure. To confirm this idea, we performed EMSA with HMGB2, a canonical cruciform DNA (Stefanovsky and Moss 2010). Consistent with our hypothesis, we observed that $Ash2L_N$ bound to HMGB2, with similar affinity than $HS2_{stru}$ (Fig. 13c). Although B-form DNA is the typical, most abundant and energetically favorable structure adopted by DNA, higher order DNA structures such as triplex, G-quartets (Lipps & Rhodes, 2009) and cruciforms (Pearson et al, 1996) have been widely reported. The significance of these structures and their biological roles *in vivo* are not well understood but cruciform DNA structures are often found at origins of replication and are postulated to regulate transcription initiation (Horwitz, 1989; Waga et al, 1990). After identifying that Ash2L binds cruciform DNA, we sought to compare the structure of Ash2L HWH domain with canonical cruciform DNA binding domains, namely the high-mobility group

3.2.1 Ash2L DNA binding domain : a hydrid between HWH and HMG?

Interestingly, the negatively charged β5-β6 wing of Ash2L HWH could serve an equivalent role. Despite these similarities, additional structural studies on the Ash2L:DNA complex will be required to fully understand Ash2L DNA binding mode. If a protein-DNA complex proves to be unfeasible, several groups have reported the structural characterization of higher order structured DNA through the use of NMR (Ambrus & Yang, 2007) or X-ray crystallography (Parkinson et al, 2002). Knowing the

structure of the $HS2_{Stru}$, coupled with further mutagenesis studies of Ash2L would permit a thorough understanding of HS2 recognition by Ash2L.

3.2.2 Links between DNA structure and β-globin expression

Several mechanisms exist at the β-globin gene locus control region to ensure the proper regulation of the globin genes (Sankaran et al, 2010). Recent studies have shown that chromatin remodeling and histone post-transcriptional modifications, including MLL mediated H3K4 tri-methylation, are important in regulating β-globin gene expression (Demers et al, 2007; Hosey et al, 2010; Kiefer et al, 2008). Additionally, other studies have found that NF-E2 recruitment to DNA (Armstrong & Emerson, 1996) triggers the eviction of nucleosomes and several transcriptional start sites for specific globin genes are depleted of nucleosomes (Kim et al, 2007). Our findings that one of the subunits associating with MLL2, Ash2L, binds bent DNA may provide a unifying mechanism between these two models, given the interaction between NF-E2 and Ash2L at the β-globin locus. Indeed, it is tempting to speculate that upon binding of NF-E2 to HS2, nucleosomes are evicted and triggers the formation of DNA secondary structures. Such secondary structures are then recognized by Ash2L and leads to the recruitment of MLL2. Recruitment of MLL2 triggers the methylation of histone H3 on Lys-4, on neighboring nucleosomes, which is recognized by ATP-dependent chromatin remodeling complexes harboring effector domains which recognize H3K4 methyl mark (Taverna et al, 2006). Such combination of events then leads to the maximal expression of the β-globin gene.

Section 3.3 - Fine-tuning the stimulation of MLL1 methyltransferase activity by a histone H3 based peptide mimetic

The interaction between MLL1 and WDR5 is key in the stimulation of MLL1 methyltransferase activity by the core complex (Dou et al, 2006; Patel et al, 2008b); hence preventing that interaction presents itself as a possible inhibitor of MLL1 methyltransferease activity (Karatas et al, 2010). Many SET domain enzymes, including MLL1, have been linked to various types of cancer (Dou et al, 2006). Consistent with these linkages, several companies have taken on the endeavor of developing therapeutic molecules targeting SET domain enzymes (Cole, 2008). However, currently, only few have been characterized. Those include 3-deazaneplanocin A (Fiskus et al, 2009) and BIX-01294 (Kubicek et al, 2007) and both have demonstrated the ability to effect *in vivo* levels of lysine methylation. Potential problems with the aforementioned inhibitors are that 3-deazaneplanocin acts as a S-adenosylhomocysteine hydrolase inhibitor and hence acts as broad range inhibitor of any AdoMet dependant methyltransferase (Miranda et al, 2009). BIX-01294 is specific for the H3K9 methyltransferase G9a, but showed mild effect on H3K9 di-methylated levels *in vivo* (Kubicek et al, 2007). These early studies demonstrated that blocking the AdoMet binding site might not present a viable avenue for developing drugs due to cross-inhibition of the other SET domain methyltransferases. To palliate some of these problems, we decided to inhibit MLL1 activity through the targeting of key protein-protein interactions (Eichler, 2008; White et al, 2008) rather than the SET domain itself. We therefore set out to develop a peptidomimetic inhibitor of MLL1 methyltransferase activity that would act by preventing the interaction of MLL1 with the core complex subunits.

Based on the sequences of histone H3 and the MLL1 WIN peptide, we found that an N-alpha acetylated histone H3 peptide (NαH3) binds WDR5 with higher affinity than the unacetylated form of the peptides. Correlatively, NαH3 is capable of inhibiting MLL1 activity *in vitro* with an IC50 value of 4.06μM (Fig. 17b). Consistent with these findings, titration studies of WDR5 with NαH3 revealed that the peptide mimetic binds with a dissociation constant of 0.13μM (Fig. 17b), an 11-fold increase in affinity over MLL WIN peptide (Patel et al, 2008b). Our crystal structure revealed the structural determinants for the high-affinity binding of NαH3 to WDR5. First, three distinct binding pockets including the peptidyl arginine binding cleft (Fig. 18 and 19), the Nα-acetyl and A1 pockets maintain NαH3 within WDR5 central toroid peptide binding cavity (Fig. 19). Second, the NαH3 acetyl moiety engages in two intra-molecular hydrogen bonds with T3 side chain and amide group. Third, the formation of a new amide bond following the Nalpha acetylation of the peptide likely reinforces the hydrogen bonds between A1 amide group and Asp-107 and Ser-91 carboxyl groups. Altogether, these interactions likely forces the peptide to adopt a more compact conformation, in comparison to MLL1 WIN (Patel et al, 2008a;Song and Kingston 2008) or an unmodified histone H3 peptides (Couture et al, 2006), within WDR5 peptidyl arginine binding.

