Structure-Function studies of the recruitment of DOT1L, an H3K79 histone methyltransferase, to MLL-AF9 and its role in leukemogenesis

Aravinda Kuntimaddi
Sterling, Virginia

B.A., Northwestern University, 2005
M.S., George Mason University, 2008
M.S., University of Virginia, 2012

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Interdisciplinary Program in Biophysics

University of Virginia
May, 2015

Abstract

The *MLL* gene is a common target of chromosomal translocations found in human leukemia. MLL-fusion leukemias consistently have poor outcomes and disrupt *HOX* gene regulation through the recruitment of transcriptional elongation factors. One of the most common translocation partners of MLL is AF9 (MLLT3). The C-terminal domain of AF9 has been shown to interact with multiple transcriptional regulators. This t(9;11) chromosomal translocation forms a chimeric MLL-AF9 protein containing the functionalities of both the N-terminal portion of MLL and the C-terminal part of AF9. The MLL-AF9 protein recruits DOT1L, a histone H3 lysine 79 (H3K79) methyltransferase which mono-, di- or tri methylates H3K79 (H3K79me1/me2/me3), leading to aberrant gene transcription. The role of the direct interaction between DOT1L and MLL-AF9 and its role in leukemogenesis was unclear.

We show that DOT1L has three AF9 binding sites with varying affinities, and that both partners are disordered and co-fold upon forming separate DOT1L-AF9 complexes. We present the NMR solution structure of the highest affinity DOT1L-AF9 complex. Based upon this structural knowledge, we generated structure-guided point mutations with graded effects on recruitment of DOT1L to MLL-AF9. ChIP-Seq analyses of H3K79me2 and H3K79me3 show that graded reduction of the DOT1L interaction with MLL-AF9 results in differential losses in H3K79me2 and me3 marks at MLL-AF9 target genes. Furthermore, the degree of MLL-AF9 hematopoietic transformation as seen in serial replating assays are dependent upon the level of DOT1L recruitment.

Additionally, the MLL-AF9 fusion protein recruits the CBX8 protein. CBX8 is a member of the PRC1 complex and reads the histone H3 lysine 27 tri-methylation mark (H3K27me3), which is characteristically involved in transcriptional silencing. Here, we present the NMR solution structure of CBX8 in complex with AF9 and we show that the CBX8-AF9 structure is nearly identical to that of the other AF9 complex structures. The novelty of this particular interaction is that it has extremely weak binding affinity and has increased backbone dynamics compared to that of the other AF9 complexes. It is still a conundrum as to why the MLL-AF9 protein recruits CBX8 and what role this interaction plays in leukemogeneis. We propose several structure-guided functional studies that will further shed light on the role of this direct interaction in MLL-AF9 leukemogenesis.

Acknowledgements

I would first like to thank my mentor, Dr. John Bushweller, for giving me enormous support and encouragement throughout my six years in the lab. Upon joining the lab in April 2009, I was very excited about delving into cancer research but was new to structural biology, had minimal experience in protein biochemistry, and clearly had much to learn. I thank John for being patient with me as I embarked upon the steep learning curve required to be successful in the Bushweller lab, and for giving me leeway to pursue my own ideas and develop my own projects. Every time I came upon a roadblock, or dead-end, John would always encourage me to come up with a number of creative ideas that would help move the project forward. In the end, I find myself lucky, as I have been given support to be able to travel and present my work at numerous conferences, apply and receive independent grants, and publish in different journals. I believe that the Bushweller lab has provided me with the best scientific experiences that I could have asked for in graduate school, and I do not think I could have asked for a better mentor. All of my future scientific success will undoubtedly be due to the support of John.

I also thank my thesis committee, Dr. Robert Nakamoto, Dr. Patrick Grant, Dr. Stefan Bekiranov and Dr. Mazhar Adli for advice and discussions over the years. Dr. Robert Nakamoto first recruited me into the Molecular Medicine program in January 2008. He has given me numerous opportunities to recruit students into the Biophysics program, run departmental retreats, and present my

work at different venues. I thank Bob for always providing me with outstanding advice and support.

Dr. Stefan Bekiranov has been an invaluable source of information and has provided me with much appreciated career advice throughout my time at UVA. Without Stefan's help I would have never been able to complete my thesis project. I thank Dr. Mazhar Adli for allowing me to work with his lab, for all the advice he has given me over the past two years, and for serving on my thesis committee. I especially thank Jeremy Thorpe and Ritambhara Singh, both in the Adli lab, who helped me greatly throughout the course of this project.

Additionally I am grateful to our collaborator Dr. Nancy Zeleznik-Le for allowing me to collaborate and publish multiple papers with their lab. This project would not have been completed without help from Alyson Lokken and Nicholas Achille, both in the Zeleznik-Le laboratory.

I could not have survived graduate school without the support of my colleagues and friends in graduate school, especially those that have been in the Bushweller lab. I could write stories of the mischief that we all got into, and while we were a dysfunctional family, at times, the members of the lab became my best friends, intramural teammates, and hangout buddies. I especially thank two former students, Dr. Ben Leach and Dr. Mike Regan, who both taught me much of the protein science protocols and NMR techniques required to be successful in the lab. Ben, in particular, started the AF9 project and continues to be a great source of help whenever I need it. In addition to these two, I thank my colleagues Yunpeng, Xuan, Chuck, Adam and Annie, and our army of current and former

undergraduate students, especially Colette Gnade, for allowing the lab to become a home away from home.

I would not have made the choice to attend a PhD program without the support of my former mentor at George Mason University, Dr. Dan Cox, who has now moved on to Georgia State University and has become extremely successful. I also would be remiss if I did not mention three of my Park View High School science teachers, Ms. Denise Wingfield, Ms. Maryann Skrzycki, and Ms. Nancy Colosi. These three were best science teachers that I have ever had, and started me on the path to a career in research.

Last but not least, my deep gratitude goes to my parents and family for their support throughout this journey. Without their continued support from early on, I would have never made it this far.

List of Abbreviations

aa	Amino Acid
AF4	ALL-Fused protein from chromosome 4
AF5	ALL-Fused protein from chromosome 5
AF9	ALL-fused protein from Chromosome 9
AF10	ALL-fused protein from Chromosome 9
AHD	ANC1 Homology Domain
ALL	Acute Lymphocytic Leukemia
AML	Acute Myeloid Leukemia
BCoR	BCL6 Corepressor
CBP	CREB Binding Protein
CBX	Chromobox homolog
CBX8	Chromobox homolog 8
CD	Circular Dichroism
ChIP	Chromatin ImmunoPrecipitation
Co-IP	Co-Immunporecipitation
CTD	C-Terminal Domain
D544R	Aspartic Acid to Arginine mutation on AF9 aa. 544
D546R	Aspartic Acid to Arginine mutation on AF9 aa. 546
DTT	Di ThioThreitol
DOT1L	Disruptor of Telomeric Silencing 1-Like
EDTA	Ethylene Diamine Tetra-Acetate
ENaC	Epithelial Sodium Channel a

ENL	Eleven-Nineteen Leukemia
FP	Fluorescence Polarization
FMR2	Fragile X Mental Retardation Syndrome protein 2
HDAC	Histone Deacetylase
H2B-K123ub	ubiquitinated histone H2B lysine 123
H3K4	histone H3 lysine 4
H3K4me3	tri-methylated histone H3 lysine 4
H3K27	histone H3 lysine 27
H3K27me3	tri-methylated histone H3 lysine 27
H4K16ac	Acetylated histone H4 lysine 16
H3K79	histone H3 lysine 79
H3K7me	methylated histone H3 lysine 79
H3K79me2/3	di- and tri-methylated histone H3 lysine 79
HMT	Histone Methyl Transferase
HOX	Homeobox
HSQC	Heteronuclear Single Quantum Coherence
IDP	Intrinsically Disordered Protein
IPTG	IsoPropyl β-D-ThioGalactoside
ITC	Isothermal Calorimetry
Kd	Dissociation constant
LEDGF	Lens Epithelium Derived Growth Factor
MBP	Maltose Binding Protein
MEIS1	Myeloid ecotropic viral integration site 1

MENIN	Multiple Endocrine Neplasia
MLL	Mixed Lineage Leukemia
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
NTD	N-Terminal Domain
PAFc	(PolII Associated Factors) complex
PHD	Plant HomeoDomain
PONDR	Predictor of Naturally Disordered Regions
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
P-TEFb	Positive Transcriptional Elongation Factor b
RI	Longitudinal Relaxation
R2	Transverse relaxation
RDI, RD2	Repression Domain 1, Repression Domain 2
RDC	Residual Dipolar Coupling
RMSD	Root Mean Square Deviation
RNA PolII	RNA Polymerase II
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SET	SuVar 3-9, E(z), Trx domain
UTR	Untranscribed Region
WT	Wild Type
WRAD	WDR5, RbBP5, Ash2L, Dpy-30

YEATS	yAF9, ENL, AF9, Tafl4, SAS5

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Mixed Lineage Leukemia

The mixed lineage leukemia (MLL) protein positively regulates gene expression during development. MLL is associated with over 5000 human promoters suggesting that it may have a global role in transcription (Sedkov et al., 1994; Yu et al., 1995; Guenther et al., 2005). The MLL protein is a human homolog of the *Drosophila* Trx and the *S. cerevisiae* Set 1/Compass complex. MLL is a histone H3 lysine 4 (H3K4) methyltransferase that is required for the maintenance of spatial patterns of *HOX* gene expression during development and hematopoiesis (Rice and Licht, 2007; Popovic and Zeleznik-Le, 2005; Yu et al., 1995, 1998). Homozygous disruption of *Mll* in mice leads to embryonic lethality at E10.5 and is associated with numerous defects in segmental identity (Yu et al., 1995). Although correct initiation of *Hox* gene expression in these homozygous embryos is achieved early in embryogenesis, subsequent maintenance of *Mll*-dependent *Hox* gene expression is lost (Yu et al., 1998). MLL is therefore not required for initiation, but is crucial for maintenance of gene expression.

The *MLL* gene, located on chromosome 11, is a common target of chromosomal translocations found in acute leukemias affecting both children (Sorensen et al., 1994) and adults (Cox et al., 2004). Chromosomal rearrangements at 11q23 fuse an N-terminal fragment of MLL to over 70 different nuclear, cytoplasmic, or membrane partners (Meyer et al., 2013). Regardless of

the fusion partner, the presence of *MLL* translocations is associated with early relapse and poor prognosis (Dimartino and Cleary, 1999). *MLL* leukemia accounts for up to 10% of acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL), but is over-represented in two subpopulations of patients: infants, where MLL is involved in up to 80% of acute leukemias, and patients previously treated for other cancers, in whom the use of drugs that target DNA topoisomerase II can trigger *MLL* rearrangements (Ford et al., 1993; Krivtsov and Armstrong, 2007; Muntean and Hess, 2012; Super et al., 1993). Canonical targets of MLL-fusion leukemias include *HOXA9* and *MEIS1*. Knockdown of HOXA9 leads to abrogation of leukemic transformation by hematopoietic progenitors and overexpression of either *HOXA9* and *MEIS1* leads to transformation. Thus, expression of *HOXA9* and *MEIS1* are the major causes of transformation in MLL-fusion leukemia (Ayton and Cleary, 2003; Wong et al., 2007). Compared to other types of leukemia, MLL-rearranged leukemias contain significantly lower cure rates (Krivtsov and Armstrong, 2007; Muntean and Hess, 2012), and further understanding of the molecular mechanisms underlying these leukemias may allow for the development of better treatment strategies.

1.2. Structural organization of the MLL Protein

MLL is a large 3970 amino acid protein consisting of many repressing and activating domains (**Figure 1-1A**). After the MLL protein is translated, wild type MLL is proteolytically cleaved via taspase cleavage into N-terminal (300 kDa) and C-terminal (180 kDa) domains. These domains then dimerize via the FYRN and FYRC domains.

The N-terminal portion of MLL forms a trimeric complex with MENIN and LEDGF whereby the PWWP domain of LEDGF is responsible for the targeting of MLL to chromatin (Yokoyama et al., 2004). The N-terminal region of MLL also contains 3 AT-hook motifs that bind to the minor groove of AT-rich genomic DNA sequences. N-terminal MLL also contains two repressive regions, RD1 and RD2. Within RD1, the CXXC domain binds to unmethylated CpG DNA, which protects CpG residues from methylation. Methylation of CpG islands is a means to promote transcriptional repression by preventing the binding of transcription factors (Erfurth et al., 2008). The RD2 domain recruits histone deacetylases HDAC1 and HDAC2. The N-terminal region of MLL contains four PHD domains. The third PHD domain (PHD3) binds to trimethylated histone H3 lysine 4 (H3K4me3). PHD3 is involved in a switch-like regulatory interaction with Cyclophilin33 (CYP33). CYP33 contains an N-terminal RNA-recognition motif (RRM) domain and a C-terminal a peptidyl-prolyl isomerase domain. The MLL PHD3 domain is proximal to a flanking bromodomain region and is connected to this region by a cis-proline linker. CYP33 is able to isomerize this linker, which disrupts the PHD3 interaction with the bromodomain region, thereby enabling binding of CYP33 to the PHD3 domain. We have previously shown that binding of H3K4me3 to PHD3 and the binding or CYP33 RRM domain to PHD3 are mutually inhibitory. Thus, a proposed mechanism is that high levels of CYP33 lead to the isomerization of the cis-proline linker and to the binding of CYP33 RRM to the PHD3 domain. This prevents the PHD3 domain from binding to

H3K4me3, which in turn leads to recruitment of H3K4 demethylases and transcriptional silencing (Park et al., 2010; Wang et al., 2010).

The C-terminal portion of MLL includes two domains. The TAD domain directly binds to the KIX domain of the CREB-binding protein (CBP) and is involved in H3K27 acetylation at promoters (Ernst et al., 2001). Lastly, this C-terminal domain of MLL contains a SET domain, a H3K4 histone methyltransferase, which is a mark associated with transcriptional activation. This domain of MLL is still poorly understood.

1.3. MLL fusion proteins

1.3.1. Structural organization of MLL fusion proteins

Rearrangements of MLL that are observed clinically include reciprocal translocations, complex translocations (De Braekeleer et al., 2010) and partial tandem duplications (Yu et al., 1996). The most common are reciprocal translocations in which the N-terminal portion of MLL is fused in frame with the C-terminal portion of the fusion protein (**Figure 1-1B**). There are only a few examples of clinically relevant translocations where the N-terminal portions of fusion proteins are fused in frame to the C-terminal portion of MLL, an example being AF4-MLL (Bursen et al., 2010). But these types of rearrangements have only been sparsely studied and much is still left to be learned about its mechanisms of action.

The structures of MLL fusion proteins are generally conserved. Upon MLL-rearrangements, the N-terminal portion of MLL of the fusion protein retains the MENIN/LEDGF interactions and the CXXC domain but loses the PHD3 and SET

domains. Disrupting the MENIN interaction with MLL fusion proteins blocks the development of acute leukemia in vivo while abrogating the oncogenicity of MLL fusions (Yokoyama et al., 2005). Recent efforts have been geared towards developing small molecules towards the MLL-MENIN interaction. These inhibitors are successfully able to block the progression of MLL-AF9 leukemia in vitro and in vivo (Borkin et al., 2015).

Inclusion of the CXXC domain has been shown to be required for transformation by MLL fusions. Previous work completed by our lab introduced structure-guided point mutations into the CXXC domain that disrupted its interaction with DNA CpG islands. This led to losses in *HOXA9* expression, increased DNA methylation at target promoters of MLL, and subsequent loss of transformation in MLL-AF9 leukemias. By protecting these CpG islands from methylation, the MLL-fusion proteins allow for expression of leukemic target genes (Cierpicki et al., 2010; Erfurth et al., 2008).

1.3.2. MLL fusions and misregulation of transcriptional elongation

Truncations of MLL do not result in hematopoietic transformation, therefore the oncogenic properties of MLL-rearrangements are directly due to the fusion partners. MLL-fusion partners AF9, ENL, AF4 and AF10 account for over two-thirds of MLL-rearrangements (Krivtsov and Armstrong, 2007). The t(9;11) rearrangement that produces the MLL-AF9 fusion are found in 2 to 5% of AML cases and up to 25% of de novo cases of AML in children. The median survival for the de novo cases hover around four years. Additionally, the t(11;19) rearrangement is associated with MLL-ENL fusions and are mostly found in

infants less than one year old who have biphenotypic or B cell ALL. Survival rates of MLL-ENL fusions are extremely poor, and are less than one year (Muntean and Hess, 2012).

MLL fusion proteins have been shown to associate with each other as members of transcriptional elongation complexes, suggesting that misregulation of transcriptional elongation is a common mechanism in MLL-dependent leukemogenesis (Biswas et al., 2011; Lin et al., 2010; Mohan et al., 2010; Mueller et al., 2007; Yokoyama et al., 2010). The ENL and AF9 proteins each have an N-terminal YEATS domain that enables them to interact with both the MLL CXXC domain and the Polymerase II associated factors complex (PAFc), a complex that contains many proteins that play roles in transcriptional elongation. This YEATS-CXXC-PAFc interaction potentially provides another mechanism for regulation of transcriptional elongation (He et al., 2011; Milne et al., 2010; Muntean et al., 2010). The N-terminal portions of these fusion partner proteins are lost in the MLL-rearrangements whereas the C-terminal domains are retained; in AF9 this domain is referred to as the ANC1 homology domain (AHD). This enables the MLL-fusion protein to constitutively interact with members of these transcriptional elongation complexes.

Over the last several years, it is accepted two important molecular events for hematopoietic transformation by MLL-fusions include:

1) Recruitment of P-TEFb, a Cyclin/CDK complex

2) Recruitment of DOT1L, a histone H3 lysine 79 methyltransferase

After transcription initiation, RNA Polymerase II pauses after transcription of approximately 50 bases (Rougvie and Lis, 1988). Phosphorylation of Serine 5 of the RNA PolII C-terminal Domain (CTD) leads to the dissociation of transcription initiation factors and the subsequent recruitment of pausing related-factors such as DSIF and NELF (Fuda et al., 2009; Peterlin and Price, 2006; Wu, 2003). This transcriptional pausing serves as another means of gene regulation. Recruitment of P-TEFb, a Cyclin/Cyclin Dependent Kinase complex, leads to the phosphorylation of Serine 2 in the RNA PolII CTD, and the phosphorylation NELF and DSIF leading to their dissociation (Ahn et al., 2004). Following these events, RNA PolII can transition to the transcription elongation phase.