3.3.1 NαH3 mode of action

We hypothesized that the inhibitory effect of the peptide is achieved through blocking of the MLL1-WDR5 interaction and hence disrupting the stimulatory effect of the core complex on MLL1 activity (Dou et al, 2006;Patel et al, 2009). Although we

have not demonstrated that our peptide is capable of disrupting the formation of the core complex, the MLL WIN peptide has been shown to abrogate core complex formation (Patel et al, 2008b). We have demonstrated the *in vitro* ability of the NαH3 peptide to inhibit MLL1 methyltransferase activity; we have not yet shown that it is capable of the same effect *in vivo.* In addition, an 11-fold increase in affinity and an IC50 value of 4.06μM may not be sufficient to produce a significant effect on H3K4 di and tri-methylation *in vivo* but our structure and mutational analysis provide an excellent starting point for further engineering and optimization of the peptide in order to produce a potent inhibitor. Increasing the affinity of the peptide for WDR5 is necessary since even at high concentrations we were not able to completely inhibit MLL1 methyltransferase activity. This residual activity may be due to MLL1 mono-methyltransferase activity (Patel et al, 2009), which does not require the presence of the core complex subunits, but without verifying the level of H3K4 methylation it is difficult to accredit the residual activity to the MLL1 protein itself.

In humans, six members of the SET1 family catalyze the di/tri-methylation of histone H3 on Lys-4 (Cosgrove & Patel, 2010). Given that the majority of the members of the SET1 family, except for MLL5, require the core complex for full activity, it is tempting to speculate that NαH3 will also block the stimulation of their enzymatic activity. However, a recent study analyzing the ability of WDR5 to bind the conserved WIN motifs of MLL1-4 demonstrated it was only capable of binding MLL1 and 4 (Song and Kingston 2008). Although this was tested through GST-pulldown assays and not verified with a quantitative method, this presents a potential roadblock in utilizing our peptide to decrease global H3K4 di and tri-methylation levels. Additionally, MLL5

catalytic activity is independent of the WDR5-RbBP5-Ash2L complex while (Fujiki et al, 2009; Wu et al, 2008) the SET1A/B mediated H3K4 methylation requires the presence of WDR82 (Wu et al, 2008). These few examples clearly indicate that NαH3 will likely prevent the association of only a subset of members of the SET1 family of methyltransferases with WDR5. On the other hand, this limitation may be advantageous given the fact that this peptide mimetic is designed to prevent cancer-related H3K4 methyl marks. In the event that only specific members of the family are inhibited by NαH3, only a subset of WDR5-MLL complex will be amenable for the NαH3-mediated blocking of the stimulation of their methyltransferase activity. This idea is further reemphasized by recent findings that each member of the SET1 family of methyltransferases are not be functionally redundant (Ansari and Mandal 2010; Eissenberg and Shilatifard 2010). If our peptide is only specific for MLL1, NαH3 may provides a unique tool for modulating MLL1 activity *in vivo* and gain additional insights into the biological roles MLL1 methyltransferase activity.

As it stands, the peptide has several characteristics that would make it a poor drug. These include high molecular weight and presence of polar groups capable of hydrogen bonding (Lipinski et al, 2001). Several peptide mimetics have been developed for disease treatment (Chen & Huerta, 2009; Janosz et al, 2009). Some of these were based on the sequence and structure of a peptide but after several rounds of refinement resembled more a small organic molecule rather than a peptide (Kikelj, 2003). In order to achieve a similar goal with the NαH3 peptide, we will have to carefully examine which chemical moieties of the peptide are essential in binding and remove any extraneous peptide components to produce a smaller and more hydrophobic drug. These

improvements will also aid in increasing cellular uptake of the drug. Ultimately, we believe that, once optimized, the NαH3 peptide will be a valuable biological tool, whether as a cancer therapeutic or as a method to study MLL mediated developmental and differentiation programs such as hematopoesis (Argiropoulos & Humphries, 2007), myogenesis (McKinnell et al, 2008) and neurogenesis (Lim et al, 2009).

3.4 CONCLUSIONS

Understanding how the core complex regulates the methyltransferase activity of MLL1 has been a difficult undertaking for the chromatin field. These difficulties stem from disparity between MLL's predicted behavior and its *in vivo* activity. Furthermore, the mechanism through which the core complex stimulates MLL1 methyltransferase activity has been difficult due to the fact that each subunit is composed of many functional domains. We identified the minimal functional domains of the core complex required for methyltransferase activity. This work has provided a basis for further understanding MLL catalysis and product specificity. By solving the structure of the Ash2L N-terminal domain, we identified novel DNA binding domain of Ash2L and provided new insights on the mechanisms controlling the recruitment of MLL to its target genes. Through analysis of WDR5's interaction with both histone H3 and the MLL1 WIN, we identified a peptidomimetic inhibitor of MLL1 methyltransferase activity. Although this inhibitor requires more development and optimization, it presents itself as a new tool in treating MLL mediated leukemias and in studying MLL dependent development and differentiation processes. The kinetic and catalytic mechanisms of several histone-modifying complexes have remained poorly understood. Given that

Ash2L, RbBP5 and WDR5 are common subunits of all SET1 family of histone methyltransferases, this work can be easily applied to other complexes and will provide a framework to help understanding the biochemical, structural and kinetic mechanisms controlling other multi-subunit histone methyltransferases.

CHAPTER 4. REFERENCES

Abrahams J, Leslie A (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallographica Section D: Biological Crystallography* **52:** 30-42

Allan J, Mitchell T, Harborne N, Bohm L, Crane-Robinson C (1986) Roles of H1 domains in determining higher order chromatin structure and H1 location. *J Mol Biol* **187:** 591-601

Allfrey VG, faulkner R, mirsky AE (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci USA* **51:** 786-794

Ambrus A, Yang D (2007) Diffusion-ordered nuclear magnetic resonance spectroscopy for analysis of DNA secondary structural elements. *Anal Biochem* **367:** 56-67

Ansari KI, Mandal SS (2010) Mixed lineage leukemia: roles in gene expression, hormone signaling and mRNA processing. *FEBS J* **277:** 1790-1804

Argiropoulos B, Humphries RK (2007) Hox genes in hematopoiesis and leukemogenesis. *Oncogene* **26:** 6766-6776

Armstrong JA, Emerson BM (1996) NF-E2 disrupts chromatin structure at human beta-globin locus control region hypersensitive site 2 in vitro. *Mol Cell Biol* **16:** 5634-5644

Arya G, Maitra A, Grigoryev SA (2010) A structural perspective on the where, how, why, and what of nucleosome positioning. *J Biomol Struct Dyn* **27:** 803-820

Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* **129:** 823-837

Basecke J, Whelan JT, Griesinger F, Bertrand FE (2006) The MLL partial tandem duplication in acute myeloid leukaemia. *Br J Haematol* **135:** 438-449

Bauer WR, Hayes JJ, White JH, Wolffe AP (1994) Nucleosome structural changes due to acetylation. *J Mol Biol* **236:** 685-690

Belmont AS (2006) Mitotic chromosome structure and condensation. *Curr Opin Cell Biol* **18:** 632-638

Beltran S, Blanco E, Serras F, Pérez-Villamil B, Guigó R, Artavanis-Tsakonas S, Corominas M (2003) Transcriptional network controlled by the trithorax-group gene ash2 in Drosophila melanogaster. *Proc Natl Acad Sci USA* **100:** 3293-3298