The AF9 AHD can recruit AF4 family members leading to the subsequent phosphorylation of RNA polymerase II via P-TEFb (Bitoun et al., 2007; Erfurth et al., 2004), as well as DOT1L, a methyltransferase responsible for histone 3 lysine 79 (H3K79) methylation, a mark associated with active transcription (Steger et al., 2008; Zhang et al., 2006). The constitutive recruitment of these proteins by MLL-AF9, combined with the gene-specific recognition binding domains of the N-terminal portion of MLL (Ayton et al., 2004; Zeleznik-Le et al., 1994) leads to dysregulated expression of MLL target genes such as *HOXA9* and *MEIS1*, decreased differentiation, and increased self-renewal (Krivtsov and Armstrong, 2007; Muntean and Hess, 2012). Interestingly, AF9 also binds to BCL6 corepressor (BCoR) (Srinivasan et al., 2003) and Polycomb 3 (CBX8/hPC3) (Hemenway et al., 2001a), proteins generally associated with transcriptional repression. While a role for BCoR is not yet fully established in mixed lineage

leukemia, CBX8 has been shown to be involved in MLL-AF9 transformation (Tan et al., 2011).

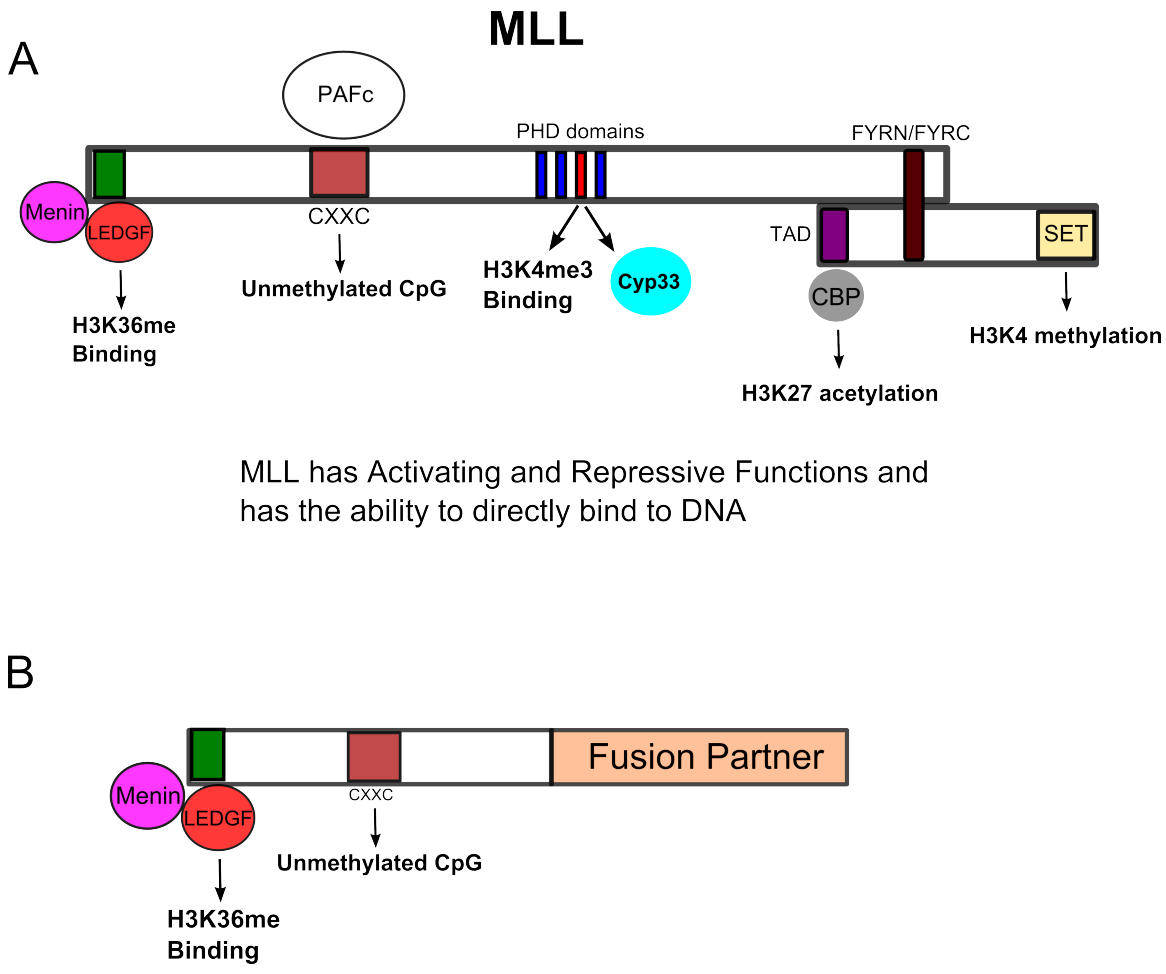


Figure 1-1. Schematic of MLL and MLL-Fusion proteins

A) MLL is cleaved and the N-terminal and C-terminal portions dimerize. The MLL proteins has many different DNA and Protein binding domains.

B) The N-terminal portion of MLL is fused in frame to over 70 different fusion partners. The fusion protein keeps the CXXC and the MENIN/LEDGF binding domains but loses the PHD and SET domains.

1.4. AF9 and its physiological roles

The *AF9* gene encodes for a serine-rich (20%), proline-rich (7%) protein that is 568 amino acids long. AF9 is an evolutionarily conserved protein in almost all eukaryotes and consists of an N-terminal YEATS domain and C-terminal ANC1 homology domain (AHD) that are both conserved. AF9 (MLLT3) and its homolog ENL are most commonly known for their roles as fusion partners in MLL-rearranged leukemias (Muntean and Hess, 2012).

The AF9 protein functions as a transcriptional activator in reporter gene assays. ANC1/TFG/TAF30, a related protein in yeast, is a component of the SWI/SNF, TFIID, and TFIIF complexes and is required for transcriptional control (Cairns et al., 1996). Homozygous deletion of *AF9* in mice leads to lethality shortly after birth and *AF9* heterozygous mice display skeletal abnormalities resulting due to inappropriate expression of genes of the *Hox* family (Collins et al., 2002). The full function of wild-type AF9, as well as the mechanistic basis of its roles, is beginning to be understood.

1.4.1. The YEATS domain of AF9 binds directly to H3K9ac

The first 140 residues of the N-terminus of AF9 comprises of the YEATS (Yaf9, Enl, AF9, Taf14, Sas5) domain, which shares a sequence conserved from yeast to human with family members found in over 70 species. The something about silencing (SAS) complex is a multiprotein complex that acetylates histone H4 lysine 16 and histone H3 lysine 14 (Osada et al., 2001). Sas5, in particular,

contains a YEATS domain, also found in two other yeast proteins, Taf14 and Yaf9. Yeast strains deficient for all three YEATS containing proteins are not viable, suggestive of the domain's essential function in *S. cerevisiae* (Zhang et al., 2004).

Members of the YEATS protein family are present in multiple protein complexes involved in histone modification, transcription regulation and chromatin remodeling (Schulze et al., 2009a). It was thought that the AF9 YEATS domain potentially bound to histone H3 and, until recently, the role of the YEATS domain was unclear. Li, et al. showed that the AF9 YEATS domain binds strongly to histone H3 lysine 9 (H3K9ac) (**Figure 1-2**) and binds, albeit more weakly, to acetylated Histone H3 lysine 27 (H3K27ac) and Histone H3 lysine 18 (H3K18ac). Additionally, they crystallized the AF9 YEATS domain in complex with H3K9ac and showed through ChIP-Seq studies that AF9 YEATS colocalizes with the H3K9ac mark. The novelty of this study is that they showed that the YEATS domain is an acetyl-lysine reader. Furthermore, they claim that reading of the H3K9ac mark by AF9 is necessary for the recruitment of DOT1L to chromatin, which leads to H3K79 methylation at target genes (Li et al., 2014).

1.4.2. The AF9 (AHD) and its interacting partners

The C-terminal of AF9, or AF9 (AHD) is a conserved region that contributes to both normal and neoplastic blood cell development. When fused to MLL, this region of AF9 is necessary and sufficient for immortalization of hematopoietic progenitor cells (Collins et al., 2000). MLL-AF9, as well as MLL-ENL, promotes leukemogenesis by dysregulating downstream genes through a

gain-of-function provided by the transcriptional effector properties of ENL or AF9 (Slany et al., 1998). Biochemical studies have shown that the AF9 (AHD) binds to AF4 and DOT1L, both transcriptional activators, and at least two co-repressor proteins, the Polycomb 3 protein (hPc3/CBX8) and the BCL-6 co-repressor (BCoR) (Erfurth et al., 2004; Hemenway et al., 2001; Srinivasan et al., 2003) (**Figure 1- 2**).

1.4.3. Transcriptional repressors, BCoR and CBX8

It is puzzling as to why AF9, and consequently, MLL-AF9 would bind to two proteins involved in transcriptional repression, BCoR and CBX8. BCoR (BCL-6 corepressor molecule) was initially identified from its interaction with BCL-6, a protein that is a key regulator of B cell development and is frequently dysregulated in non-Hodgkin's lymphoma. This protein is a key transcriptional regulator of early embryonic development, stem cell function and hematopoiesis. In binding to BCL-6, BCoR augments the ability of BCL6 to act as a transcriptional repressor (Huynh et al., 2000).

BCoR is expressed in human tissues as one of several different splice variants, but only two of four isoforms in mice are capable of binding to AF9. It has been shown to be involved in a complex with histone ubiquitination activity and histone 3 lysine 36 (H3K36) demethylase activity (Fan et al., 2009). More importantly, methylated H3K36 is correlated with transcriptional elongation, thus binding of BCoR may lead to the recruitment of an H3K36 demethylase and subsequently transcriptional repression (Srinivasan et al., 2003). The role of the direct interaction with BCoR to AF9 is still unknown and future work in our lab by

my colleagues is geared towards elucidating the role of this direct interaction in MLL-AF9 leukemogenesis.

The role of CBX8, and its direct interaction with AF9, is discussed in (**Chapter 4**). Briefly, CBX8 is a member of the polycomb group (PcG) proteins that participate in gene silencing and transcriptional repression by modifying chromatin structure (Morey and Helin, 2010). The Polycomb repressive complex 2 catalyzes trimethylation of Histone 3 Lysine 27 (H3K27me3), which mediates transcriptional repression. Displacement of PcG proteins from promoters leads to a switch between H3K27 trimethylation and H3K27 acetylation suggesting that preventing H3K27 acetylation is a mechanism in which PcG proteins control transcription (Francis et al., 2004; Müller et al., 1995; Simon and Kingston, 2009).

1.4.4. Transcriptional activators, AF4 and DOT1L

The role of DOT1L is discussed in the next section. AF4 is a member of the ALF(AF4, LAF4, FMR2) family, is rich in serines and prolines and includes a nuclear targeting sequence (Erfurth et al., 2004; Ma and Staudt, 1996; Nakamura et al., 1993). As previously mentioned, AF4 is one of the most common MLL translocation partners. Homologous deletions of *Af4* in mice lead to impaired development of B and T-cells (Isnard et al., 2000). An autosomal dominant mutation in *Af4* leads to its accumulation in cerebellum purkinje cells and neurological disorders (Bitoun and Davies, 2005).

We recently solved an NMR solution structure of the C-terminal portion of AF4 in complex with the AF9 AHD (Leach et al., 2013). The role of AF4 in transcriptional elongation has been shown through its interaction with P-TEFb,

(Bitoun et al., 2007) and the crystal structure of P-TEFb with the N-terminal portion of AF5q31, a member of the AF4 family, was recently solved (Schulze-Gahmen et al., 2013). These studies structurally show a direct link between the recruitment of AF4 through the AF9 (AHD), and the subsequent recruitment of P-TEFb through the N-terminal domain of AF4.

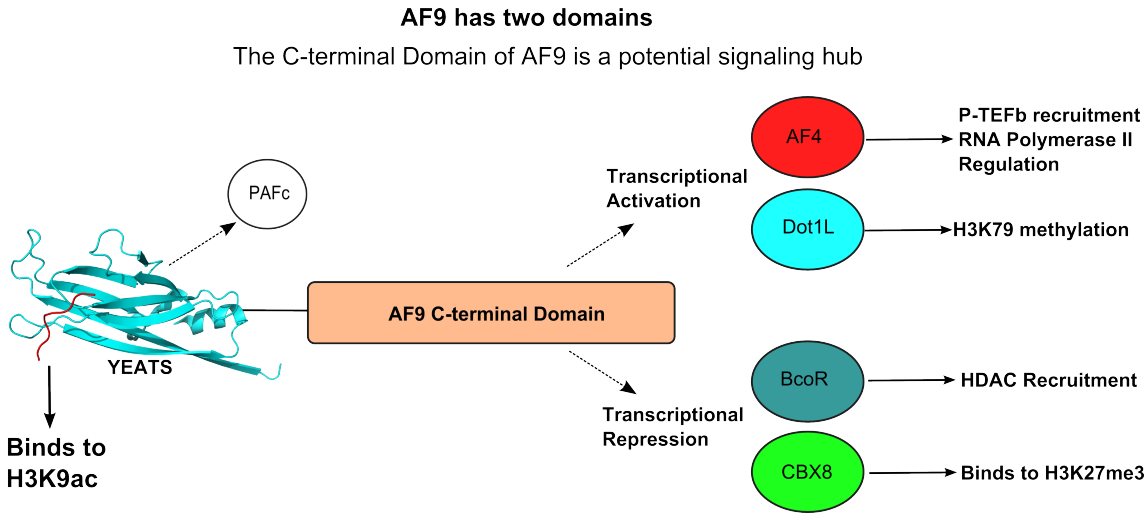


Figure 1-2. AF9 has an N-terminal YEATS and a C-terminal AHD domain. The crystal structure of the YEATS domain with acetylated histone H3K9 is shown (PDB: 4TMP). The AF9 (AHD) binds to two proteins in involved in transcriptional activation and two involved in transcriptional repression.

1.5. DOT1L is a histone H3 lysine 79 (H3K79) methyltransferase

Disruptor of telomeric silencing (Dot1) is the only known histone H3 lysine 79 methyltransferase and is highly conserved. Dot1 was originally discovered in *S. cerevisiae* as a regulator of telomeric silencing, yet its role was conflicting as

both overexpression and deletion of Dot1 led to losses in telomeric silencing (Feng et al., 2002; Ng et al., 2003; Singer et al., 1998). Nonetheless, Dot1L is essential in early development, as mouse embryos carrying a homologous Dot1L deletion do not survive (Chang et al., 2010; Jo et al., 2011).

The mammalian DOT1L (Disruptor of telomeric silencing like) protein is a large 1739 amino acid protein where the N-terminal portion of the protein (aa. 1-416) contains the methyltransferase domain. The DOT1L C-terminal domain has been implicated in protein-protein interactions with MLL-fusion partners AF10 and AF9 (Nguyen and Zhang, 2011), has been shown to directly interact with RNA polymerase II (Kim et al., 2012) and studies suggest that it preferentially associates with Wnt pathways (Gibbons et al., 2015; Mohan et al., 2010).

DOT1L is the only known enzyme to catalyze the mono-, di-, and tri-methylation of the globular domain of histone H3 at lysine 79 (H3K79me1, H3K79me2, H3K79me3) (Feng et al., 2002). Knockout of Dot1 results in complete losses in H3K79 methylation in most organisms (Jones et al., 2008; van Leeuwen et al., 2002; Shanower et al., 2005). Structurally, H3K79 lies between the H3/H4 tetramer and the H2A/H2B dimer and is solvent accessible (Luger et al., 1997). H3K79 methylation is enriched in gene coding regions (Kouskouti and Talianidis, 2005) and appears to be dependent on monubiquitiniation of histone H2B, lysine 120 (H2B-K120) through the ubiquitin-conjugating enzyme, Rad6, and its E3 ubiquitin ligase Bre1 (Robzyk et al., 2000; Sun and Allis, 2002; Wood et al., 2003). This suggests a level of histone cross-talk between H3K79 and H2B-K120. The mechanisms are unclear but it is

possible that there are other proteins such as COMPASS complex that interact with both H2B-K120 and DOT1L, or that the chromatin architecture changes based on this H2B-K120 ubiquitination to allow for H3K79 methylation (Nguyen and Zhang, 2011).

DOT1L-mediated H3K79 methylation marks are coupled with gene transcription (Steger et al., 2008). A number of studies have conducted genome-wide profile analyses in different *Drosophila*, mice and human cell lines that support a role for H3K79 methylation in active transcription (Nguyen and Zhang, 2011; Schübeler et al., 2004; Steger et al., 2008).

1.5.1. Physiological Roles of DOT1L

DOT1L plays a role in many different cellular processes from cell cycle regulation to differentiation and has been implicated in leukemogenesis, kidney injury, and cardiac disorders (Nguyen and Zhang, 2011). DOT1L knockout embryos show heart enlargement, reduced red blood cells and other phenotypes associated with deficiencies in the cardiovascular system (Jones et al., 2008). A novel role for DOT1L was shown in mice where loss of function led to dilated cardiomyopathy, congestive heart failure and lethality in adult mice. This study showed that loss of DOT1L mediated H3K79 methylation led to the down-regulation of dystrophin transcription. Loss of dystrophin causes misformation of the Dystrophin-glycoprotein complex which is important for the stability and viability of cardiomyocytes (Nguyen et al., 2011a).

DOT1 is believed to play a role in responses to DNA damage check point function and repair. The human 53BP1 protein is recruited to double stranded

breaks in DNA and can also bind to methylated H3K79. Mutations of this protein that prevent binding to H3K79, or mutations in Dot1, lead to the abrogation of 53BP1 recruitment to DNA double stranded breaks (Huyen et al., 2004).

Not all of the functions of this enzyme are completely understood. In mice, Dot1 has five splice variants, denoted as Dot1 a-e. Puzzlingly, Dot1a been shown to be directly involved in the repression of epithelial sodium channels (ENaC). Dot1a directly interacts with AF9, which leads to H3K79 methylation and repression of ENaC transcription. This study further shows that that aldosterone relieves Dot1a-AF9-mediated repression by downregulating expression of Dot1a and AF9 (Zhang et al., 2006).

1.5.2. DOT1L is a nonprocessive enzyme

Generally histone methyltransferases contain SET-domains that are able add up to three methyl groups to its lysine substrate. These histone methyltransferases kinetically carry out their activities through a processive mechanism (Dirk et al., 2007; Patnaik et al., 2004). Processive enzymes, such as DNA polymerase, stay attached to its substrate, where as nonprocessive enzymes, dissociate and reassociate. Unlike all other known lysine methyltransferases, DOT1L is unique in that the enzymatic region does not contain a SET domain, and the DOT1L structure is similar to that of arginine, not lysine, methyltransferases. The enzymatic region of DOT1L is composed of an α helical domain of an open α/β structure, including a 7-stranded β sheet, a characteristic fold found in arginine methylases (**Figure 1-3**) (Min et al., 2003; Nguyen and Zhang, 2011).

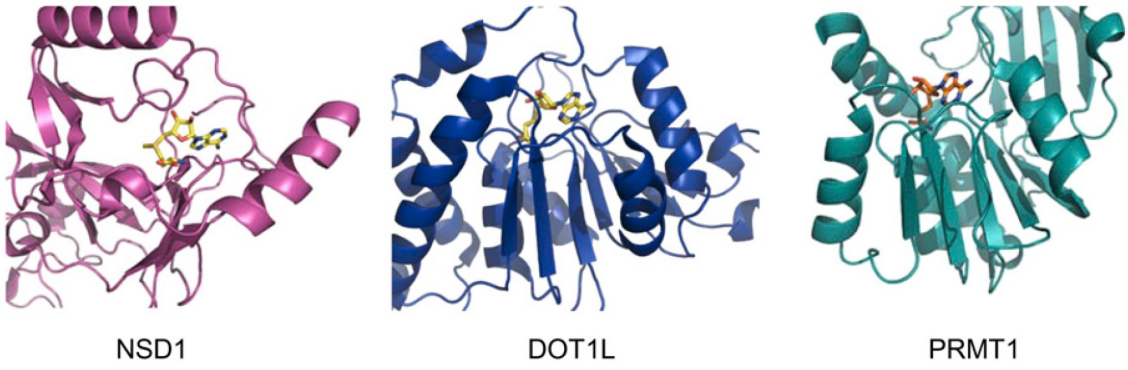


Figure 1-3. The catalytic region of DOT1L is most similar to an arginine methyltransferase.

Crystal structures of NSD1 (PDB: 3OOI), a lysine methyltransferase, DOT1L (PDB: 1NW3) and PRMT1 (PDB: 1ORI), an arginine methyltransferase. This figure was adapted from Nguyen et al. Figure 1-B (Nguyen and Zhang, 2011).

Frederiks *et al*. (2008) show that the relative levels of each mono- di- and tri- methyl mark are dependent upon the concentration of Dot1. At low Dot1 concentrations in chromatin isolated from yeast, they only observe the appearance of the H3K79me1 mark. As they increase the concentrations of Dot1, they then observe accumulations of H3K79me2 followed by H3K79me3. Therefore, they show that Dot1 catalyzes the mono- di- and tri- methylation of H3K79 in a distributive or nonprocessive manner. If this enzyme were processive, increasing concentrations of Dot1 would lead to a pattern where only a single methylation state would appear, and the particular methylation state that appears would be dependent on the time of incubation of Dot1 with chromatin (Frederiks et al., 2008).

1.5.3. DOT1L requirement for MLL-fusion leukemias

The first evidence for the role of DOT1L in MLL-fusion leukemias occurred when Okada *et al*. (2005) used yeast-two-hybrid studies to broadly map the interaction of the AF10 octapeptide-leucine zipper motif (OM-LZ) with DOT1L. They showed that this protein-protein interaction was necessary for MLL-AF10 transformation (Okada et al., 2005). Other groups have demonstrated the requirement of DOT1L for transformation by MLL-AF4 and MLL-ENL leukemias (Krivtsov et al., 2008; Mueller et al., 2007).

A number of *in vitro, in vivo*, and small molecule inhibitor studies have recently shown that DOT1L is essential for MLL-AF9 leukemia (Bernt et al., 2011; Chang et al., 2010; Daigle et al., 2011; Jo et al., 2011; Nguyen et al., 2011). Furthermore, the occupancy of the MLL-AF9 fusion protein is correlated with elevated H3K79me2 levels at target genes (Bernt et al., 2011; Nguyen et al., 2011b). MLL-fusion partners interact directly or indirectly with DOT1L resulting in aberrant H3K79 methylation at MLL-target genes such as the *HoxA* cluster and *Meis1.* Hypermethylation of H3K79 at these specific genes leads to increases in gene expression followed by leukemic transformation. While these studies show that DOT1L plays a major role in MLL-rearranged leukemogenesis, the exact mechanisms are unclear.

A small-molecule inhibitor, EPZ-5676, was recently developed that targets the enzymatic region of DOT1L, thereby inhibiting its activity. EPZ-5676 leads to losses of H3K79me2 at MLL-fusion loci and has shown efficacy in mouse models of MLL-fusion leukemia (Daigle et al., 2013). A recent study has found that

DOT1L blocks SIRT1 mediated epigenetic silencing and that this blockage maintains expression of MLL-fusion target genes. This study by Armstrong and colleagues show that this DOT1L recruitment maintains an “open chromatin state” in which MLL-fusion target genes show high levels of H3K79 methylation, H3K9 acetylation and H4K16 acetylation marks and low repressive H3K9 methylation levels. A SIRT1 agonist enhances the sensitivity of DOT1L inhibition in several MLL-fusion cell lines, suggesting this type of combination therapy could be effective (Chen et al., 2015).

1.6. Many proteins contain intrinsically disordered regions

For many years, it was thought that all proteins needed to adapt a three-dimensional structure in order to function. Recently, a number of studies have shown that a large portion of the genome contains proteins that do not necessarily adopt such a defined structure, but are critical for cellular functionality (Dyson and Wright, 2005). These proteins are referred to as intrinsically disordered proteins (IDPs). IDPs lack a unique 3D structure either entirely or exist as separate regions in larger structured proteins. The structure of IDPs resembles that of the denatured states of structured proteins. Essentially they are an ensemble of structures that interconvert at fast timescales (Tompa, 2005). As most structural studies have been reliant on proteins that have been crystallized, scientists did not believe in the existence of IDPs as the nature of IDPs prevent them from being analyzed as static crystal structures. Due to the advancement of other biophysical techniques from solution state NMR to molecular dynamicsover the last 15 years, they have gained widespread acceptance in the scientific

community and are only now beginning to be understood (reviewed in Dyson and Wright, 2005; Wright and Dyson, 2014). The frequency of IDPs appears to increase with increasing complexity of the organisms. More than a third of eukaryotic proteins contain regions of intrinsic disorder that are over 30 residues in length (Ward et al., 2004).

1.6.1 Intrinsically disordered proteins play a critical role in different biological systems

IDPs play many different biological roles such as the organization of chromatin and ribosomes, transport through the nuclear pore, and are critical in transcription, translation (Iakoucheva et al., 2002), and the cell-cycle (Galea et al., 2008; Guharoy et al., 2013; Tompa, 2005; Wright and Dyson, 2014). These proteins are often in important cellular processes as the majority of transcription factors contain disordered regions (Liu et al., 2006). As these IDPs are prevalent in many important proteins, they are tightly regulated (Babu et al., 2011). Underexpression of the IDP p27 has been linked with various types of cancers (Grimmler et al., 2007). Misfolding of IDPs such as α-synuclein and tau lead to neurological disorders (Breydo et al., 2012; Mukrasch et al., 2009).

1.6.2. The importance of intrinsically disordered proteins in signal transduction pathways

IDPs function as, or interact with, hubs in signal transduction pathways. Hub proteins tend to be more disordered than other proteins within the proteome (Kim et al., 2008). The disordered regions of proteins involved in signaling pathways contain short linear binding motifs that are conserved and are often

regulated through post-translational modifications. These motifs are located within intrinsically disordered regions in proteins, and proteomics studies have suggested that over 100,000 of these motifs exist (Tompa et al., 2014). The advantage of these motifs are that the same amino acid sequence within a protein can used to cause different cellular responses by turning on or off different pathways due their ability to be post-transitionally regulated (Dyson and Wright, 2002). Furthermore, the unique ability to bind to different targets via these motifs allows IDPs to play an integral part in the dynamic assembly of higher-order complexes and scaffolds. For example, the CREB-binding protein contains a nuclear co-activator-binding domain that is disordered when not bound to a partner, but forms two different helical structures with different topologies when bound to multiple proteins such as interferon regulatory factor 3 (Qin et al., 2005) or p160 nuclear receptor coactivators (Demarest et al., 2002; Waters et al., 2006). Another example is that the disordered region near the C-terminus of p53 can form helical, β-strand or extended structures upon binding to their protein partners (**Figure 1-4**) (Dunker et al., 2008). Interaction with a wide variety of targets appears to be a major characteristic of functional disordered proteins within signaling pathways.

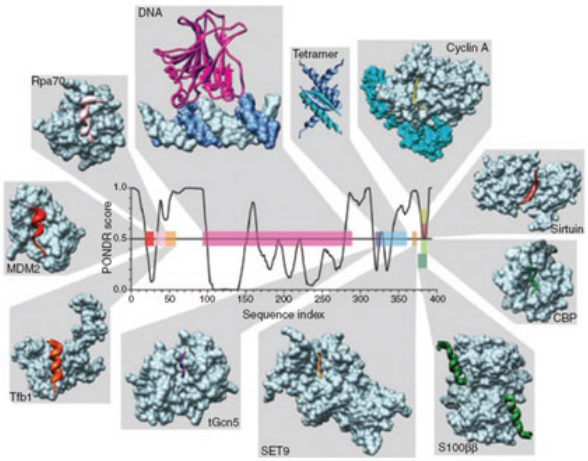


Figure 1-4. IDPs can form different structures upon binding.

The disordered region near the C-terminus of p53 forms different structures upon binding. Adapted from Figure 2b (Fuxreiter et al., 2008).

1.6.3. The biophysical mechanisms of IDP interactions

It is still difficult to understand why IDPs bind to other proteins when these interactions have a highly unfavorable entropic contribution that needs to be driven by a very large enthalpic contribution. Coupled folding and binding is the broad term used to denote the binding events involving intrinsically disordered proteins (Dyson and Wright, 2002). Essentially there are two proposed mechanisms used to explain the binding of IDPs.

The first proposed mechanism is that the disordered protein can find its target, bind and subsequently fold, otherwise known as induced folding. In this case a transient and weakly bound encounter complex is formed prior to the formation of a structural complex. Sugase, et al. showed that the phosphorylated kinase-inducible domain (pKID) of cyclic AMP binds to the KIX interaction domain

of CBP through an induced folding mechanism. In this study they show that pKID binds in a disordered state and subsequently folds on the surface of KIX. They use 2D NMR titration experiments and NMR relaxation dispersion experiments to demonstrate the formation of an encounter complex (Sugase et al., 2007)

The second mechanism, known as conformation selection, suggests that IDPs exist as a varying ensemble of states that constitute different degrees of protein folding. The target protein then recognizes one of these transient states. An example of this mechanism is the existence of pre-formed helical structures in the linker region of p27, a cyclin-dependent kinase and an IDP. The p27 protein plays a role in cell-cycle regulation through binding and inhibition of Cdk2/cyclin A and Cdk2/cyclin E resulting in cell cycle arrest at the G_1 to S transition. Mutations that prevented the formation of these nascent helicies also blocked the binding of p27, which subsequently lost its ability to inhibit the Cdk2/cyclin complexes (Otieno and Kriwacki, 2012).

Both of these mechanisms lie on the opposite ends of the spectrum and it is most likely that proteins use some sort of combination of both of proposed approaches. Indeed many examples have been seen in which IDPs remain disordered even upon binding to their target protein, denoted as “fuzzy” complexes (Fuxreiter et al., 2008; Tompa and Fuxreiter, 2008).

1.7. Preliminary Data: The AF9 (AHD) is an intrinsically disordered protein[1]

Our initial goal was to solve the structure of the C-terminal domain of AF9, known as the AF9 (AHD), and conduct structure-function studies. We expressed the C-terminal 79 amino acids of AF9 that have been shown to be required for oncogenic activity of an MLL-AF9 fusion protein in colony-forming assays (Prasad et al., 1995), and roughly corresponds to the minimal domain observed in a clinical case of leukemia with an MLL-AF9 translocation (Mitterbaur et al., 1999). Our initial efforts to express and purify the AF9 (AHD) were extremely unsuccessful. The protein had limited solubility and was prone to degradation as was observed on SDS-PAGE gels and Size Exclusion Chromatography. It was difficult to get this protein to high enough concentrations to conduct NMR studies. Even at minimal protein concentrations (50μM) 2D ^{15}N-^{1}H heteronuclear single quantum coherence spectroscopy (HSQC) spectra showed extreme broadening, with an insufficient number of peaks for the amino acids in the domain (Figure 1. A). As the AF4 protein interaction with AF9 had been finely mapped, we ordered an AF4 peptide that comprised of the AF9 interacting regions (aa. 761-774) and titrated this unlabeled peptide into a ^{15}N labeled AF9 NMR sample. Intriguingly, we observed improved chemical shift dispersion and the appearance of AF9 resonance peaks that were previously unseen (**Figure 1-5B**). Further CD spectroscopy experiments showed that AF9 was a random coil with some residual secondary structure, but when titrated with the AF4 peptide, this led to the formation of a mixed alpha-beta structure (**Figure 1-5C**). Thus, the data show

[1] The experiments in this section were performed in our lab by Ben Leach (Leach et al., 2013). (Leach et al., 2013)

that the AF9 (AHD) is intrinsically disordered and undergoes coupled folding and binding to form a structured AF4-AF9 complex. It is presumable to believe that the AF9 (AHD) behaves in a similar manner with each of its other binding partners.

1.8. Thesis summary

The fact that DOT1L interacts with transcriptional elongation complexes consisting of MLL fusion partner proteins suggests that the recruitment of DOT1L is an important common mechanism central to MLL-fusion leukemias (Biswas et al., 2011; Mohan et al., 2010; Nguyen and Zhang, 2011; Okada et al., 2005; Yokoyama et al., 2010). The protein-protein interaction between AF9 and DOT1L has been roughly mapped (Biswas et al., 2011; Shen et al., 2013; Yokoyama et al., 2010; Zhang et al., 2006), but there is a lack of structural characterization of this interaction and functional effects of the direct recruitment of DOT1L to MLL-AF9. To that end, herein we show that there are three separate regions in DOT1L that interact with AF9 and fold into structurally similar complexes. We also present the first structure of a DOT1L-AF9 complex (**Chapter 2**). Using structure-guided mutagenesis, we developed point mutations that reduce DOT1L binding to AF9 in a graded manner. Functional characterization of these point mutations in the context of MLL-AF9 shows that the degree of DOT1L recruitment to the MLL-AF9 fusion protein differentially affects H3K79me2 and H3K79me3 levels at specific target genes and that direct recruitment of DOT1L is essential for the transforming potential of MLL-AF9 (**Chapter 3**).

Lastly, we discuss the direct interaction between CBX8 and AF9 and present the NMR solution structure of the CBX8-AF9 complex (**Chapter 4**). We show that AF9 forms nearly identical structures with each of its binding partners. Based on this CBX8-AF9 structure, we propose several biological experiments that will help to further elucidate the role of the CBX8 interaction with MLL-AF9 in leukemogenesis.

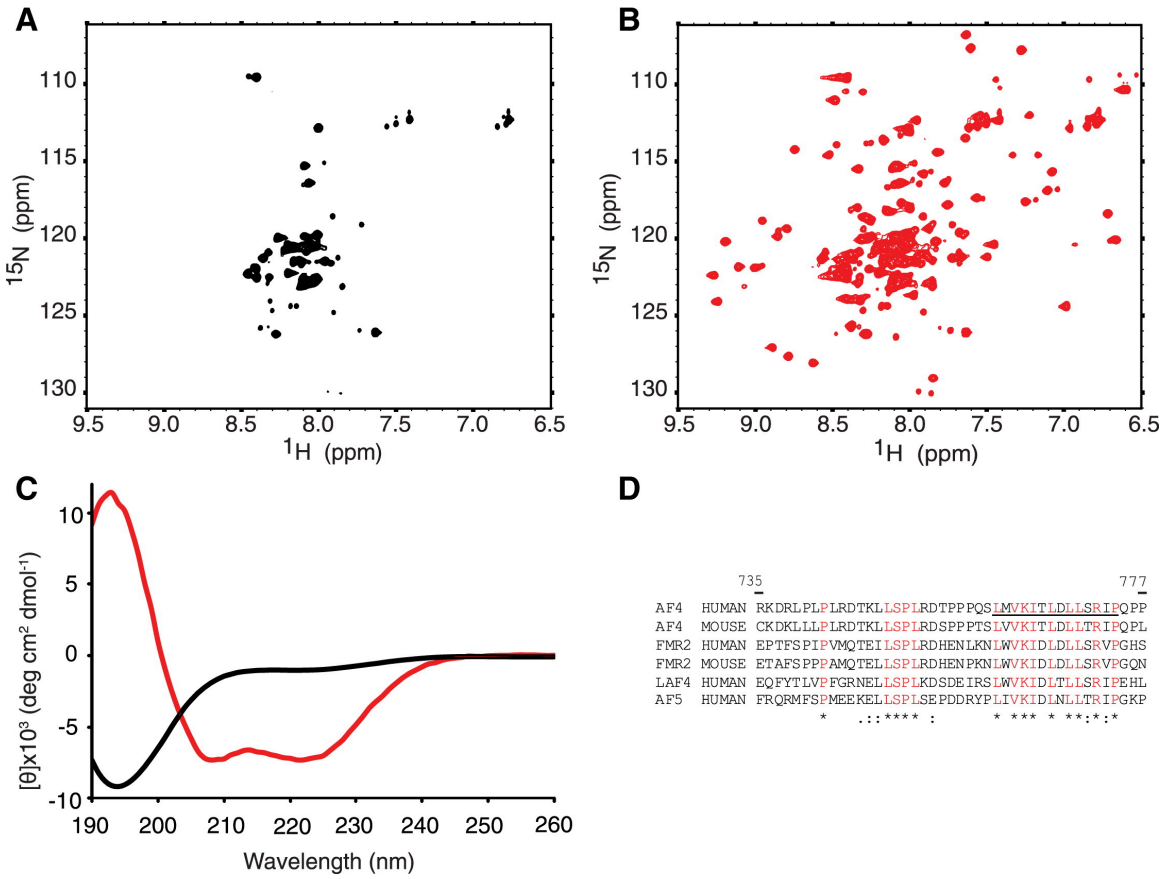


Figure 1-5. AF9 (AHD) is an intrinsically disordered protein[2]

A) ^{15}N-^{1}H HSQC of AF9 (AHD) shows limited peak dispersion. **B)** Titration of AF4 peptide into ^{15}N AF9 (AHD) leads to dispersion of resonance peaks. **C)** AF9 (AHD) is a random coil but becomes a missed alpha-beta structure upon titration of AF4. **D)** Conservation of the AF4 interacting site with AF9.