Bergink S, Salomons FA, Hoogstraten D, Groothuis TAM, de Waard H, Wu J, Yuan L, Citterio E, Houtsmuller AB, Neefjes J, Hoeijmakers JHJ, Vermeulen W, Dantuma NP (2006) DNA damage triggers nucleotide excision repair-dependent monoubiquitylation of histone H2A. *Genes Dev* **20:** 1343-1352

Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, Kouzarides T, Schreiber SL (2002) Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc Natl Acad Sci USA* **99:** 8695-8700

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ, 3rd, Gingeras TR, Schreiber SL, Lander ES (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120:** 169-181

Bernstein BE, Liu CL, Humphrey EL, Perlstein EO, Schreiber SL (2004) Global nucleosome occupancy in yeast. *Genome Biol* **5:** R62

Birke M, Schreiner S, García-Cuéllar M-P, Mahr K, Titgemeyer F, Slany RK (2002) The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Res* **30:** 958-965

Blais A, van Oevelen C, Margueron R, Acosta-Alvear D, Dynlacht B (2007) Retinoblastoma tumor suppressor protein–dependent methylation of histone H3 lysine 27 is associated with irreversible cell cycle exit. *The Journal of cell biology* **179:** 1399

Boeger H, Griesenbeck J, Strattan JS, Kornberg RD (2003) Nucleosomes unfold completely at a transcriptionally active promoter. *Molecular cell* **11:** 1587-1598

Brent MM, Anand R, Marmorstein R (2008) Structural basis for DNA recognition by FoxO1 and its regulation by posttranslational modification. *Structure* **16:** 1407-1416

Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, Allis CD (1996) Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84:** 843-851

Campos EI, Reinberg D (2009) Histones: Annotating Chromatin. *Annu Rev Genet* **43:** 559-599

Carter D, Chakalova L, Osborne C, Dai Y, Fraser P (2002) Long-range chromatin regulatory interactions in vivo. *Nat Genet* **32:** 623-626

Chang P-Y, Hom RA, Musselman CA, Zhu L, Kuo A, Gozani O, Kutateladze TG, Cleary ML (2010) Binding of the MLL PHD3 Finger to Histone H3K4me3 Is Required for MLL-Dependent Gene Transcription. *J Mol Biol* **400:** 137-144

Chen DJ, Huerta S (2009) Smac mimetics as new cancer therapeutics. *Anticancer Drugs* **20:** 646-658

Chen V, Arendall W, Headd J, Keedy D, Immormino R, Kapral G, Murray L, Richardson J, Richardson D (2009) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **66:** 12-21

Cheng X, Collins RE, Zhang X (2005) Structural and sequence motifs of protein (histone) methylation enzymes. *Annu Rev Biophys Biomol Struct* **34:** 267-294

Chonghaile TN, Letai A (2008) Mimicking the BH3 domain to kill cancer cells. *Oncogene* **27 Suppl 1:** S149-157

Cierpicki T, Risner LE, Grembecka J, Lukasik SM, Popovic R, Omonkowska M, Shultis DD, Zeleznik-Le NJ, Bushweller JH (2010) Structure of the MLL CXXC domain–DNA complex and its functional role in MLL-AF9 leukemia. *Nat Struct Mol Biol* **17:** 62-68

Cole PA (2008) Chemical probes for histone-modifying enzymes. *Nat Chem Biol* **4:** 590-597

Corona DFV, Siriaco G, Armstrong JA, Snarskaya N, McClymont SA, Scott MP, Tamkun JW (2007) ISWI regulates higher-order chromatin structure and histone H1 assembly in vivo. *PLoS Biol* **5:** e232

Cosgrove MS, Patel A (2010) Mixed lineage leukemia: a structure-function perspective of the MLL1 protein. *FEBS J* **277:** 1832-1842

Couture J, Collazo E, Trievel R (2006a) Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat Struct Mol Biol* **13:** 698-703

Couture J-F, Collazo E, Brunzelle JS, Trievel RC (2005) Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes Dev* **19:** 1455-1465

Couture J-F, Dirk LMA, Brunzelle JS, Houtz RL, Trievel RC (2008) Structural origins for the product specificity of SET domain protein methyltransferases. *Proc Natl Acad Sci USA* **105:** 20659-20664

Couture J-F, Trievel RC (2006) Histone-modifying enzymes: encrypting an enigmatic epigenetic code. *Curr Opin Struct Biol* **16:** 753-760

Couture JF, Collazo E, Ortiz-Tello PA, Brunzelle JS, Trievel RC (2007) Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat Struct Mol Biol* **14:** 689-695

Couture JF, Collazo E, Trievel RC (2006b) Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat Struct Mol Biol* **13:** 698-703

Cowtan K, Main P (1993) Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. *Acta Crystallographica Section D: Biological Crystallography* **49:** 148-157

Cragg MS, Harris C, Strasser A, Scott CL (2009) Unleashing the power of inhibitors of oncogenic kinases through BH3 mimetics. *Nat Rev Cancer* **9:** 321-326

Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J Mol Biol* **319:** 1097-1113

de La Fortelle E, Bricogne G (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods in enzymology* **276:** 472-494

Dehe PM, Dichtl B, Schaft D, Roguev A, Pamblanco M, Lebrun R, Rodriguez-Gil A, Mkandawire M, Landsberg K, Shevchenko A, Rosaleny LE, Tordera V, Chavez S, Stewart AF, 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:** 35404-35412

Del Rizzo PA, Couture J-F, Dirk LMA, Strunk BS, Roiko MS, Brunzelle JS, Houtz RL, Trievel RC (2010) SET7/9 catalytic mutants reveal the role of active site water molecules in lysine multiple methylation. *J Biol Chem*

Demers C, Chaturvedi C-P, 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

Dennis Jr G, Sherman B, Hosack D, Yang J, Gao W, Lane H, Lempicki R (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* **4:** P3

Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, Zhou MM (1999) Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399:** 491-496

Dillon SC, Zhang X, Trievel RC, Cheng X (2005) The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol* **6:** 227

Dorrance AM, Liu S, Yuan W, Becknell B, Arnoczky KJ, Guimond M, Strout MP, Feng L, Nakamura T, Yu L, Rush LJ, Weinstein M, Leone G, Wu L, Ferketich A, Whitman SP, Marcucci G, Caligiuri MA (2006) Mll partial tandem duplication induces aberrant Hox expression in vivo via specific epigenetic alterations. *J Clin Invest* **116:** 2707-2716

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

Dou Y, Milne TA, Tackett AJ, Smith ER, Fukuda A, Wysocka J, Allis CD, Chait BT, Hess JL, Roeder RG (2005) Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **121:** 873-885