[2] This figure was directly taken from Figure 1 in (Leach et al., 2013).

1.9. References for Chapter 1

Ahn, S.H., Kim, M., and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. Mol. Cell 13, 67–76.

Ayton, P.M., and Cleary, M.L. (2003). Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9. Genes Dev. 17, 2298–2307.

Ayton, P.M., Chen, E.H., and Cleary, M.L. (2004). Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. Mol. Cell. Biol. 24, 10470–10478.

Babu, M.M., van der Lee, R., de Groot, N.S., and Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. Curr. Opin. Struct. Biol. 21, 432–440.

Bernt, K.M., Zhu, N., Sinha, A.U., Vempati, S., Faber, J., Krivtsov, A. V, Feng, Z., Punt, N., Daigle, A., Bullinger, L., et al. (2011). MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. Cancer Cell 20, 66–78.

Biswas, D., Milne, T. a, Basrur, V., Kim, J., Elenitoba-Johnson, K.S.J., Allis, C.D., and Roeder, R.G. (2011). Function of leukemogenic mixed lineage leukemia 1

(MLL) fusion proteins through distinct partner protein complexes. Proc. Natl. Acad. Sci. U. S. A. 108, 15751–15756.

Bitoun, E., and Davies, K.E. (2005). The robotic mouse: unravelling the function of AF4 in the cerebellum. Cerebellum 4, 250–260.

Bitoun, E., Oliver, P.L., and Davies, K.E. (2007). The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. Hum. Mol. Genet. 16, 92–106.

Borkin, D., He, S., Miao, H., Kempinska, K., Pollock, J., Chase, J., Purohit, T., Malik, B., Zhao, T., Wang, J., et al. (2015). Pharmacologic Inhibition of the Menin-MLL Interaction Blocks Progression of MLL Leukemia In Vivo. Cancer Cell.

De Braekeleer, E., Meyer, C., Douet-Guilbert, N., Morel, F., Le Bris, M.-J., Berthou, C., Arnaud, B., Marschalek, R., Férec, C., and De Braekeleer, M. (2010). Complex and cryptic chromosomal rearrangements involving the MLL gene in acute leukemia: A study of 7 patients and review of the literature. Blood Cells, Mol. Dis. 44, 268–274.

Breydo, L., Wu, J.W., and Uversky, V.N. (2012). α-Synuclein misfolding and Parkinson’s disease. Biochim. Biophys. Acta - Mol. Basis Dis. 1822, 261–285.

Bursen, A., Schwabe, K., Rüster, B., Henschler, R., Ruthardt, M., Dingermann, T., and Marschalek, R. (2010). The AF4.MLL fusion protein is capable of inducing ALL in mice without requirement of MLL.AF4. Blood 115, 3570–3579.

Cairns, B., Henry, N., and Kornberg, R. (1996). TFG/TAF30/ANC1, a component of the yeast SWI/SNF complex that is similar to the leukemogenic proteins ENL and AF-9. Mol. Cell. Biol. 16, 3308–3316.

Chang, M.-J., Wu, H., Achille, N.J., Reisenauer, M.R., Chou, C.-W., Zeleznik-Le, N.J., Hemenway, C.S., and Zhang, W. (2010). Histone H3 lysine 79 methyltransferase Dot1 is required for immortalization by MLL oncogenes. Cancer Res. 70, 10234–10242.

Chen, C.-W., Koche, R.P., Sinha, A.U., Deshpande, A.J., Zhu, N., Eng, R., Doench, J.G., Xu, H., Chu, S.H., Qi, J., et al. (2015). DOT1L inhibits SIRT1-mediated epigenetic silencing to maintain leukemic gene expression in MLL-rearranged leukemia. Nat. Med. 21, 335–343.

Cierpicki, T., Risner, L.E., Grembecka, J., Lukasik, S.M., Popovic, R., Omonkowska, M., Shultis, D.D., Zeleznik-Le, N.J., and Bushweller, J.H. (2010). Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. Nat. Struct. Mol. Biol. 17, 62–68.

Collins, E.C., Pannell, R., Simpson, E.M., Forster, A., and Rabbitts, T.H. (2000). Inter-chromosomal recombination of Mll and Af9 genes mediated by cre-loxP in mouse development. EMBO Rep. 1, 127–132.

Collins, E.C., Appert, A., Ariza-McNaughton, L., Pannell, R., Yamada, Y., and Rabbitts, T.H. (2002). Mouse Af9 is a controller of embryo patterning, like Mll, whose human homologue fuses with Af9 after chromosomal translocation in leukemia. Mol. Cell. Biol. 22, 7313–7324.

Cox, M.C., Panetta, P., Lo-Coco, F., Del Poeta, G., Venditti, A., Maurillo, L., Del Principe, M.I., Mauriello, A., Anemona, L., Bruno, A., et al. (2004). Chromosomal aberration of the 11q23 locus in acute leukemia and frequency of MLL gene translocation: results in 378 adult patients. Am. J. Clin. Pathol. 122, 298–306.

Daigle, S.R., Olhava, E.J., Therkelsen, C.A., Majer, C.R., Sneeringer, C.J., Song, J., Johnston, L.D., Scott, M.P., Smith, J.J., Xiao, Y., et al. (2011). Selective Killing of Mixed Lineage Leukemia Cells by a Potent Small-Molecule DOT1L Inhibitor. Cancer Cell 20, 53–65.

Daigle, S.R., Olhava, E.J., Therkelsen, C.A., Basavapathruni, A., Jin, L., Boriack-Sjodin, P.A., Allain, C.J., Klaus, C.R., Raimondi, A., Scott, M.P., et al. (2013). Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. Blood 122, 1017–1025.

Demarest, S.J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H.J., Evans, R.M., and Wright, P.E. (2002). Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. Nature 415, 549–553.

Deshpande, A.J., Chen, L., Fazio, M., Sinha, A.U., Bernt, K.M., Banka, D., Dias, S., Chang, J., Olhava, E.J., Daigle, S.R., et al. (2013). Leukemic transformation by the MLL-AF6 fusion oncogene requires the H3K79 methyltransferase Dot1l. Blood 121, 2533–2541.

Dimartino, J.F., and Cleary, M.L. (1999). Mll rearrangements in haematological malignancies: lessons from clinical and biological studies. Br. J. Haematol. 106, 614–626.

Dirk, L.M.A., Flynn, E.M., Dietzel, K., Couture, J.-F., Trievel, R.C., and Houtz, R.L. (2007). Kinetic manifestation of processivity during multiple methylations catalyzed by SET domain protein methyltransferases. Biochemistry 46, 3905–3915.

Dunker, A.K., Oldfield, C.J., Meng, J., Romero, P., Yang, J.Y., Chen, J.W., Vacic, V., Obradovic, Z., and Uversky, V.N. (2008). The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics 9 Suppl 2, S1.

Dyson, H.J., and Wright, P.E. (2002). Coupling of folding and binding for unstructured proteins. Curr. Opin. Struct. Biol. 12, 54–60.

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6, 197–208.

Erfurth, F., Hemenway, C.S., de Erkenez, A.C., and Domer, P.H. (2004). MLL fusion partners AF4 and AF9 interact at subnuclear foci. Leukemia 18, 92–102.

Erfurth, F.E., Popovic, R., Grembecka, J., Cierpicki, T., Theisler, C., Xia, Z.-B., Stuart, T., Diaz, M.O., Bushweller, J.H., and Zeleznik-Le, N.J. (2008). MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression. Proc. Natl. Acad. Sci. U. S. A. 105, 7517–7522.

Ernst, P., Wang, J., Huang, M., Goodman, R.H., and Korsmeyer, S.J. (2001). MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. Mol. Cell. Biol. 21, 2249–2258.

Fan, Z., Yamaza, T., Lee, J.S., Yu, J., Wang, S., Fan, G., Shi, S., and Wang, C.-Y. (2009). BCOR regulates mesenchymal stem cell function by epigenetic mechanisms. Nat. Cell Biol. 11, 1002–1009.

Feng, Q., Wang, H., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002). Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain. Curr. Biol. 12, 1052–1058.

Ford, A.M., Ridge, S.A., Cabrera, M.E., Mahmoud, H., Steel, C.M., Chan, L.C., and Greaves, M. (1993). In utero rearrangements in the trithorax-related oncogene in infant leukaemias. Nature 363, 358–360.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science 306, 1574–1577.

Frederiks, F., Tzouros, M., Oudgenoeg, G., van Welsem, T., Fornerod, M., Krijgsveld, J., and van Leeuwen, F. (2008). Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states. Nat. Struct. Mol. Biol. 15, 550–557.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature 461, 186–192.

Fuxreiter, M., Tompa, P., Simon, I., Uversky, V.N., Hansen, J.C., and Asturias, F.J. (2008). Malleable machines take shape in eukaryotic transcriptional regulation. Nat. Chem. Biol. 4, 728–737.

Galea, C.A., Wang, Y., Sivakolundu, S.G., and Kriwacki, R.W. (2008). Regulation of Cell Division by Intrinsically Unstructured Proteins: Intrinsic Flexibility, Modularity, and Signaling Conduits †. Biochemistry 47, 7598–7609.

Gibbons, G.S., Owens, S.R., Fearon, E.R., and Nikolovska-Coleska, Z. (2015). Regulation of Wnt signaling target gene expression by the histone methyltransferase DOT1L. ACS Chem. Biol. 10, 109–114.

Grimmler, M., Wang, Y., Mund, T., Cilenšek, Z., Keidel, E.-M., Waddell, M.B., Jäkel, H., Kullmann, M., Kriwacki, R.W., and Hengst, L. (2007). Cdk-Inhibitory Activity and Stability of p27Kip1 Are Directly Regulated by Oncogenic Tyrosine Kinases. Cell 128, 269–280.

Guenther, M.G., Jenner, R.G., Chevalier, B., Nakamura, T., Croce, C.M., Canaani, E., and Young, R.A. (2005). Global and Hox-specific roles for the MLL1 methyltransferase. Proc. Natl. Acad. Sci. U. S. A. 102, 8603–8608.

Guharoy, M., Szabo, B., Contreras Martos, S., Kosol, S., and Tompa, P. (2013). Intrinsic structural disorder in cytoskeletal proteins. Cytoskeleton (Hoboken). 70, 550–571.

He, N., Chan, C.K., Sobhian, B., Chou, S., Xue, Y., Liu, M., Alber, T., Benkirane, M., and Zhou, Q. (2011). Human Polymerase-Associated Factor complex (PAFc)

connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin. Proc. Natl. Acad. Sci. U. S. A. 108, E636–45.

Hemenway, C.S., de Erkenez, A.C., and Gould, G.C. (2001). The polycomb protein MPc3 interacts with AF9, an MLL fusion partner in t(9;11)(p22;q23) acute leukemias. Oncogene 20, 3798–3805.

Huyen, Y., Zgheib, O., Ditullio, R.A., Gorgoulis, V.G., Zacharatos, P., Petty, T.J., Sheston, E.A., Mellert, H.S., Stavridi, E.S., and Halazonetis, T.D. (2004). Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature 432, 406–411.

Huynh, K.D., Fischle, W., Verdin, E., and Bardwell, V.J. (2000). BCoR, a novel corepressor involved in BCL-6 repression. Genes Dev. 14, 1810–1823.

Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradović, Z., and Dunker, A.K. (2002). Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. J. Mol. Biol. 323, 573–584.

Isnard, P., Coré, N., Naquet, P., and Djabali, M. (2000). Altered lymphoid development in mice deficient for the mAF4 proto-oncogene. Blood 96, 705–710.

Jo, S.Y., Granowicz, E.M., Maillard, I., Thomas, D., and Hess, J.L. (2011). Requirement for Dot1l in murine postnatal hematopoiesis and leukemogenesis by MLL translocation. Blood 117, 4759–4768.

Jones, B., Su, H., Bhat, A., Lei, H., Bajko, J., Hevi, S., Baltus, G.A., Kadam, S., Zhai, H., Valdez, R., et al. (2008). The Histone H3K79 Methyltransferase Dot1L Is Essential for Mammalian Development and Heterochromatin Structure. PLoS Genet. 4, e1000190.

Kim, P.M., Sboner, A., Xia, Y., and Gerstein, M. (2008). The role of disorder in interaction networks: a structural analysis. Mol. Syst. Biol. 4, 179.

Kim, S.-K., Jung, I., Lee, H., Kang, K., Kim, M., Jeong, K., Kwon, C.S., Han, Y.-M., Kim, Y.S., Kim, D., et al. (2012). Human histone H3K79 methyltransferase DOT1L protein [corrected] binds actively transcribing RNA polymerase II to regulate gene expression. J. Biol. Chem. 287, 39698–39709.

Kouskouti, A., and Talianidis, I. (2005). Histone modifications defining active genes persist after transcriptional and mitotic inactivation. EMBO J. 24, 347–357.

Krivtsov, A. V, and Armstrong, S.A. (2007). MLL translocations, histone modifications and leukaemia stem-cell development. Nat. Rev. Cancer 7, 823–833.

Krivtsov, A. V, Feng, Z., Lemieux, M.E., Faber, J., Vempati, S., Sinha, A.U., Xia, X., Jesneck, J., Bracken, A.P., Silverman, L.B., et al. (2008). H3K79 methylation profiles define murine and human MLL-AF4 leukemias. Cancer Cell 14, 355–368.

Leach, B.I., Kuntimaddi, A., Schmidt, C.R., Cierpicki, T., Johnson, S. a, and Bushweller, J.H. (2013). Leukemia fusion target AF9 is an intrinsically disordered transcriptional regulator that recruits multiple partners via coupled folding and binding. Structure 21, 176–183.

Van Leeuwen, F., Gafken, P.R., and Gottschling, D.E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell 109, 745–756.

Li, Y., Wen, H., Xi, Y., Tanaka, K., Wang, H., Peng, D., Ren, Y., Jin, Q., Dent, S.Y.R., Li, W., et al. (2014). AF9 YEATS Domain Links Histone Acetylation to DOT1L-Mediated H3K79 Methylation. Cell 159, 558–571.

Lin, C., Smith, E.R., Takahashi, H., Lai, K.C., Martin-Brown, S., Florens, L., Washburn, M.P., Conaway, J.W., Conaway, R.C., and Shilatifard, A. (2010). AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. Mol. Cell 37, 429–437.

Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, A.K. (2006). Intrinsic disorder in transcription factors. Biochemistry 45, 6873–6888.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251–260.

Ma, C., and Staudt, L.M. (1996). LAF-4 encodes a lymphoid nuclear protein with transactivation potential that is homologous to AF-4, the gene fused to MLL in t(4;11) leukemias. Blood 87, 734–745.

Meyer, C., Hofmann, J., Burmeister, T., Gröger, D., Park, T.S., Emerenciano, M., Pombo de Oliveira, M., Renneville, A., Villarese, P., Macintyre, E., et al. (2013). The MLL recombinome of acute leukemias in 2013. Leukemia 27, 2165–2176.

Milne, T.A., Kim, J., Wang, G.G., Stadler, S.C., Basrur, V., Whitcomb, S.J., Wang, Z., Ruthenburg, A.J., Elenitoba-Johnson, K.S.J., Roeder, R.G., et al. (2010). Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. Mol. Cell 38, 853–863.

Min, J., Feng, Q., Li, Z., Zhang, Y., and Xu, R.-M. (2003). Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. Cell 112, 711–723.

Mitterbaur, G., Zimmer, C., Fonatsch, C., Haas, O., Thalhammer-Scherrer, R., Schwarzinger, I., Kalhs, P., Jaeger, U., Lechner, K., and Mannhalter, C. (1999). Monitoring of minimal residual leukemia in patients with MLL-AF9 positive acute myeloid leukemia by RT-PCR. Leukemia 13, 1519–1524.

Mohan, M., Herz, H.-M., Takahashi, Y.-H., Lin, C., Lai, K.C., Zhang, Y., Washburn, M.P., Florens, L., and Shilatifard, A. (2010). Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). Genes Dev. 24, 574–589.

Morey, L., and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. Trends Biochem. Sci. 35, 323–332.

Mueller, D., Bach, C., Zeisig, D., Garcia-Cuellar, M.-P., Monroe, S., Sreekumar, A., Zhou, R., Nesvizhskii, A., Chinnaiyan, A., Hess, J.L., et al. (2007). A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. Blood 110, 4445–4454.

Mukrasch, M.D., Bibow, S., Korukottu, J., Jeganathan, S., Biernat, J., Griesinger, C., Mandelkow, E., and Zweckstetter, M. (2009). Structural polymorphism of 441-residue tau at single residue resolution. PLoS Biol. 7, e34.

Müller, J., Gaunt, S., and Lawrence, P.A. (1995). Function of the Polycomb protein is conserved in mice and flies. Development 121, 2847–2852.

Muntean, A.G., and Hess, J.L. (2012). The pathogenesis of mixed-lineage leukemia. Annu. Rev. Pathol. 7, 283–301.

Muntean, A.G., Tan, J., Sitwala, K., Huang, Y., Bronstein, J., Connelly, J.A., Basrur, V., Elenitoba-Johnson, K.S.J., and Hess, J.L. (2010). The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. Cancer Cell 17, 609–621.

Nakamura, T., Alder, H., Gu, Y., Prasad, R., Canaani, O., Kamada, N., Gale, R.P., Lange, B., Crist, W.M., and Nowell, P.C. (1993). Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs. Proc. Natl. Acad. Sci. U. S. A. 90, 4631–4635.

Ng, H.H., Ciccone, D.N., Morshead, K.B., Oettinger, M.A., and Struhl, K. (2003). Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation. Proc. Natl. Acad. Sci. U. S. A. 100, 1820–1825.