Eichler J (2008) Peptides as protein binding site mimetics. *Curr Opin Chem Biol* **12:** 707-713

Eissenberg JC, Shilatifard A (2010) Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Dev Biol* **339:** 240-249

Elborough KM, West SC (1988) Specific binding of cruciform DNA structures by a protein from human extracts. *Nucleic Acids Res* **16:** 3603-3616

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60:** 2126-2132

Ercan S, Simpson RT (2004) Global chromatin structure of 45,000 base pairs of chromosome III in a- and alpha-cell yeast and during mating-type switching. *Mol Cell Biol* **24:** 10026-10035

Ferraiuolo MA, Rousseau M, Miyamoto C, Shenker S, Wang XQD, Nadler M, Blanchette M, Dostie J (2010) The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res***:** 1-13

Filippakopoulos P, Low A, Sharpe TD, Uppenberg J, Yao S, Kuang Z, Savitsky P, Lewis RS, Nicholson SE, Norton RS, Bullock AN (2010) Structural Basis for Par-4 Recognition by the SPRY Domain- and SOCS Box-Containing Proteins SPSB1, SPSB2, and SPSB4. *J Mol Biol*

Fiskus W, Wang Y, Sreekumar A, Buckley KM, Shi H, Jillella A, Ustun C, Rao R, Fernandez P, Chen J, Balusu R, Koul S, Atadja P, Marquez VE, Bhalla KN (2009) Combined epigenetic therapy with the histone methyltransferase EZH2 inhibitor 3-deazaneplanocin A and the histone deacetylase inhibitor panobinostat against human AML cells. *Blood***:** 1-11

Flanagan JF, Mi L-Z, Chruszcz M, Cymborowski M, Clines KL, Kim Y, Minor W, Rastinejad F, Khorasanizadeh S (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438:** 1181-1185

Fujiki R, Chikanishi T, Hashiba W, Ito H, Takada I, Roeder RG, Kitagawa H, Kato S (2009) GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis. *Nature* **459:** 455-459

Gangaraju VK, Bartholomew B (2007) Mechanisms of ATP dependent chromatin remodeling. *Mutat Res* **618:** 3-17

Gao J, Bosco DA, Powers ET, Kelly JW (2009) Localized thermodynamic coupling between hydrogen bonding and microenvironment polarity substantially stabilizes proteins. *Nat Struct Mol Biol* **16:** 684-690

Gariglio M, Ying GG, Hertel L, Gaboli M, Clerc RG, Landolfo S (1997) The high-mobility group protein T160 binds to both linear and cruciform DNA and mediates DNA bending as determined by ring closure. *Exp Cell Res* **236:** 472-481

Goto NK, Zor T, Martinez-Yamout M, Dyson HJ, Wright PE (2002) Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. *J Biol Chem* **277:** 43168-43174

Grant PA (2001) A tale of histone modifications. *Genome Biol* **2:** REVIEWS0003

Greer EL, Maures TJ, Hauswirth AG, Green EM, Leeman DS, Maro GS, Han S, Banko MR, Gozani O, Brunet A (2010) Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in C. elegans. *Nature* **466:** 383-387

Greeson N, Sengupta R, Arida A, Jenuwein T, Sanders S (2008) Di-methyl H4 lysine 20 targets the checkpoint protein CRB2 to sites of DNA damage. *Journal of Biological Chemistry* **283:** 33168

Halbach A, Zhang H, Wengi A, Jablonska Z, Gruber IM, Halbeisen RE, Dehe PM, Kemmeren P, Holstege F, Geli V, Gerber AP, Dichtl B (2009) Cotranslational assembly of the yeast SET1C histone methyltransferase complex. *EMBO J* **28:** 2959-2970

Han Z, Guo L, Wang H, Shen Y, Deng XW, Chai J (2006a) Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. *Molecular cell* **22:** 137-144

Han Z, Guo L, Wang H, Shen Y, Deng XW, Chai J (2006b) Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. *Mol Cell* **22:** 137-144

Hess J (2004) MLL: a histone methyltransferase disrupted in leukemia. *Trends in molecular medicine* **10:** 500-507

Hollmann M, Simmerl E, Schäfer U, Schäfer MA (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

Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24:** 2780-2781

Honda BM, Dixon GH, Candido EP (1975) Sites of in vivo histone methylation in developing trout testis. *J Biol Chem* **250:** 8681-8685

Horwitz MS (1989) Transcription regulation in vitro by an E. coli promoter containing a DNA cruciform in the '-35' region. *Nucleic Acids Res* **17:** 5537-5545

Hosey AM, Chaturvedi C-P, Brand M (2010) Crosstalk between histone modifications maintains the developmental pattern of gene expression on a tissue-specific locus. *Epigenetics* **5:** 273-281

Hsieh JJ-D, Ernst P, Erdjument-Bromage H, Tempst P, Korsmeyer SJ (2003) Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. *Mol Cell Biol* **23:** 186-194

Hsu T (1962) Differential rate in RNA synthesis between euchromatin and heterochromatin. *Exp Cell Res* **27:** 332

Huang Y, Fang J, Bedford MT, Zhang Y, Xu R-M (2006) Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312:** 748-751

Janosz KEN, Zalesin KC, Miller WM, McCullough PA (2009) Treating type 2 diabetes: incretin mimetics and enhancers. *Therapeutic advances in cardiovascular disease* **3:** 387-395

Jonathan C, Shilatifard A (2005) Recruitment of MLL by HMG-domain protein iBRAF promotes neural differentiation. *Nature cell biology*

Jones RS, Gelbart WM (1993) The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. *Mol Cell Biol* **13:** 6357-6366

Jude CD, Climer L, Xu D, Artinger E, Fisher JK, Ernst P (2007) Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. *Cell Stem Cell* **1:** 324-337

Kapetanaki MG, Guerrero-Santoro J, Bisi DC, Hsieh CL, Rapić-Otrin V, Levine AS (2006) The DDB1-CUL4ADDB2 ubiquitin ligase is deficient in xeroderma pigmentosum group E and targets histone H2A at UV-damaged DNA sites. *Proc Natl Acad Sci USA* **103:** 2588-2593

Karatas H, Townsend EC, Bernard D, Dou Y, Wang S (2010) Analysis of the Binding of Mixed Lineage Leukemia 1 (MLL1) and Histone 3 Peptides to WD Repeat Domain 5 (WDR5) for the Design of Inhibitors of the MLL1-WDR5 Interaction. *J Med Chem* **53:** 5179-5185

Kiefer CM, Hou C, Little JA, Dean A (2008) Epigenetics of beta-globin gene regulation. *Mutat Res* **647:** 68-76

Kikelj D (2003) Peptidomimetic thrombin inhibitors. *Pathophysiol Haemost Thromb* **33:** 487-491