Nguyen, A.T., and Zhang, Y. (2011). The diverse functions of Dot1 and H3K79 methylation. Genes Dev. 25, 1345–1358.

Nguyen, A.T., Xiao, B., Neppl, R.L., Kallin, E.M., Li, J., Chen, T., Wang, D.-Z., Xiao, X., and Zhang, Y. (2011a). DOT1L regulates dystrophin expression and is critical for cardiac function. Genes Dev. 25, 263–274.

Nguyen, A.T., Taranova, O., He, J., and Zhang, Y. (2011b). DOT1L, the H3K79 methyltransferase, is required for MLL-AF9-mediated leukemogenesis. Blood 117, 6912–6922.

Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V.M., Su, L., Xu, G., and Zhang, Y. (2005). hDOT1L links histone methylation to leukemogenesis. Cell 121, 167–178.

Osada, S., Sutton, A., Muster, N., Brown, C.E., Yates, J.R., Sternglanz, R., and Workman, J.L. (2001). The yeast SAS (something about silencing) protein complex contains a MYST-type putative acetyltransferase and functions with chromatin assembly factor ASF1. Genes Dev. 15, 3155–3168.

Otieno, S., and Kriwacki, R. (2012). Probing the Role of Nascent Helicity in p27 Function as a Cell Cycle Regulator. PLoS One 7, e47177.

Park, S., Osmers, U., Raman, G., Schwantes, R.H., Diaz, M.O., and Bushweller, J.H. (2010). The PHD3 domain of MLL acts as a CYP33-regulated switch between MLL-mediated activation and repression . Biochemistry 49, 6576–6586.

Patnaik, D., Chin, H.G., Estève, P.-O., Benner, J., Jacobsen, S.E., and Pradhan, S. (2004). Substrate specificity and kinetic mechanism of mammalian G9a histone H3 methyltransferase. J. Biol. Chem. 279, 53248–53258.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol. Cell 23, 297–305.

Popovic, R., and Zeleznik-Le, N.J. (2005). MLL: how complex does it get? J. Cell. Biochem. 95, 234–242.

Prasad, R., Yano, T., Sorio, C., Nakamura, T., Rallapalli, R., Gu, Y., Leshkowitz, D., Croce, C.M., and Canaani, E. (1995). Domains with transcriptional regulatory activity within the ALL1 and AF4 proteins involved in acute leukemia. Proc. Natl. Acad. Sci. U. S. A. 92, 12160–12164.

Qin, B.Y., Liu, C., Srinath, H., Lam, S.S., Correia, J.J., Derynck, R., and Lin, K. (2005). Crystal structure of IRF-3 in complex with CBP. Structure 13, 1269–1277.

Rice, K.L., and Licht, J.D. (2007). HOX deregulation in acute myeloid leukemia. J. Clin. Invest. 117, 865–868.

Robzyk, K., Recht, J., and Osley, M.A. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. Science 287, 501–504.

Rougvie, A.E., and Lis, J.T. (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell 54, 795–804.

Schübeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J., et al. (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev. 18, 1263–1271.

Schulze, J.M., Wang, A.Y., and Kobor, M.S. (2009). YEATS domain proteins: a diverse family with many links to chromatin modification and transcription. Biochem. Cell Biol. 87, 65–75.

Schulze-Gahmen, U., Upton, H., Birnberg, A., Bao, K., Chou, S., Krogan, N.J., Zhou, Q., and Alber, T. (2013). The AFF4 scaffold binds human P-TEFb adjacent to HIV Tat. Elife 2, e00327.

Sedkov, Y., Tillib, S., Mizrokhi, L., and Mazo, A. (1994). The bithorax complex is regulated by trithorax earlier during Drosophila embryogenesis than is the Antennapedia complex, correlating with a bithorax-like expression pattern of distinct early trithorax transcripts. Development 120, 1907–1917.

Shanower, G.A., Muller, M., Blanton, J.L., Honti, V., Gyurkovics, H., and Schedl, P. (2005). Characterization of the grappa gene, the Drosophila histone H3 lysine 79 methyltransferase. Genetics 169, 173–184.

Shen, C., Jo, S.Y., Liao, C., Hess, J.L., and Nikolovska-Coleska, Z. (2013). Targeting recruitment of disruptor of telomeric silencing 1-like (DOT1L): characterizing the interactions between DOT1L and mixed lineage leukemia (MLL) fusion proteins. J. Biol. Chem. 288, 30585–30596.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat. Rev. Mol. Cell Biol. 10, 697–708.

Singer, M.S., Kahana, A., Wolf, A.J., Meisinger, L.L., Peterson, S.E., Goggin, C., Mahowald, M., and Gottschling, D.E. (1998). Identification of high-copy disruptors of telomeric silencing in Saccharomyces cerevisiae. Genetics 150, 613–632.

Slany, R.K., Lavau, C., and Cleary, M.L. (1998). The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX. Mol. Cell. Biol. 18, 122–129.

Sorensen, P.H., Chen, C.S., Smith, F.O., Arthur, D.C., Domer, P.H., Bernstein, I.D., Korsmeyer, S.J., Hammond, G.D., and Kersey, J.H. (1994). Molecular rearrangements of the MLL gene are present in most cases of infant acute

myeloid leukemia and are strongly correlated with monocytic or myelomonocytic phenotypes. J. Clin. Invest. 93, 429–437.

Srinivasan, R.S., de Erkenez, A.C., and Hemenway, C.S. (2003). The mixed lineage leukemia fusion partner AF9 binds specific isoforms of the BCL-6 corepressor. Oncogene 22, 3395–3406.

Steger, D.J., Lefterova, M.I., Ying, L., Stonestrom, A.J., Schupp, M., Zhuo, D., Vakoc, A.L., Kim, J.-E., Chen, J., Lazar, M.A., et al. (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Mol. Cell. Biol. 28, 2825–2839.

Sugase, K., Dyson, H.J., and Wright, P.E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447, 1021–1025.
Sun, Z.-W., and Allis, C.D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature 418, 104–108.

Super, H.J., McCabe, N.R., Thirman, M.J., Larson, R.A., Le Beau, M.M., Pedersen-Bjergaard, J., Philip, P., Diaz, M.O., and Rowley, J.D. (1993). Rearrangements of the MLL gene in therapy-related acute myeloid leukemia in patients previously treated with agents targeting DNA-topoisomerase II. Blood 82, 3705–3711.

Tan, J., Jones, M., Koseki, H., Nakayama, M., Muntean, A.G., Maillard, I., and Hess, J.L. (2011). CBX8, a Polycomb Group Protein, Is Essential for MLL-AF9-Induced Leukemogenesis. Cancer Cell 20, 563–575.

Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett. 579, 3346–3354.

Tompa, P., and Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem. Sci. 33, 2–8.

Tompa, P., Davey, N.E., Gibson, T.J., and Babu, M.M. (2014). A Million Peptide Motifs for the Molecular Biologist. Mol. Cell 55, 161–169.

Wang, Z., Song, J., Milne, T.A., Wang, G.G., Li, H., Allis, C.D., and Patel, D.J. (2010). Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. Cell 141, 1183–1194.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J. Mol. Biol. 337, 635–645.

Waters, L., Yue, B., Veverka, V., Renshaw, P., Bramham, J., Matsuda, S., Frenkiel, T., Kelly, G., Muskett, F., Carr, M., et al. (2006). Structural diversity in

p160/CREB-binding protein coactivator complexes. J. Biol. Chem. 281, 14787–14795.

Wong, P., Iwasaki, M., Somervaille, T.C.P., So, C.W.E., So, C.W.E., and Cleary, M.L. (2007). Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential. Genes Dev. 21, 2762–2774.

Wood, A., Krogan, N.J., Dover, J., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Golshani, A., Zhang, Y., Greenblatt, J.F., et al. (2003). Bre1, an E3 Ubiquitin Ligase Required for Recruitment and Substrate Selection of Rad6 at a Promoter. Mol. Cell 11, 267–274.

Wright, P.E., and Dyson, H.J. (2014). Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. 16, 18–29.

Wu, C.-H. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. Genes Dev. 17, 1402–1414.

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

Yokoyama, A., Somervaille, T.C.P., Smith, K.S., Rozenblatt-Rosen, O., Meyerson, M., and Cleary, M.L. (2005). The Menin Tumor Suppressor Protein Is an Essential Oncogenic Cofactor for MLL-Associated Leukemogenesis. Cell 123, 207–218.

Yokoyama, A., Lin, M., Naresh, A., Kitabayashi, I., and Cleary, M.L. (2010). A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. Cancer Cell 17, 198–212.

Yu, B.D., Hess, J.L., Horning, S.E., Brown, G.A., and Korsmeyer, S.J. (1995). Altered Hox expression and segmental identity in Mll-mutant mice. Nature 378, 505–508.

Yu, B.D., Hanson, R.D., Hess, J.L., Horning, S.E., and Korsmeyer, S.J. (1998). MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. Proc. Natl. Acad. Sci. U. S. A. 95, 10632–10636.

Yu, M., Honoki, K., Andersen, J., Paietta, E., Nam, D.K., and Yunis, J.J. (1996). MLL tandem duplication and multiple splicing in adult acute myeloid leukemia with normal karyotype. Leukemia 10, 774–780.

Zeleznik-Le, N.J., Harden, A.M., and Rowley, J.D. (1994). 11q23 translocations split the “AT-hook” cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. Proc. Natl. Acad. Sci. U. S. A. 91, 10610–10614.

Zhang, H., Richardson, D.O., Roberts, D.N., Utley, R., Erdjument-Bromage, H., Tempst, P., Côté, J., and Cairns, B.R. (2004). The Yaf9 component of the SWR1 and NuA4 complexes is required for proper gene expression, histone H4 acetylation, and Htz1 replacement near telomeres. Mol. Cell. Biol. 24, 9424–9436.

Zhang, W., Xia, X., Reisenauer, M.R., Hemenway, C.S., and Kone, B.C. (2006). Dot1a-AF9 complex mediates histone H3 Lys-79 hypermethylation and repression of ENaCalpha in an aldosterone-sensitive manner. J. Biol. Chem. 281, 18059–18068.

Chapter 2. Biochemical and Structural characterization of the DOT1L – AF9 interaction

2.1. Introduction

The only evidence for a direct interaction between DOT1L and AF9 came from several biochemical studies that broadly delineated interacting regions of both DOT1L as well as AF9. Our initial goal was to biochemically validate the DOT1L-AF9 interaction and we hoped that this would lead to future structure studies, which would provide better understanding as to how the AF9 AHD could mechanistically bind to multiple proteins.

Coexpression of the previously published interacting regions of AF9 and DOT1L with one another resulted in stable complexes but very poor NMR spectra, unsuitable for structural studies. We spent the next year creating deletion mutants in order to narrow down these large co-expressed constructs to identify the minimal interacting domains for this protein-protein interaction. Our end goal was to generate a finely mapped interaction suitable for use for structure determination either by X-ray crystallography or NMR. Intriguingly, this process led not only to high quality ^{15}N-^{1}H HSQC NMR spectra but also the identification of three separate DOT1L motifs for binding with AF9, a completely novel finding. The ^{15}N-^{1}H HSQC spectra of the three complexes of AF9 with the different DOT1L motifs show a strong similarity indicating that AF9 folds in a very similar manner with each of the separate DOT1L binding motifs. We proceeded to solve the NMR solution structure of the highest affinity DOT1L binding site in

complex with AF9. These results were the crux of our DOT1L-AF9 structure-function studies.

2.2. Identification and mapping of three separate DOT1L binding sites to AF9

Previous biochemical studies have broadly determined regions of both DOT1L (aa. 479-659 and aa. 828-1095) as well as AF9 (aa. 392-568) which mediate their interaction (Biswas et al., 2011; Mueller et al., 2009; Yokoyama et al., 2010; Zhang et al., 2006). Using Mueller et al., as a starting point, we co-expressed DOT1L residues (aa. 826-1095) with the entire C-terminal domain of AF9 (aa. 392-568). SDS-PAGE and size exclusion chromatography data validated this protein-protein interaction. However, 2D ^{15}N-^{1}H HSQC NMR spectrum of the complex showed extreme broadening, and an insufficient number of peaks with differing intensities for the amino acids (**Figure 2-1A)**. Additionally, we observed poor chemical shift dispersion as evidenced by the domination of strong resonance peaks appearing between ^{1}H 7.5 and 8.5 ppm, which are characteristic of unstructured flexible residues. While we did notice a small number of weak, but dispersed peaks, this was a disappointing result. At this time we were conducting structural studies with AF4-AF9 solution structure, but as the minimal protein binding fragments for both AF4 and AF9 proteins were already published, it was possible that either the DOT1L-AF9 interaction had not been precisely mapped or this DOT1L-AF9 protein complex was not even suitable for NMR studies.

We created a library of deletion constructs and analyzed them using heteronuclear triple resonance NMR to assign resonances and {^{15}N}-^{1}H heteronuclear NOE measurements to assess the dynamic behavior of the residues. These experiments enabled us to identify what residues were flexible, not involved in the protein-protein interaction, and remove them (**Figures 2-1B - E**). This was carried out in an iterative fashion with removal of flexible residues from the constructs to optimize the behavior of the DOT1L-AF9 complexes. This process led to improved NMR spectra, and the appearance of double sets of resonances for a majority of residues (**Figure 2-1C)**. Upon examination of the DOT1L sequence (aa. 863-900) we noticed a similar motif that is repeated: LXIXIXL (aa. 863-869) versus LXVXIXL (aa. 879-885) where the only differing residue is the Isoleucine (aa. 865) versus a Valine (aa. 881) (**Figure 2-2A)**. Thus, our ^{15}N-^{1}H HSQC with double sets of resonances was representative of two separate DOT1L interaction sites with AF9. Additionally, another study showed an interaction between AF9 and an upstream portion of DOT1L, nearly 400 residues away from the other two binding sites (aa. 479-659) (Zhang et al., 2006). We employed a similar process to remove the flexible residues in the interaction and were able to identify a third DOT1L binding site to AF9. Strikingly, this course of action led not only to high quality ^{15}N-^{1}H HSQC NMR spectra but also the identification of three separate DOT1L motifs with AF9: Site 1 (aa. 628-653), Site 2 (aa. 863-878), and Site 3 (aa. 877-900). For AF9, we identified a functional domain (aa. 499-568) that produces optimal NMR spectra with all three DOT1L motifs (**Figure 2-2)**. Indeed, this region of AF9 defined, is the same

region for which we have determined the structure of an AF4 peptide-AF9 complex (Leach et al., 2013).

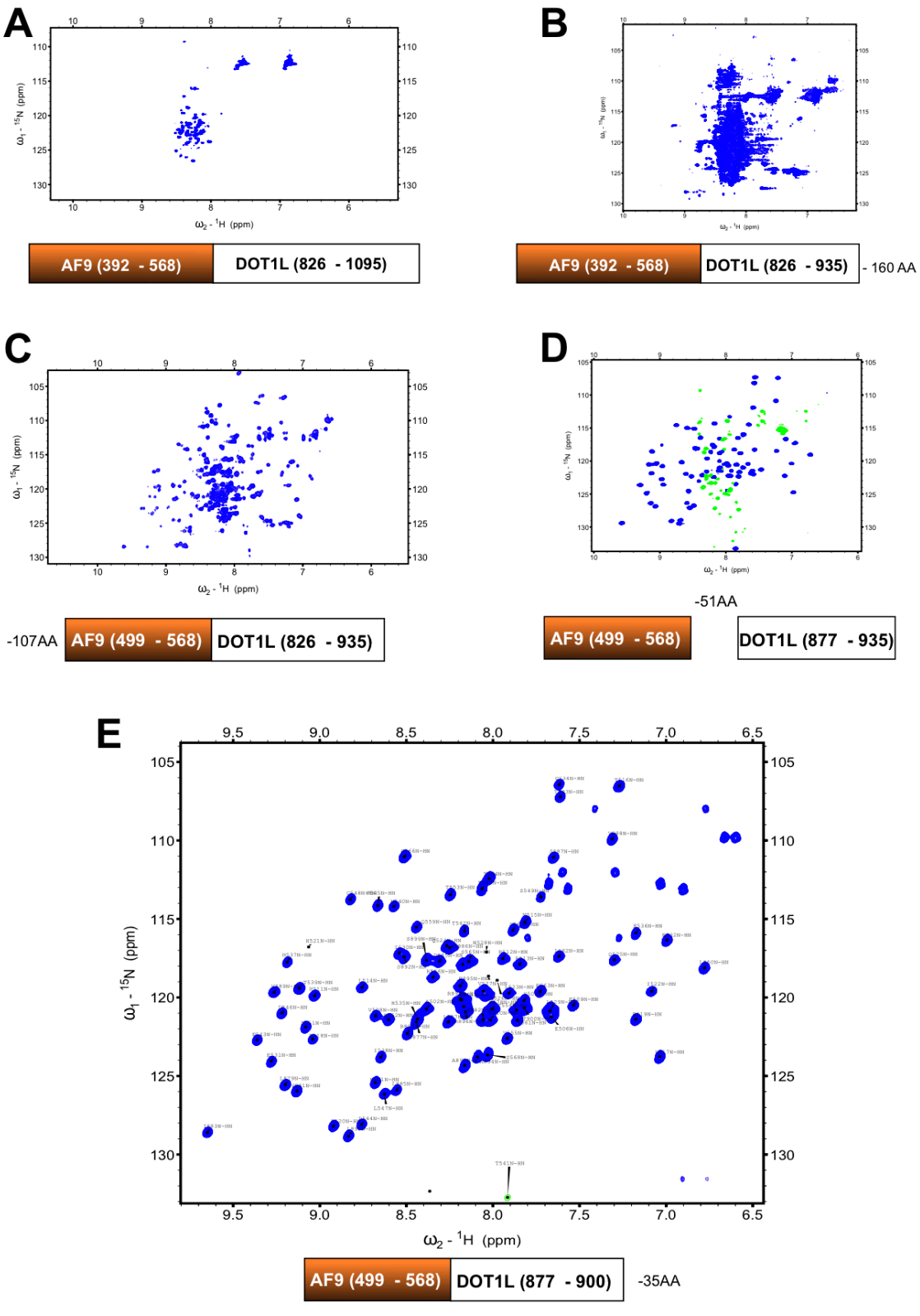

A
B
C
D
E
AF9 (392 - 568)
DOT1L (826 - 1095)
AF9 (392 - 568)
DOT1L (826 - 935)
- 160 AA
-107AA
AF9 (499 - 568)
DOT1L (826 - 935)
-51AA
AF9 (499 - 568)
DOT1L (877 - 935)
AF9 (499 - 568)
DOT1L (877 - 900)
-35AA
ω1 - 15N (ppm)
ω2 - 1H (ppm)

Figure 2-1. Identification of the minimal interaction sites of DOT1L-AF9. **A)** ^{15}N-^{1}H HSQC NMR spectrum of the coexpressed AF9 and DOT1L complex using sites identified by (Mueller et al., 2009) shows minimal dispersion. Both AF9 and DOT1L are ^{15}N labeled. **B-C)** We observe the appearance of dispersed resonance peaks when removing flexible residues from both DOT1L and AF9 that are not involved in the protein-protein interaction. **D)** Representative {^{15}N}-^{1}H heteronuclear NOE spectrum where green resonance peaks are flexible. These peaks were assigned and removed from the DOT1L-AF9 construct. **E)** Final ^{15}N-^{1}H HSQC assigned NMR spectrum of the DOT1L (Site 3) - AF9 complex.