Kim A, Song S-h, Brand M, Dean A (2007) Nucleosome and transcription activator antagonism at human beta-globin locus control region DNase I hypersensitive sites. *Nucleic Acids Res* **35:** 5831-5838

Kim J, Daniel J, Espejo A, Lake A, Krishna M, Xia L, Zhang Y, Bedford MT (2006) Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep* **7:** 397-403

Kornberg RD (1977) Structure of chromatin. *Annu Rev Biochem* **46:** 931-954

Kouzarides T (2002) Histone methylation in transcriptional control. *Current opinion in genetics & development* **12:** 198-209

Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128:** 693-705

Kubicek S, O'Sullivan RJ, August EM, Hickey ER, Zhang Q, Teodoro ML, Rea S, Mechtler K, Kowalski JA, Homon CA, Kelly TA, Jenuwein T (2007) Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase. *Molecular cell* **25:** 473-481

Kukreti S, Kaur H, Kaushik M, Bansal A, Saxena S, Kaushik S, Kukreti R (2010) Structural polymorphism at LCR and its role in beta-globin gene regulation. *Biochimie* **92:** 1199-1206

Kwon H, Imbalzano AN, Khavari PA, Kingston RE, Green MR (1994) Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature* **370:** 477-481

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

Lee C-K, Shibata Y, Rao B, Strahl BD, Lieb JD (2004) Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet* **36:** 900-905

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* **39:** 1235-1244

Libertini LJ, Ausió J, van Holde KE, Small EW (1988) Histone hyperacetylation. Its effects on nucleosome core particle transitions. *Biophys J* **53:** 477-487

Lim DA, Huang Y-C, Swigut T, Mirick AL, Garcia-Verdugo JM, Wysocka J, Ernst P, Alvarez-Buylla A (2009) Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells. *Nature* **458:** 529-533

Linhart C, Halperin Y, Shamir R (2008) Transcription factor and microRNA motif discovery: The Amadeus platform and a compendium of metazoan target sets. *Genome Res* **18:** 1180

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **46:** 3-26

Lipps HJ, Rhodes D (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol* **19:** 414-422

Littlefield O, Nelson H (1999) A new use for the'wing'of the'winged'helix-turn-helix motif in the HSF–DNA cocrystal. *Nat Struct Mol Biol* **6:** 464-470

Littler DR, Alvarez-Fernandez M, Stein A, Hibbert RG, Heidebrecht T, Aloy P, Medema RH, Perrakis A (2010) Structure of the FoxM1 DNA-recognition domain bound to a promoter sequence. *Nucleic Acids Res* **38:** 4527-4538

Luger K (2006) Dynamic nucleosomes. *Chromosome Res* **14:** 5-16

Luger K, Hansen JC (2005) Nucleosome and chromatin fiber dynamics. *Curr Opin Struct Biol* **15:** 188-196

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389:** 251-260

Lusser A, Kadonaga JT (2003) Chromatin remodeling by ATP-dependent molecular machines. *Bioessays* **25:** 1192-1200

Martens JH, Stunnenberg HG (2010) The molecular signature of oncofusion proteins in acute myeloid leukemia. *FEBS Lett* **584:** 2662-2669

Martens JHA, Stunnenberg HG (2165) The molecular signature of oncofusion proteins in acute myeloid leukemia. *FEBS Lett* **584:** 2662-2669

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

Martin ME, Milne TA, Bloyer S, Galoian K, Shen W, Gibbs D, Brock HW, Slany R, Hess JL (2003) Dimerization of MLL fusion proteins immortalizes hematopoietic cells. *Cancer Cell* **4:** 197-207

Matharu NK, Hussain T, Sankaranarayanan R, Mishra RK (2103) Vertebrate Homologue of Drosophila GAGA Factor. *J Mol Biol* **400:** 434-447

Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC, Gilmour DS, Albert I, Pugh BF (2008) Nucleosome organization in the Drosophila genome. *Nature* **453:** 358-362

McGhee JD, Wood WI, Dolan M, Engel JD, Felsenfeld G (1981) A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion. *Cell* **27:** 45-55

McKinnell I, Ishibashi J, Le Grand F, Punch V, Addicks G, Greenblatt J, Dilworth F, Rudnicki M (2007) Pax7 activates myogenic genes by recruitment of a histone methyltransferase complex. *Nature cell biology* **10:** 77-84

McKinnell IW, Ishibashi J, Le Grand F, Punch VGJ, Addicks GC, Greenblatt JF, Dilworth FJ, Rudnicki MA (2008) Pax7 activates myogenic genes by recruitment of a histone methyltransferase complex. *Nature cell biology* **10:** 77-84

Mendez AJ (2010) The promise of apolipoprotein A-I mimetics. *Curr Opin Endocrinol Diabetes Obes* **17:** 171-176

Mersfelder EL, Parthun MR (2006) The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. *Nucleic Acids Res* **34:** 2653-2662

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 USA* **98:** 12902-12907

Milne TA, Dou Y, Martin ME, Brock HW, Roeder RG, Hess JL (2005) MLL associates specifically with a subset of transcriptionally active target genes. *Proc Natl Acad Sci USA* **102:** 14765-14770

Miranda TB, Cortez CC, Yoo CB, Liang G, Abe M, Kelly TK, Marquez VE, Jones PA (2009) DZNep is a global histone methylation inhibitor that reactivates developmental genes not silenced by DNA methylation. *Mol Cancer Ther* **8:** 1579-1588

Mizuguchi G, Tsukiyama T, Wisniewski J, Wu C (1997) Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. *Mol Cell* **1:** 141-150

Mulholland NM, King IFG, Kingston RE (2003) Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA. *Genes Dev* **17:** 2741-2746

Muramoto T, Muller I, Thomas G, Melvin A, Chubb JR (2010) Methylation of H3K4 Is Required for Inheritance of Active Transcriptional States. *Current Biology* **20:** 397-406

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

Nakanishi S, Sanderson BW, Delventhal KM, Bradford WD, Staehling-Hampton K, Shilatifard A (2008) A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. *Nat Struct Mol Biol* **15:** 881-888

Ng HH, Robert F, Young RA, Struhl K (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular cell* **11:** 709-719

Nowak SJ, Corces VG (2000) Phosphorylation of histone H3 correlates with transcriptionally active loci. *Genes Dev* **14:** 3003-3013

Obsil T, Obsilova V (2008) Structure/function relationships underlying regulation of FOXO transcription factors. *Oncogene* **27:** 2263-2275

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

Osley MA (2004) H2B ubiquitylation: the end is in sight. *Biochim Biophys Acta* **1677:** 74-78

Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in enzymology* **276:** 307-326

Paik WK, Paik DC, Kim S (2007) Historical review: the field of protein methylation. *Trends Biochem Sci* **32:** 146-152

Parkinson GN, Lee MPH, Neidle S (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417:** 876-880

Passarge E (1979) Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. *Am J Hum Genet* **31:** 106-115

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

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

Patel A, Vought VE, Dharmarajan V, Cosgrove MS (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

Pearson C, Zorbas H, Price G, Zannis-Hadjopoulos M (1996) Inverted repeats, stem-loops, and cruciforms: significance for initiation of DNA replication. *Journal of cellular biochemistry* **63:** 1-22

Peña PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R, Kutateladze TG (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442:** 100-103

Pennings S, Meersseman G, Bradbury EM (1991) Mobility of positioned nucleosomes on 5 S rDNA. *J Mol Biol* **220:** 101-110

Pérez-Cadahía B, Drobic B, Davie JR (2009) H3 phosphorylation: dual role in mitosis and interphase. *Biochem Cell Biol* **87:** 695-709

Perrakis A, Harkiolaki M, Wilson K, Lamzin V (2001) ARP/wARP and molecular replacement. *Acta Crystallographica Section D: Biological Crystallography* **57:** 1445-1450

Pflugrath J (1999) The finer things in X-ray diffraction data collection. *Acta Crystallographica Section D: Biological Crystallography* **55:** 1718-1725

Pinskaya M, Morillon A (2009) Histone H3 lysine 4 di-methylation: a novel mark for transcriptional fidelity? *Epigenetics* **4:** 302-306

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122:** 517-527

Popovic R, Zeleznik-Le NJ (2005) MLL: how complex does it get? *Journal of cellular biochemistry* **95:** 234-242

Probst AV, Dunleavy E, Almouzni G (2009) Epigenetic inheritance during the cell cycle. *Nat Rev Mol Cell Biol* **10:** 192-206

Qian C, Zhou M-M (2006) SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell Mol Life Sci* **63:** 2755-2763

Raghuram N, Carrero G, Th'ng J, Hendzel MJ (2009) Molecular dynamics of histone H1. *Biochem Cell Biol* **87:** 189-206

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

Rando OJ, Ahmad K (2007) Rules and regulation in the primary structure of chromatin. *Curr Opin Cell Biol* **19:** 250-256

Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406:** 593-599

Richmond TJ, Davey CA (2003) The structure of DNA in the nucleosome core. *Nature* **423:** 145-150

Ruthenburg AJ, Allis CD, Wysocka J (2007) Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular cell* **25:** 15-30

Ruthenburg AJ, Wang W, Graybosch DM, Li H, Allis CD, Patel DJ, Verdine GL (2006) Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nat Struct Mol Biol* **13:** 704-712

Sankaran VG, Xu J, Orkin SH (2010) Advances in the understanding of haemoglobin switching. *Br J Haematol* **149:** 181-194

Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NCT, Schreiber SL, Mellor J, Kouzarides T (2002) Active genes are tri-methylated at K4 of histone H3. *Nature* **419:** 407-411

Schalch T, Duda S, Sargent DF, Richmond TJ (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436:** 138-141

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

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

Schneider R, Bannister AJ, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T (2004b) Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol* **6:** 73-77

Schotta G, Ebert A, Krauss V, Fischer A, Hoffmann J, Rea S, Jenuwein T, Dorn R, 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

Schübeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, Gottschling DE, O'Neill LP, Turner BM, Delrow J, Bell SP, Groudine M (2004) The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev* **18:** 1263-1271

Schuetz A, Allali-Hassani A, Martín F, Loppnau P, Vedadi M, Bochkarev A, Plotnikov AN, Arrowsmith CH, Min J (2006) Structural basis for molecular recognition and presentation of histone H3 by WDR5. *The EMBO journal* **25:** 4245-4252

Sheffield P, Garrard S, Derewenda Z (1999) Overcoming Expression and Purification Problems of RhoGDI Using a Family of. *Protein expression and purification* **15:** 34-39

Shi X, Gozani O (2005) The fellowships of the INGs. *J Cell Biochem* **96:** 1127-1136

Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, Carney D, Peña P, Lan F, Kaadige MR, Lacoste N, Cayrou C, Davrazou F, Saha A, Cairns BR, Ayer DE, Kutateladze TG, Shi Y, Côté J, Chua KF, Gozani O (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442:** 96-99

Shukla A, Chaurasia P, Bhaumik SR (2008) Histone methylation and ubiquitination with their cross-talk and roles in gene expression and stability. *Cell Mol Life Sci*

Sims RJ, Reinberg D (2006) Histone H3 Lys 4 methylation: caught in a bind? *Genes Dev* **20:** 2779-2786

Slany RK (2009) The molecular biology of mixed lineage leukemia. *Haematologica* **94:** 984-993

So CW, Lin M, Ayton PM, Chen EH, Cleary ML (2003) Dimerization contributes to oncogenic activation of MLL chimeras in acute leukemias. *Cancer Cell* **4:** 99-110

Song J-J, Kingston RE (2008a) WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *J Biol Chem* **283:** 35258-35264

Song JJ, Kingston RE (2008b) WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *J Biol Chem* **283:** 35258-35264

Southall SM, Wong P-S, Odho Z, Roe SM, Wilson JR (2009a) Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol Cell* **33:** 181-191

Southall SM, Wong PS, Odho Z, Roe SM, Wilson JR (2009b) Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol Cell* **33:** 181-191

Stassen M, Bailey D, Nelson S, Chinwalla V, Harte P (1995) The Drosophila trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. *Mechanisms of development* **52:** 209-223

Steward MM, Lee J-S, O'Donovan A, Wyatt M, Bernstein BE, Shilatifard A (2006a) Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat Struct Mol Biol* **13:** 852-854

Steward MM, Lee JS, O'Donovan A, Wyatt M, Bernstein BE, Shilatifard A (2006b) Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat Struct Mol Biol* **13:** 852-854

Stoller JZ, Huang L, Tan CC, Huang F, Zhou DD, Yang J, Gelb BD, Epstein JA (2010) Ash2l interacts with Tbx1 and is required during early embryogenesis. *Exp Biol Med (Maywood)* **235:** 569-576

Stott K, Tang GSF, Lee K-B, Thomas JO (2006) Structure of a complex of tandem HMG boxes and DNA. *J Mol Biol* **360:** 90-104

Stros M, Launholt D, Grasser KD (2007) The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell Mol Life Sci* **64:** 2590-2606

Stros M, Muselíková E (2000) A role of basic residues and the putative intercalating phenylalanine of the HMG-1 box B in DNA supercoiling and binding to four-way DNA junctions. *J Biol Chem* **275:** 35699-35707

Takahashi YH, Lee JS, Swanson SK, Saraf A, Florens L, Washburn MP, Trievel RC, 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