2.3. DOT1L binding sites have differing binding affinities to AF9

To measure binding affinities we utilized a fluorescence anisotropy peptide binding assay. Fluorescently tagged DOT1L peptides were titrated with Maltose Binding Protein (MBP) tagged AF9. We were forced to use MBP-AF9 as this construct yielded far more soluble protein than expressing AF9 on its own, which we have previously seen was difficult to purify and was prone to aggregation. Comparison of the DOT1L motifs across species shows that they are conserved and exhibit a periodic pattern of hydrophobic residues consistent with that of a β-strand and are predicted to be intrinsically disordered (**Figure 2-3**). The DOT1L (Site 2) and (Site 3) motifs are nearly identical, with the exception of a bulkier Isoleucine residue in the third position of the DOT1L (Site 2) binding site as opposed to a Valine in DOT1L (Site 3) (**Figures 2-2A and 2-4)**. We hypothesized that this would lead to weaker binding due to the spatial constraints within the DOT1L-AF9 binding pocket. Indeed, fluorescence polarization based binding studies revealed a 17 fold weaker affinity for DOT1L (Site 2) with AF9 (**Figure 2-4)**. In the same position, DOT1L (Site 1) has a bulkier Leucine residue as well as a Leucine in the 10^{th} position, as opposed to a Valine, as seen in the other two binding sites (**Figures 2-2A and 2-4)**. Our binding studies show weak affinity for this binding site (K_d > 2000 nM) (**Figure 2-4)**. As ENL, another fusion partner with MLL, is highly homologous to AF9, we tested binding of each of these DOT1L peptides with ENL, which yielded similar binding affinities for ENL as seen for AF9 (**Figure 2-5)**.

Our binding measurements of the DOT1L (Site 2) with AF9 yielded K_d values significantly different from those reported in a recent publication that biochemically mapped the interaction of this particular DOT1L site with AF9, but did not identify the interactions of the other DOT1L sites (Shen et al., 2013). This publication reported the binding of DOT1L (Site 2) with AF9 to be on the order of 65nM using Fluorsecence Polarization, but 1100 nM using ITC, compared to our measurement of 26.7nM for this binding site. It is surprising that they obtain vastly different binding affinities using two separate techniques. As is the case with many intrinsically disordered proteins, the AF9 AHD has a propensity to aggregate, requiring significant care in the concentrations and conditions employed for binding measurements, perhaps suggesting a rationale for the difference in the measured binding affinities.

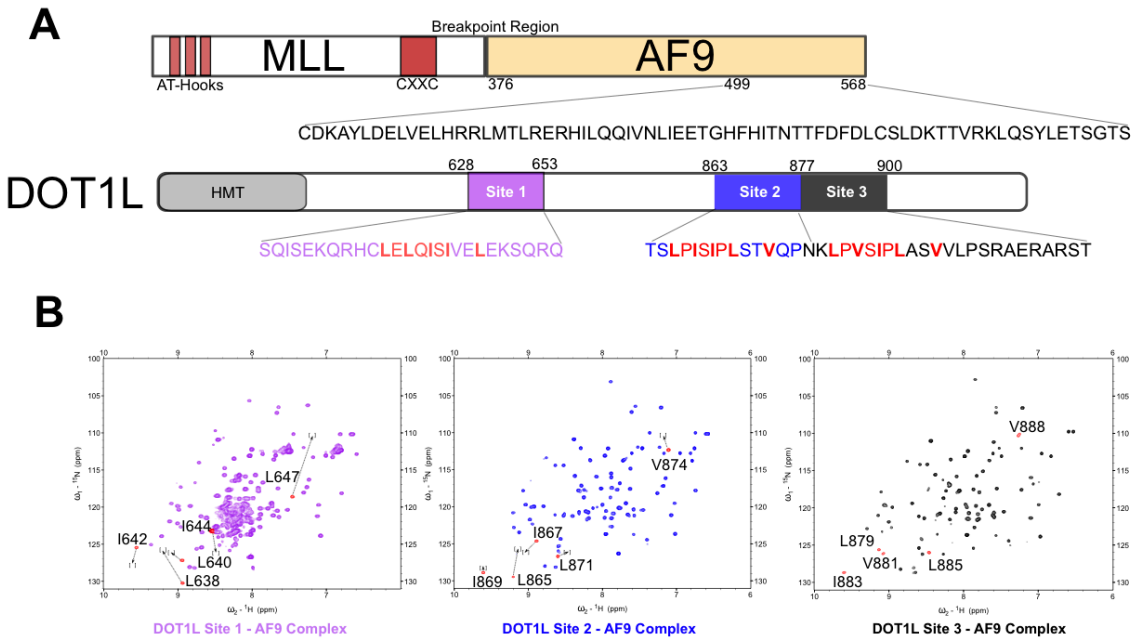


Figure 2-2. DOT1L has three separate sites of interaction with AF9.

A) The minimal interacting sites of DOT1L, Site 1 (Purple), Site 2 (Blue) and Site 3 (Black) with MLL-AF9 (aa. 499-568). Identified in red are the similar motifs making up each of the binding sites. **B)** ^{15}N-^{1}H HSQC NMR spectra of the minimal interacting sites of DOT1L with AF9. Labeled in red are the chemical shifts of hydrophobic residues within the separate DOT1L motifs compared to that of the same residues within DOT1L Site 3 (displayed as brackets).

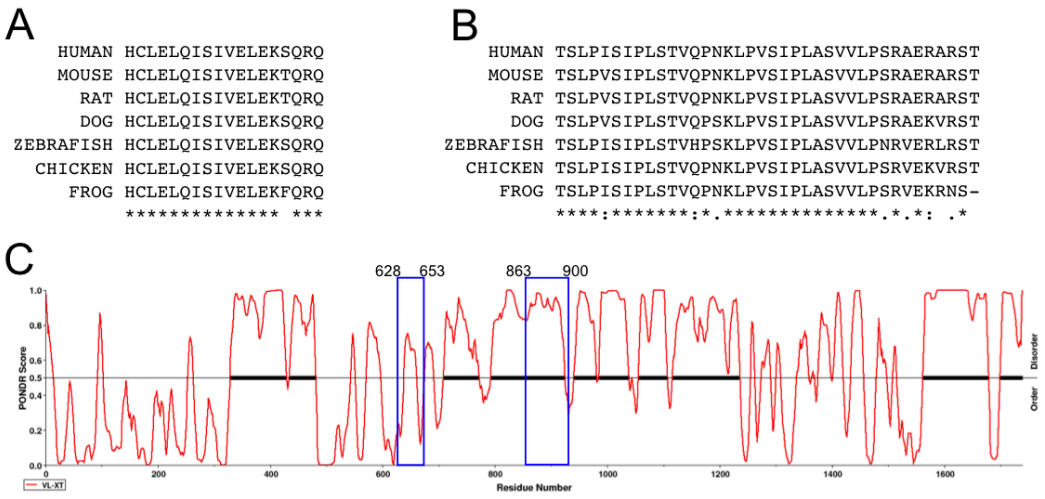
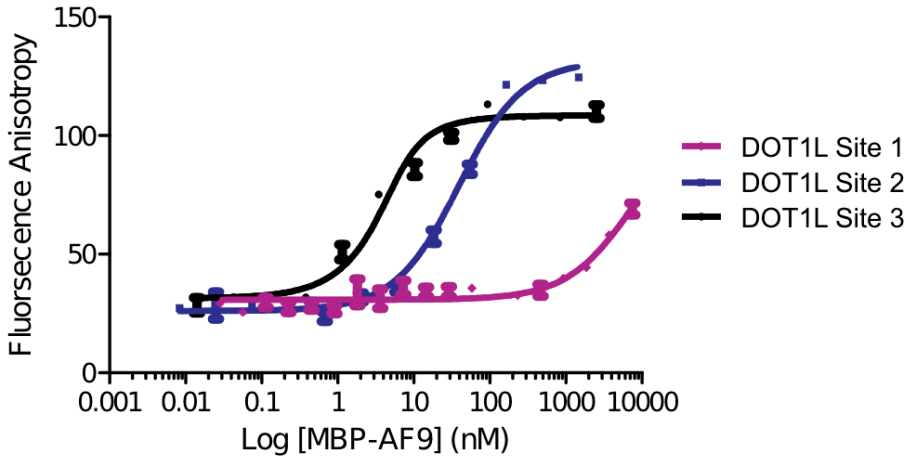


Figure 2-3. The three DOT1L sites for binding to AF9 are conserved and are intrinsically disordered.

A) Site 1 and **B)** Site 2 and Site 3, show that these binding sites are conserved across species. **C)** PONDR plot of full length DOT1L protein. Highlighted in blue are the DOT1L binding sites to AF9. Each binding site is predicted to be intrinsically disordered.

A

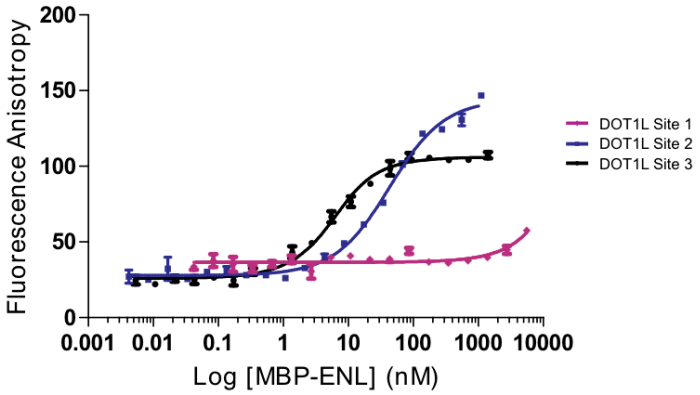


B

DOT1L Binding Sites	Binding Affinities with AF9 (Kd in nM)
Site 1 LELQISIVEL	> 2000
Site 2 LPISIPLSTV	26.7 ± 6.7
Site 3 LPVSIPLASV	1.56 ± .02

Figure 2-4. The three DOT1L binding motifs also bind to the AF9 AHD. **A)** Results of fluorescence polarization assay for determination of the Kd values for binding of MBP-AF9 AHD to each of the three DOT1L binding motifs. **B)** Table of Kd values for the three DOT1L sites with their respective primary sequences. Shown in red are hydrophobic residues of these DOT1L motifs. Highlighted is the third position of this motif that differs between each of the binding sites.

A



B

DOT1L Binding Sites	Binding Affinities with ENL (Kd in nM)
Site 1 LELQISIVEL	> 5500
Site 2 LPISIPLSTV	41 ± 3.4
Site 3 LPVSIPLASV	3.6 ± .46

Figure 2-5. The three DOT1L binding motifs also bind to the ENL AHD.

A) Results of fluorescence polarization assay measurement of binding of MBP-ENL AHD binding to the three different peptide binding motifs from DOT1L. DOT1L Site 1 binding was too weak to measure. **B)** Table of binding affinities for DOT1L motifs to MBP-ENL AHD are very similar to those obtained for binding to AF9.

2.4. Structure of the highest affinity DOT1L-AF9 Complex

We solved the NMR solution structure of the highest affinity DOT1L-AF9 complex, (DOT1L Site 3 877-900) (PDB ID: 2MV7) using dihedral angle, NOE, and residual dipolar coupling (RDC) restraints without any significant constraint violations (**Table 2.1**). To calculate and refine the structure, we used dihedral angle, NOEs, and residual dipolar couplings (RDCs) as restraints (**Figure 2-6**). The NMR ensemble of the 10 lowest energy structures of DOT1L-AF9 shows a well-formed complex (**Figure 2-7A**). The DOT1L-AF9 complex forms a mixed alpha-beta structure and the DOT1L residues (879-884) form a β strand followed by a β turn (aa. 885-888). Immediately C-terminal, residues 889-895 make contacts with AF9, but not as significant as the preceding DOT1L residues and the following C-terminal amino acids, 896-900, are unstructured (**Figure 2-7B**). The interface between the two proteins is largely hydrophobic, as DOT1L L879, V881, I883, L885, and V888 are critical hydrophobic residues that are buried within the DOT1L-AF9 interface (**Figure 2-7C**). AF9 forms three helices around the DOT1L peptide (α1, α2, α3) and a β hairpin (β1 and β2), which forms a three-stranded antiparallel β sheet with the β strand from DOT1L; the C-terminus of AF9 (aa. 563-568) is unstructured (**Figure 2-7B**). The structure of the DOT1L-AF9 complex is remarkably similar to our previously solved AF4-AF9 complex (PDB ID: 2LM0). Both DOT1L and AF4 share a similar consensus hydrophobic motif (**Figure 2-8A**) and superposition of the backbone residues yields an RMSD of 1.4 Å (**Figure 2-7D**).

Table 2.1 Statistics for NMR Data Collection and Structure Calculations of the DOT1L-AF9 Complex	Value
NMR Distance and Dihedral Constraints	
Distance Constraints	
Total NOE	1667
Intraresidue	821
Interresidue	846
Sequential ([|i-j| = 1)	400
Medium range (|i-j| <= 4)	216
Long range (|i-j| > 5)	230
Intermolecular	136
Total dihedral angle restraints	139
ϕ	69
ψ	70
Structure Statistics	
Violations (mean and SD)	
Distance constraints (Å)	.021 +/- .001
Dihedral angle constraints (°)	.450 +/- .070
Maximum dihedral angle violation (°)	3.749
Maximum distance constraint violation (Å)	0.388
Deviations from idealized geometry	
Bond lengths (Å)	.003 +/- 0.000
Bond angles (°)	.465 +/- 0.005
Impropers (°)	.296 +/- 0.010
Average pairwise rmsd (Å)	
Heavy	0.916
Backbone	0.3
Total RDCs	
50 +/-MA	
HN	49
Cα-C	44
50+M	
Cα-C	42
RDCs used for validation but not for structure calculation	
HN	43
Qfree (%)	24.4

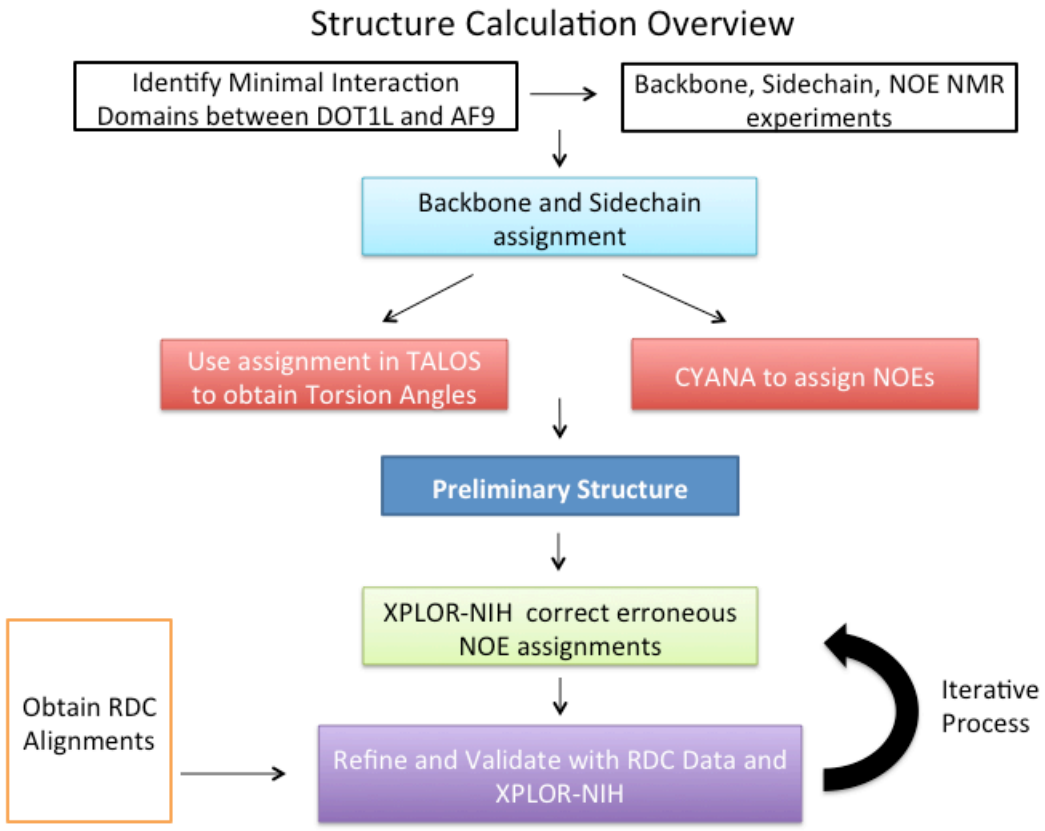


Figure 2-6. Schematic of the NMR Structure Calculation process.