Tan CC, Sindhu KV, Li S, Nishio H, Stoller JZ, Oishi K, Puttreddy S, Lee TJ, Epstein JA, Walsh MJ, Gelb BD (2008) Transcription factor Ap2delta associates with Ash2l and ALR, a trithorax family histone methyltransferase, to activate Hoxc8 transcription. *Proc Natl Acad Sci USA* **105:** 7472-7477

Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14:** 1025-1040

Tenney K, Shilatifard A (2005) A COMPASS in the voyage of defining the role of trithorax/MLL-containing complexes: linking leukemogensis to covalent modifications of chromatin. *J Cell Biochem* **95:** 429-436

Terranova R, Agherbi H, Boned A, Meresse S, Djabali M (2006a) Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc Natl Acad Sci USA* **103:** 6629-6634

Terranova R, Agherbi H, Boned A, Meresse S, Djabali M (2006b) Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc Natl Acad Sci U S A* **103:** 6629-6634

Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular cell* **10:** 1453-1465

Travers A, Caserta M, Churcher M, Hiriart E, Di Mauro E (2009) Nucleosome positioning--what do we really know? *Mol Biosyst* **5:** 1582-1592

Travers AA, Klug A (1987) The bending of DNA in nucleosomes and its wider implications. *Philos Trans R Soc Lond, B, Biol Sci* **317:** 537-561

Trievel R, Shilatifard A (2009) WDR5, a complexed protein. *Nat Struct Mol Biol* **16:** 678-680

Trojer P, Reinberg D (2007) Facultative heterochromatin: is there a distinctive molecular signature? *Molecular cell* **28:** 1-13

Tsai K-L, Sun Y-J, Huang C-Y, Yang J-Y, Hung M-C, Hsiao C-D (2007) Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification. *Nucleic Acids Res* **35:** 6984-6994

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. *The EMBO journal* **13:** 3822-3831

Vagin A, Steiner R, Lebedev A, Potterton L, McNicholas S, Long F, Murshudov G (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica Section D: Biological Crystallography* **60:** 2184-2195

Vagin A, Teplyakov A (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallographica Section D: Biological Crystallography* **56:** 1622-1624

Valeyev N, Downing A, Sondek J, Deane C (2008) Electrostatic and Functional Analysis of the Seven-Bladed WD $# x003b2;-Propellers. *Evolutionary Bioinformatics Online* **4:** 203

Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, Sidow A, Fire A, Johnson SM (2008) A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. *Genome Res* **18:** 1051-1063

Waga S, Mizuno S, Yoshida M (1990) Chromosomal protein HMG1 removes the transcriptional block caused by the cruciform in supercoiled DNA. *J Biol Chem* **265:** 19424-19428

Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, Zhang Y (2004) Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431:** 873-878

Wang P, Lin C, Smith ER, Guo H, Sanderson BW, Wu M, Gogol M, Alexander T, Seidel C, Wiedemann LM, Ge K, Krumlauf R, 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:** 6074-6085

Wang Q, Zeng M, Wang W, Tang J (2007) The HMGB1 acidic tail regulates HMGB1 DNA binding specificity by a unique mechanism. *Biochemical and biophysical research communications* **360:** 14-19

Wang Z, Song J, Milne TA, Wang GG, Li H, Allis CD, Patel DJ (2010) Pro Isomerization in MLL1 PHD3-Bromo Cassette Connects H3K4me Readout to CyP33 and HDAC-Mediated Repression. *Cell* **141:** 1183-1194

Weigelt J, Climent I, Dahlman-Wright K, Wikström M (2001) Solution structure of the DNA binding domain of the human forkhead transcription factor AFX (FOXO4). *Biochemistry* **40:** 5861-5869

White AW, Westwell AD, Brahemi G (2008) Protein-protein interactions as targets for small-molecule therapeutics in cancer. *Expert Rev Mol Med* **10:** e8

Woo JS, Suh HY, Park SY, Oh BH (2006) Structural basis for protein recognition by B30.2/SPRY domains. *Mol Cell* **24:** 967-976

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

Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, Roeder RG, Brivanlou AH, Allis CD (2005) WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121:** 859-872

Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P, Wu C, Allis CD (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442:** 86-90

Xiao B, Jing C, Kelly G, Walker PA, Muskett FW, Frenkiel TA, Martin SR, Sarma K, Reinberg D, Gamblin SJ, Wilson JR (2005a) Specificity and mechanism of the histone methyltransferase Pr-Set7. *Genes Dev* **19:** 1444-1454

Xiao B, Tarricone C, Lin K, Kelly… G (2007) Optimizing Protein Complexes for Crystal Growth†. *Crystal Growth and ...*

Xiao B, Wilson JR, Gamblin SJ (2003) SET domains and histone methylation. *Curr Opin Struct Biol* **13:** 699-705

Xiao T, Kao C-F, Krogan NJ, Sun Z-W, Greenblatt JF, Osley MA, Strahl BD (2005b) Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Mol Cell Biol* **25:** 637-651

Yang H, Mizzen CA (2009) The multiple facets of histone H4-lysine 20 methylation. *Biochem Cell Biol* **87:** 151-161

Yokoyama A, Wang Z, Wysocka J, Sanyal M, Aufiero DJ, Kitabayashi I, Herr W, Cleary ML (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

Yu BD, Hanson RD, Hess JL, Horning SE, Korsmeyer SJ (1998) MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. *Proc Natl Acad Sci USA* **95:** 10632-10636

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

Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. *Science* **309:** 626-630

Zeng L, Zhou MM (2002) Bromodomain: an acetyl-lysine binding domain. *FEBS Lett* **513:** 124-128

Zhang X, Bruice TC (2008) Mechanism of product specificity of AdoMet methylation catalyzed by lysine methyltransferases: transcriptional factor p53 methylation by histone lysine methyltransferase SET7/9. *Biochemistry* **47:** 2743-2748

Zhang X, Yang Z, Khan SI, Horton JR, Tamaru H, Selker EU, Cheng X (2003) Structural basis for the product specificity of histone lysine methyltransferases. *Molecular cell* **12:** 177-185

Zhou W, Wang X, Rosenfeld MG (2009) Histone H2A ubiquitination in transcriptional regulation and DNA damage repair. *Int J Biochem Cell Biol* **41:** 12-15

CHAPTER 5 - STATEMENT OF CONTRIBUTORS AND COLLABORATORS

5.1 – Contributions to "Identification of the minimal structural determinants underlying the allosteric regulation of Myeloid Lymphoma Leukemia protein methyltransferase activity."

Authors: Vanja Avdic*, Pamela Zhang*, Sylvain Lanouette, Adam Groulx, Sylvain Lanouette, Veronique Tremblay, and Jean-Francois Couture
***These authors contributed equally to the article.**

Avdic V. completed work on cloning, expressing and purifying full length and different protein constructs used in study. Additionaly, she worked on mapping the interaction between RbBP5 and WDR5. Finally, she developed and optimized methyltransferase assay used in the study.

5.2 Contribtions to "An atypical helix-wing-helix domain mediates Ash2L binding to the *beta-globin* gene".

Authors: Vanja Avdic*, Sabina Sarvan*, Véronique Tremblay*, Chandra-Prakash Chaturvedi, Sylvain Lanouette, Alexandre Blais, Joseph S. Brunzelle, Marjorie Brand Jean-François Couture
***These authors contributed equally to the article.**

Avdic V. completed work on cloning, expressing and purifying full length and different protein constructs and mutants used in study. She also genereated the Ash2L-DNA model and used gel shift to map the domain of Ash2L responsible for the DNA binding activity. Finally, she developed a method to test binding to the different structures of DNA that the probe for gel shift undertook.

5.3 – Fine-tuning the stimulation of MLL1 methyltransferase activity by a histone H3 based peptide mimetic

Authors: Vanja Avdic*, Pamela Zhang*, Sylvain Lanouette, Anastassia Voronova, Ilona Skerjanc and Jean-Francois Couture

***These authors contributed equally to the article .**

Avdic V. was responsible for protein expression and purification. She also grew crystals of and solved the structure of WDR5 in complex with the NaH3 peptide. Finally, she was responsible for developing an assay to test the *in vitro* inhibitory effect of the NaH3 peptide.

CHAPTER 7 – SUPPLEMENTARY MATERIALS

7.1 – Supplementary materials for "An atypical helix-wing-helix domain mediates Ash2L binding to the *beta-globin* gene"

7.1.1 – Supplementary Figures

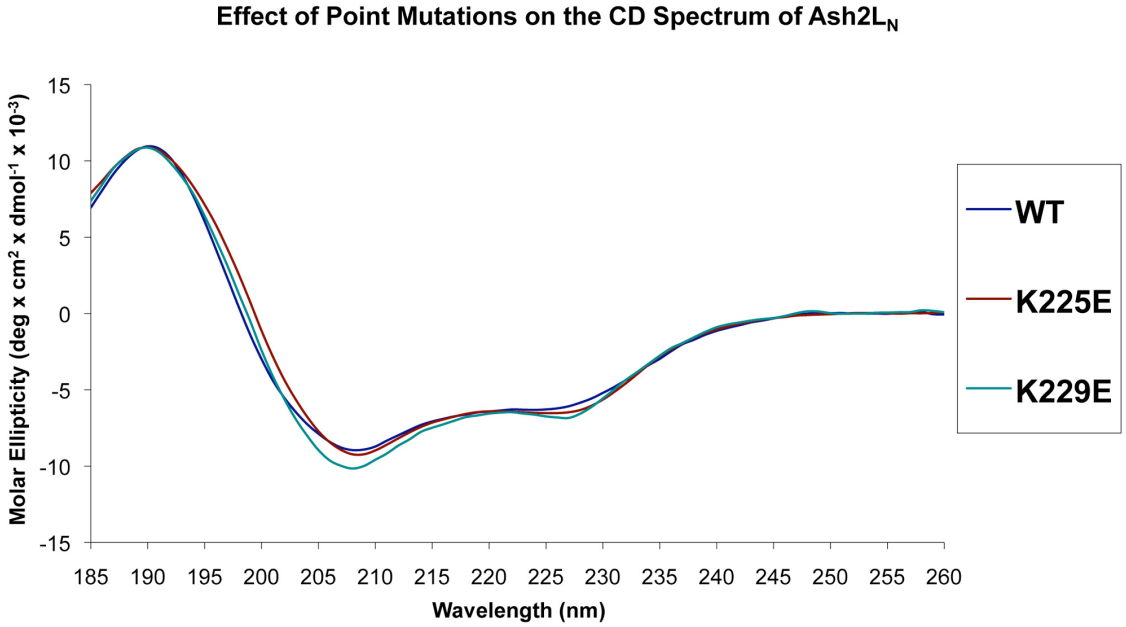


Figure S1. Mutation of Ash2L$_N$ does not impair protein folding. Collection of circular dichroism spectra for the wild type and mutant proteins was performed using a Chirascan Circular Dichroism Spectrometer (Applied Photophysics Ltd., Leatherhead, Surrey, U.K.). Each protein (or buffer) sample was placed in a 0.1mm pathlength cylindrical cuvette for analysis, and seven replicate scans were obtained. Processing of raw spectra was performed using the Pro-data Suite software package (version 4.1.1, Applied Photophysics Ltd., Leatherhead, Surrey, U.K.). The raw spectrum for each protein sample was corrected for baseline anomalies by subtraction of the spectrum for a corresponding buffer blank, and replicate scans were averaged. These corrected spectra were smoothed using a 3-point window, which showed no discernable systemic variability in the residuals.

7.1.2 – Supplementary Tables

Table S1 : Primers for chromatin immunoprecipitation experiments of Ash2L target genes in Hela cells.	
Neg. Ctl.-forward	5’-TTCTGTCCTGCATCGTCTTG-3’
Neg. Ctl.-reverse	5’-TTTGCAGTGTGATTCCCAAC-3’
TPP1-forward	5’-TCAGAGGGTGGTGGACTAGG-3’
TPP1-reverse	5’-TCCTAGGACTTGTTCGGGC-3’
CAPN3-forward	5’-TAAGCCGAAAAAGTTCCAGG-3’
CAPN3-reverse	5’-TTGCAAACTTCTCAGCATGG-3’
ALX1-forward	5’-CATTAGAACTGGGAGGCTGC-3’
ALX1-reverse	5’-CCGGTCCCTTTAATACGTCC-3’
NHEJ1-forward	5’-AAGGCCAATTCCACTCACC-3’
NHEJ1-reverse	5’-GAGGCATCTAGGATTCTTGTGG-3’

7.1.3 – Supplementary Methods

Circular dichroism. Collection of circular dichroism spectra for the wild type and mutant proteins was performed using a Chirascan Circular Dichroism Spectrometer (Applied Photophysics Ltd., Leatherhead, Surrey, U.K.). Each protein (or buffer) sample was placed in a 0.1mm path length cylindrical cuvette for analysis, and seven replicate scans were obtained. Processing of raw spectra was performed using the Pro-data Suite software package (version 4.1.1, Applied Photophysics Ltd., Leatherhead, Surrey, U.K.). The raw spectrum for each protein sample was corrected for baseline anomalies by subtraction of the spectrum for a corresponding buffer blank, and replicate scans were averaged. These corrected spectra were smoothed using a 3-point window, which showed no discernable systemic variability in the residuals.