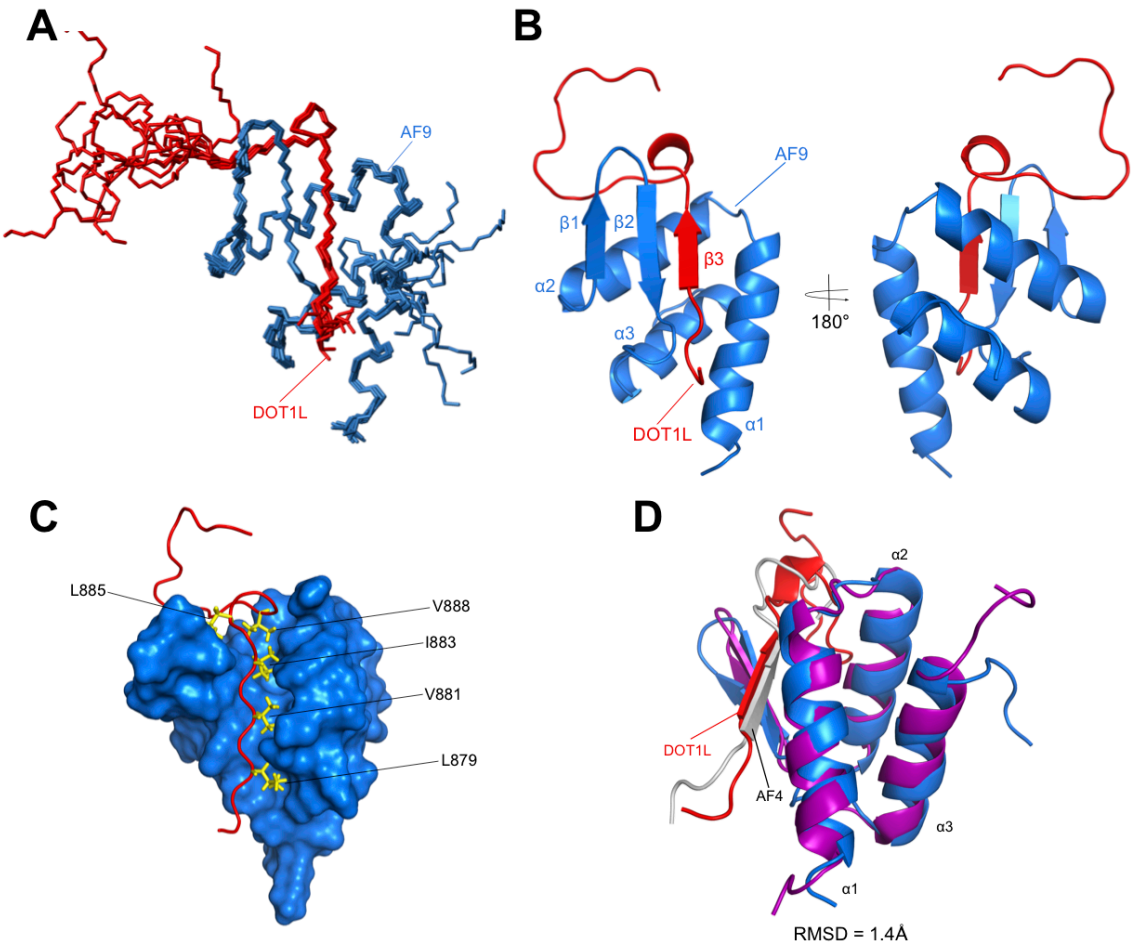


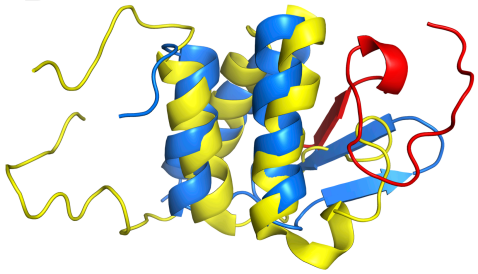
Figure 2-7. Structure of the DOT1L-AF9 complex.

A) Ensemble of the 10 lowest energy conformers. DOT1L is shown in red and AF9 in blue. **B)** Cartoon representation of the lowest energy conformer. DOT1L (red) forms a β strand, along with the β hairpin from AF9 (blue), while AF9 additionally forms three α helices. **C)** Surface representation of the DOT1L-AF9 complex. Shown are hydrophobic residues from DOT1L that are buried within the protein-protein interface and are critical for the interaction. **D)** Superposition of DOT1L-AF9 with our previously solved AF4-AF9 complex (PDB code: 2LM0; AF9 is purple and AF4 is white). RMSD = 1.4Å.

A

AF4	LMVKITLDLL
Site 1	LELQISIVEL
Site 2	LPISIPLSTV
Site 3	LPVSIPLASV

B



RMSD = 3.98 Å

Figure 2-8. Similarity of DOT1L-AF9 to other proteins.

A) AF4 has a similar hydrophobic motif with all of the DOT1L binding sites. The hydrophobic residues are highlighted in red. **B)** Superposition of BRD4 ET domain (yellow) with DOT1L-AF9 (RMSD=3.98 Å). The BRD4 fold is similar to the helical fold of AF9 (blue).

According to the Dali server, the BRD4 ET domain (PDB ID: 2JNS) has a similar helical fold to the helical portion of AF9 (RMSD = 3.98Å) (**Figure 2-8B**). My colleagues in the lab attempted to observe whether the BRD4 ET domain binds to DOT1L or AF4, but it appears that these proteins do not interact.

2.5. The DOT1L-AF9 complexes are structurally similar

In order to explore the solution dynamics properties of DOT1L-AF9, and to define regions properly suitable for RDC refinement, we conducted a suite of NMR dynamics experiments to probe protein backbone motion. This information provides insights into the local structural rigidity, which may be related to protein function and stability. $\{^1H\}^{15}N$ heteronuclear NOE experiments on the two DOT1L high affinity-binding sites (Site 2) (**Figure 2-9B**) and (Site 3) (**Figure 2-9D**) complexes show that both of the N-terminal (aa. 500-501), C-terminal domains of AF9 (aa. 563-568) and the C-terminal regions of both DOT1L peptides (aa. 877-878 in Site 2) and (aa. 896-900 in Site 3) are flexible, confirming the residue boundaries we used to define the protein construct. We also recorded R_1 (Spin-Lattice) and R_2 (Spin-Spin) relaxation rates and examined the R_1R_2 product of both high affinity complexes (**Figures 2-9A and 2-9C**). The advantage of the R_1R_2 product is that it can distinguish between motional anisotropy and chemical exchange, allowing for the identification of residues with altered dynamics. Chemical exchange on slow time scales (μs-ms) results in increased $R_1 * R_2$ values whereas fast time scale motions (ns and sub-ns) result in decreased values (Kneller et al., 2002). The high affinity binding sites both show similar patterns. Consistent with our heteronuclear NOE data, the majority

of the AF9 residues were ordered but we observe generally elevated $R_1 * R_2$ values near the loop regions on AF9 (aa. 516-521, 532-534, 540-543, 549-554). Additionally, the β strand region of DOT1L and the following extended strand (aa. 878-882, 889-890, 895-900) generally show depressed $R_1 * R_2$ values indicating fast time scale motions. Residues 892 and 893 which are located following the β turn and prior to the extended β strand exhibit elevated $R_1 * R_2$ values (**Figure 2-10A**). The β strand of DOT1L Site 2 generally shows depressed $R_1 * R_2$ values but, intriguingly, shows a gradual increase towards the end of the β strand followed by a gradual decrease after the β strand (**Figure 2-9A**).

The ^{15}N-^{1}H HSQC spectra of the three complexes of AF9 with the different DOT1L motifs show a strong similarity. Indeed, comparison of the individual AF9 amide chemical shifts among the three complexes shows relatively small changes. As chemical shifts are highly dependent on protein structure and local environment, this indicates that AF9 folds in a very similar manner with each of the separate DOT1L binding motifs. The chemical shift differences are minimal between the three DOT1L-AF9 complexes, but larger chemical shift changes are seen around AF9 residues adjacent to the DOT1L peptide (aa. 537-547) (**Figures 2-10B and 2-10C**). Our NMR data and chemical shift mapping results suggest that the three sites of DOT1L all form the same mixed alpha-beta structure. The DOT1L peaks show larger chemical shift changes but are located in similar regions in the NMR spectra (**Figure 2-2B**).

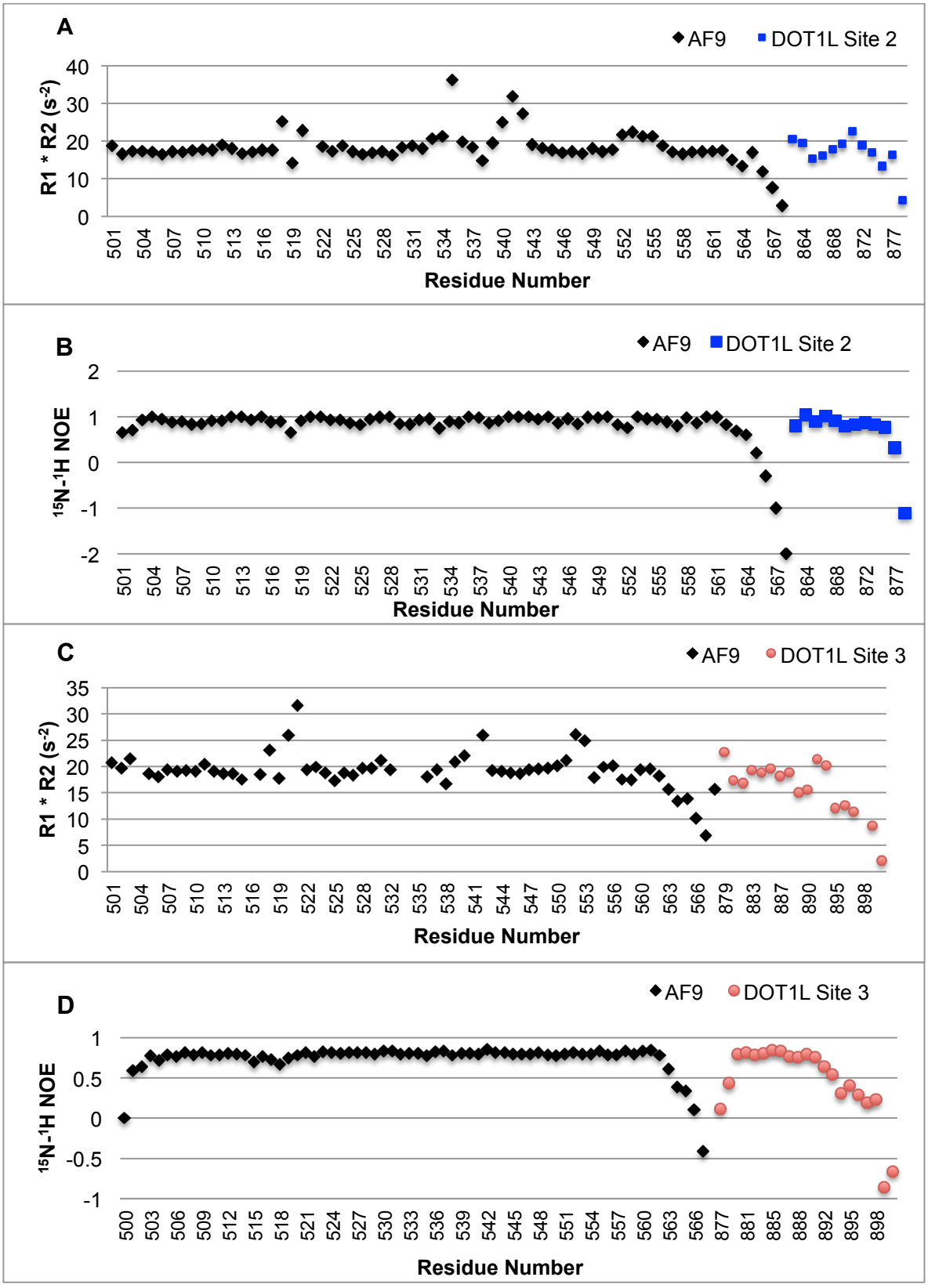


Figure 2-9. Backbone Dynamics of the high affinity DOT1L-AF9 complexes
AF9 residues are shown in Black, DOT1L Site 2 in Blue, DOT1L Site 3 in Red.

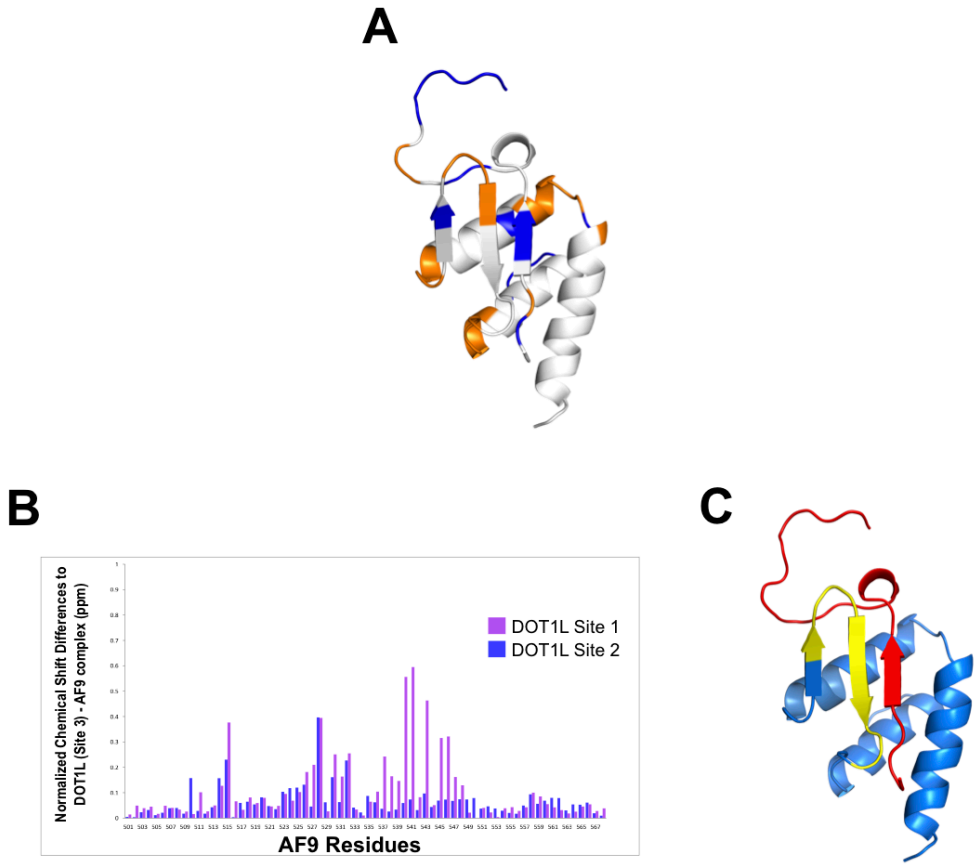


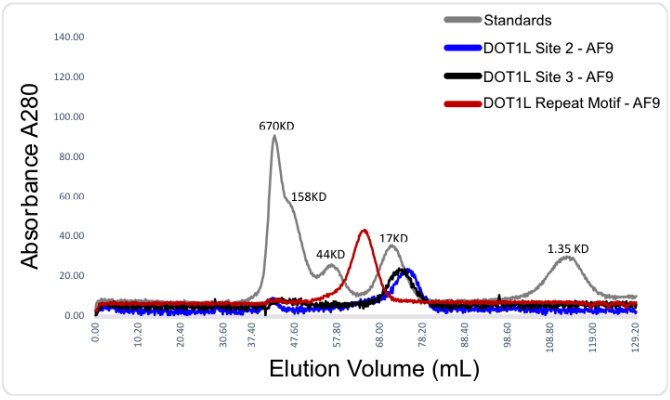
Figure 2-10. Backbone Dynamics and Chemical Shift Perturbations.

A) Mapping of elevated (Orange) and depressed (Blue) $R_1 * R_2$ values onto the DOT1L-AF9 structure. **B)** Chemical shift difference between AF9 amide NH resonances from the DOT1L (Site 1)-AF9 complex (Purple) and DOT1L (Site 2)-AF9 complex (Blue) compared to DOT1L (Site 3)-AF9. **C)** The most significant differences in chemical shifts of AF9 (aa. 537 – 547) mapped on the DOT1L (Site 3) - AF9 NMR structure (Yellow), are located near the DOT1L binding site.

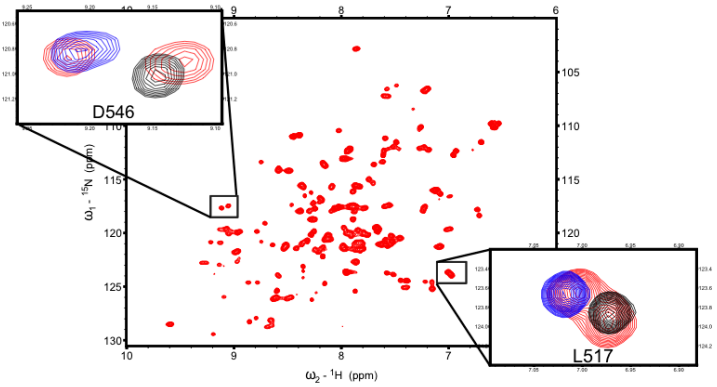
2.6. AF9 can simultaneously bind to each DOT1L site within the repeat motif

Intriguingly, DOT1L (Site 2) and DOT1L (Site 3) are high affinity binding motifs, which are only separated by a 4 amino acid spacer sequence, i.e. they form a repeat motif in DOT1L. To test whether it was sterically feasible for two AF9 AHDs to bind simultaneously to both of these DOT1L sites, we co-expressed the entire repeat motif consisting of both DOT1L (Sites 2 and 3) (aa. 863-900) with the AF9 AHD. Size-exclusion chromatography shows the formation of a complex consistent with two AF9 AHDs bound. The ^{15}N-^{1}H HSQC NMR spectrum of this complex shows that there are duplicate AF9 peaks for each AF9 residue which overlay very closely with the AF9 amide resonances from the DOT1L (Site 2) - AF9 and DOT1L (Site 3) - AF9 individual complexes (Figure 1H). Additionally, the DOT1L resonances in this entire repeat motif overlap with the resonances from each of the individual DOT1L (Site 2) and (Site 3) complexes, indicating that the complexes on the repeat motif closely resemble those of the individual complexes.

A


Standards
DOT1L Site 2 - AF9
DOT1L Site 3 - AF9
DOT1L Repeat Motif - AF9
Absorbance A280
670KD
158KD
44KD
17KD
1.35 KD
Elution Volume (mL)

B


D546
L517
ω1 - 15N (ppm)
ω2 - 1H (ppm)

C

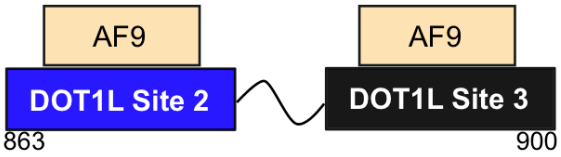

AF9
AF9
DOT1L Site 2
DOT1L Site 3
863
900

Figure 2-11. The DOT1L repeat motif can simultaneously bind to two AF9 AHDs.

A) Size exclusion profile of complexes of AF9 with DOT1L Site 2 (aa. 863-878; blue), DOT1L Site 3 (aa. 877-900; black), and with the repeat element (aa. 863-900; red) bound to AF9 overlaid with protein standards (grey). The size of AF9 bound to the DOT1L repeat motif is consistent with that of two AF9 proteins bound to the repeat motif. **B)** ^{15}N-^{1}H HSQC NMR spectrum of DOT1L repeat motif (aa. 863-900; red) bound to AF9. Shown are examples of two AF9 resonance peaks, overlaid with the same AF9 amide peak from DOT1L Site 2(aa. 863-878) – AF9 (blue) and DOT1L Site 3 (aa. 877-900) – AF9 (black). **C)** Cartoon depicting that two separate AF9 proteins can bind simultaneously to both high affinity DOT1L binding sites.

2.7. Discussion

Our study provides important insights into the detailed structural mechanism of the interactions of DOT1L with MLL-AF9. We have shown for the first time that there are three AF9 binding motifs in DOT1L. Additionally, we have characterized the binding to all three and solved the first 3D structure of a DOT1L-AF9 complex. Furthermore, our NMR studies show that each separate DOT1L interacting site forms a similarly structured DOT1L-AF9 complex. We previously showed that the ^{15}N-^{1}H HSQC NMR spectra of AF9 in complex with the other AF9 binding partners (AF4, CBX8, BCOR) closely resemble one another, displaying minimal differences in the AF9 chemical shifts, strongly suggesting that all of the AF9 complex structures are very similar (Leach et al., 2013). Our results provide a structural basis for findings that interactions of AF9 are mutually exclusive with its different binding partners (Biswas et al., 2011; Leach et al., 2013; Yokoyama et al., 2010). Indeed, our NMR solution structure of the highest affinity DOT1L-AF9 complex is remarkably similar to that of AF4-AF9, as both form nearly identical mixed α/β structures. The minimal differences between these structures stem from the dynamic loops in both protein complexes and the non-hydrophobic residues on both DOT1L and AF4 β-strands. These residues are not buried within the protein-protein complex and form different residue-specific interactions with AF9. From a therapeutic perspective, this suggests targeting of AF9, or ENL will likely block all binding partners, which might be favorable for achieving inhibition of MLL fusion protein activity.

The AF9 AHD is intrinsically disordered, a characteristic seen in numerous signaling hubs, potentially allowing the hub protein to gain kinetic advantages from the binding of one interacting partner and its exchange for another (Dunker et al., 2005; Dyson and Wright, 2005). Each of the AF9 binding proteins contains a similar binding motif, consistent with the observation that many proteins are enriched with short linear motifs, or eukaryotic linear motifs, which are small intrinsically disordered regions of functional modules from 3 to 11 amino acid residues (Tompa et al., 2014). These play a large role in many hub proteins that regulate diverse cellular processes. Eukaryotic linear motifs are enriched with post-translational modification sites that serve as a means of dynamic regulation and control of their activity (Van Roey et al., 2014; Tompa et al., 2014). Indeed, we have previously shown that phosphorylation of AF4 at a site that has been shown to be phosphorylated in cells (Beausoleil et al., 2004) reduces its affinity for AF9 (Leach et al., 2013). DOT1L is unique among the AF9 binding partners in that there are multiple DOT1L motifs that bind to AF9, and each is predicted to be disordered. Phosphorylation of serines within each of the DOT1L binding motifs (aa. 643, 868, 882) could also lead to reduction in binding, but only S882 has been shown to be phosphorylated *in vivo* (Hornbeck et al., 2012).

As the high affinity binding sites both contain exposed prolines, another means of regulation could be from proline isomerization. Prolines are the most disorder-promoting residues and the overall proline content in IDPs is 1.4-times higher than on surfaces of folded proteins (Radivojac et al., 2007).
Proline *cis/trans* isomerization reactions catalyzed by peptidyl-prolyl isomerases

(PPIases) can impact protein function and many IDPs are PPIase substrates (Theillet et al., 2014). Thus, proline *cis/trans* isomerization could regulate the number of DOT1L high-affinity motifs that are bound to AF9 at any given time. While this is purely speculative, this isomerization event could be combined with phosphorylation to provide a 2-way control over the protein's function. Intriguingly, enzymes such as Pin1, a PPIase, establish faster inter-conversion rates upon phosphorylation. The phosphorylation status of the pSer7-Pro8 motif of the intrinsically disordered, C-terminal domain (CTD) of RNA polymerase II correlates with transcriptional activity. Indeed, proline *cis/trans* isomerization has been identified as the rate-limiting step in Ser7 dephosphorylation as the dephosphorylation in the CTD pSer7-Pro8 sequence occurred much faster when Pin1 was present (Bataille et al., 2012; Werner-Allen et al., 2011). Other examples exist but it is clear that PPIase activites can represent another level of protein regulation, and may have an important role in IDP function.

It has been shown that many short linear motifs may fold into regular secondary structures upon binding to their partners and may be repeated within a protein sequence. One advantage of repeat, or bivalent, intrinsically disordered protein (IDP) interactions that include two or more (or three) individual IDP motifs is that they are able to compensate for the entropic cost of an unfavorable single IDP motif binding event (Barbar and Nyarko, 2014). These multiple motifs and binding events enable the linker regions to stay flexible and are tunable by post-translational modifications. Future experiments with this system entail validation of these multiple binding events *in vitro* and a further understanding of the

underlying kinetics. Even though a large number of eukaryotic motifs in IDPs have been characterized in various diseases, the biological function and necessity of multiple eukaryotic linear motifs within one protein is only beginning to be unraveled (Tompa et al., 2014).

2.8. Experimental Procedures

2.8.1. Protein Cloning, Expression, and Purification

The pETDuet-1 vector (EMD Millipore) was modified to include an N-terminal thioredoxin tag, for expression and solubility enhancement, followed by an inserted TEV protease site and a 6x Histidine tag directly upstream of the first pETDuet-1 cloning site. DOT1L was cloned into in the first cloning site and AF9 was cloned in the second cloning site of pETDuet-1. Proteins were co-expressed in Rosetta 2(DE3) cells (EMD Millipore) in European Molecular Biology Laboratory media supplemented with either 10ml/L of unlabeled, ^{15}N, or ^{15}N and ^{13}C Bioexpress (Cambridge Isotope Labs). Also added to the media as necessary were ^{15}N-labeled ammonium sulfate and ^{13}C-labeled glucose for either ^{15}N-labeled protein samples or ^{15}N,^{13}C-labeled samples for NMR data collection. Proteins were expressed for 6 hours at 30°C with 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG). The DOT1L-AF9 protein complex was purified using Ni-NTA chromatography. The thioredoxin and histidine tags were removed by TEV protease digestion overnight, and the protein further purified by a second round of passage through the Ni-NTA column. As a final step, any residual aggregates and impurities were removed using size exclusion chromatography with Superdex 75 column (GE Life Sciences). All subsequent DOT1L-AF9 constructs were expressed and purified in the same manner.

2.8.2. Optimization and Comparison of DOT1L-AF9 Complexes

All protein samples were exchanged into the same buffer employed for NMR data collection containing 25mM Bis-Tris/MES (pH 6.0), 100mM NaCl,

1mM DTT. NMR experiments were conducted using either a Varian Inova 600 MHz or a Bruker 600 MHz magnet equipped with cryogenically cooled probes. All NMR data were processed and analyzed using NMRPipe (Delaglio et al., 1995) and Sparky (T.D. Goddard and J.M. Kneller, University of California, San Francisco). $\{^{1}H\}^{15}N$ heteronuclear NOE experiments were recorded with 3 second relaxation delays. Residues identified as flexible based on depressed heteronuclear NOE values were removed from the construct by inserting a and the structured residues were cloned into pETDuet-1 to generate the three final DOT1L-AF9 constructs: DOT1L residues: 1) (aa. 628-653) 2) (aa. 863-877) 3) (aa. 877-900) in the first pETDuet-1 cloning site, and AF9 residues (aa. 499-568) in the second cloning site. To express the DOT1L Site 2 and Site 3 repeat motifs with AF9, we cloned DOT1L containing both high affinity binding motifs (aa. 863-900) into the first site of petDuet-1 along with AF9 (aa. 499-568) in the second site. We collected 3D HNCACB and CBCA(CO)NH data to assign backbone protein resonances in each of the different DOT1L-AF9 protein complexes. To examine chemical shift differences between the DOT1L-AF9 complexes, weighted chemical shift changes in parts per million were calculated by using the equation: $\delta\Delta = (((0.2 \times \Delta N)^2 + \Delta H^2))^{½}$.

2.8.3. DOT1L-AF9 Complex Resonance Assignments

All NMR experiments for structure determination, double labeled (^{15}N ^{13}C) DOT1L (877-900) - AF9 complex at a concentration of 750 µM was used. To fully assign protein resonances for structure determination we ran the following experiments: HNHA, HNCO, HNCACB, HNCA, CBCA(CO)NH, CC(CO)NH-

TOCSY, HCCH-TOCSY, 3D ^{15}N-edited NOESY (80 ms mixing time), 3D ^{13}C-edited aliphatic NOESY (80 ms mixing time), and a 3D ^{13}C-edited aromatic NOESY (80 ms mixing time). Stereospecific assignments for the methyl groups of Val and Leu were made using a 10% ^{13}C sample (Neri et al., 1989).

2.8.4. DOT1L-AF9 Complex Structure Determination and Refinement

Calculation of the DOT1L-AF9 structure was completed in two steps. First, we calculated preliminary structures using the XPLOR-NIH simulated annealing protocol with restraints derived from dihedral angles and NOE distances (Schwieters et al., 2003). Φ and ψ dihedral angle restraints were generated by TALOS+ based on Cα, Cβ, C′, and N chemical shifts (Shen et al., 2009). Initial NOE assignments were performed using CYANA (Güntert et al., 1997) and were checked and corrected using SPARKY. Distance restraints were then generated based on NOE cross-peak intensities and placed into four categories: 1.8−2.8, 1.8−3.3, 1.8−4.2, and 1.8−5.5 Å for structure calculations. NOE violations greater than 0.2 Å were analyzed and appropriately corrected. The lowest energy structure conformation was then selected for a second stage of refinement in XPLOR-NIH. Here, the temperature annealing step was lowered to 4°C and Residual Dipolar Coupling (RDC) data were incorporated in addition to the dihedral angle and NOE restraints. RDC data were recorded in positively charged (50+M) and zwitterionic (50+-MA) compressed polyacrylamide gels (Cierpicki and Bushweller, 2004). The 50+M gels provided a different alignment from the 50+-MA gels. The RDC $^{1}D_{HN}$, and $^{1}D_{C'C\alpha}$ measurements were made using 3D HNCO based IPAP pulse sequences (Permi et al., 2000). A histogram

of the RDC distribution was used to estimate initial axial (D_a) and rhombic (R) components of the alignment tensor followed by final optimization of the D_a and R values by fitting of the measured RDCs to the preliminary structure using PALES (Zweckstetter and Bax, 2000). 50+- MA $^1D_{HN}$, and $^1D_{C'C\alpha}$ and 50+M $^1D_{C'C\alpha}$ data were used in the structure refinement calculations, however, 50+M $^1D_{HN}$ data were not used in the structure refinement process but were kept separately for structure validation and Q_{free} factor calculation using PALES. NMR structures were viewed using MOLMOL (Koradi et al., 1996). The 10 lowest total energy structures calculated out of 100 were used to represent the ensemble conformation. The ensemble structures were displayed using MOLMOL, and the ribbon and surface structures were displayed with PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC). Surface electrostatic potential was calculated using PyMOL APBS (http://www.poissonboltzmann.org/apbs/) tools.

2.8.5. NMR Relaxation Experiments

T1 and T2 NMR relaxation experiments were conducted to study protein dynamics using relaxation delays of 10, 180, 300, 500, 900, 1300 and 1800ms (T1) and 10, 30, 50, 90, 130, 170 and 230ms (T2). Experiments were conducted on the Varian Inova 600 MHz magnet. Data were fitted using Sparky.

2.8.6. Fluorescence Polarization (FP) Binding Measurements

AF9 (475-568) was cloned into pMAL C2 (NEB), expressed in Rosetta 2(DE3) cells in LB medium, and the MBP-AF9 protein was purified using Amylose resin (NEB). This protein proved to be prone to aggregation and was

immediately purified using size exclusion chromatography with a Superdex 75 column (GE Life Sciences) at 4°C to remove the significant amount of aggregate protein. We used three separate DOT1L peptides corresponding to each AF9 binding site, which were synthesized with an N-terminal Fluorescein tag. For binding studies, the pure MBP-AF9 protein was titrated into 5 nM DOT1L peptides, the 5nM CBX8 or 1 nM AF4 synthesized peptides, in 50mM Tris-HCl, 150mM KCl, 1mM DTT, pH 7.5 buffer and incubated for 30 minutes at room temperature on 96-well black COSTAR (Corning Life Sciences, Lowell, MA) plates. A PHERAstar microplate reader (BMG Labtech, Durham, NC) was used to measure fluorescence polarization with excitation at 494 nm and emission at 525 nm. A binding curve was generated by fitting data to a single site binding model accounting for ligand depletion (Veiksina et al., 2010).

We used the following peptides purchased from Biosyn:

DOT1L Site 1: FITC-AHX- SQISEKQRHCLELQISIVELEKSQRQ

DOT1L Site 2: FITC-AHX-TSLPISIPLSTVQPNK

DOT1L Repeat Motif: FITC-AHX-

TSLPISIPLSTVQPNKLPVSIPLASVVLPSRAERARST

2.9. References for Chapter 2

Barbar, E., and Nyarko, A. (2014). Polybivalency and disordered proteins in ordering macromolecular assemblies. Semin. Cell Dev. Biol.

Bataille, A.R., Jeronimo, C., Jacques, P.-É., Laramée, L., Fortin, M.-È., Forest, A., Bergeron, M., Hanes, S.D., and Robert, F. (2012). A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. Mol. Cell *45*, 158–170.

Beausoleil, S.A., Jedrychowski, M., Schwartz, D., Elias, J.E., Villén, J., Li, J., Cohn, M.A., Cantley, L.C., and Gygi, S.P. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc. Natl. Acad. Sci. U. S. A. *101*, 12130–12135.

Biswas, D., Milne, T. a, Basrur, V., Kim, J., Elenitoba-Johnson, K.S.J., Allis, C.D., and Roeder, R.G. (2011). Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. Proc. Natl. Acad. Sci. U. S. A. *108*, 15751–15756.

Cierpicki, T., and Bushweller, J.H. (2004). Charged gels as orienting media for measurement of residual dipolar couplings in soluble and integral membrane proteins. J. Am. Chem. Soc. *126*, 16259–16266.

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*.

Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M., and Uversky, V.N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J. *272*, 5129–5148.

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. *6*, 197–208.

Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. *273*, 283–298.

Hornbeck, P. V, Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. *40*, D261–70.

Kneller, J.M., Lu, M., and Bracken, C. (2002). An Effective Method for the Discrimination of Motional Anisotropy and Chemical Exchange. J. Am. Chem. Soc. *124*, 1852–1853.

Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. J. Mol. Graph. *14*, 51–55.

Leach, B.I., Kuntimaddi, A., Schmidt, C.R., Cierpicki, T., Johnson, S. a, and Bushweller, J.H. (2013). Leukemia fusion target AF9 is an intrinsically disordered

transcriptional regulator that recruits multiple partners via coupled folding and binding. Structure *21*, 176–183.

Mueller, D., García-Cuéllar, M.-P., Bach, C., Buhl, S., Maethner, E., and Slany, R.K. (2009). Misguided transcriptional elongation causes mixed lineage leukemia. PLoS Biol. *7*, e1000249.

Neri, D., Szyperski, T., Otting, G., Senn, H., and Wuethrich, K. (1989). Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional carbon-13 labeling. Biochemistry *28*, 7510–7516.

Permi, P., Rosevear, P.R., and Annila, A. (2000). A set of HNCO-based experiments for measurement of residual dipolar couplings in 15N, 13C, (2H)-labeled proteins. J. Biomol. NMR *17*, 43–54.

Radivojac, P., Iakoucheva, L.M., Oldfield, C.J., Obradovic, Z., Uversky, V.N., and Dunker, A.K. (2007). Intrinsic disorder and functional proteomics. Biophys. J. *92*, 1439–1456.

Van Roey, K., Uyar, B., Weatheritt, R.J., Dinkel, H., Seiler, M., Budd, A., Gibson, T.J., and Davey, N.E. (2014). Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. Chem. Rev. *114*, 6733–6778.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Marius Clore, G. (2003). The Xplor-NIH NMR molecular structure determination package. J. Magn. Reson. *160*, 65–73.

Shen, C., Jo, S.Y., Liao, C., Hess, J.L., and Nikolovska-Coleska, Z. (2013). Targeting recruitment of disruptor of telomeric silencing 1-like (DOT1L): characterizing the interactions between DOT1L and mixed lineage leukemia (MLL) fusion proteins. J. Biol. Chem. *288*, 30585–30596.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR *44*, 213–223.

Theillet, F.-X., Kalmar, L., Tompa, P., Han, K.-H., Selenko, P., Dunker, A.K., Daughdrill, G.W., and Uversky, V.N. (2014). The alphabet of intrinsic disorder. Intrinsically Disord. Proteins *1*, e24360.

Tompa, P., Davey, N.E., Gibson, T.J., and Babu, M.M. (2014). A Million Peptide Motifs for the Molecular Biologist. Mol. Cell *55*, 161–169.

Veiksina, S., Kopanchuk, S., and Rinken, A. (2010). Fluorescence anisotropy assay for pharmacological characterization of ligand binding dynamics to melanocortin 4 receptors. Anal. Biochem. *402*, 32–39.

Werner-Allen, J.W., Lee, C.-J., Liu, P., Nicely, N.I., Wang, S., Greenleaf, A.L., and Zhou, P. (2011). cis-Proline-mediated Ser(P)5 dephosphorylation by the

RNA polymerase II C-terminal domain phosphatase Ssu72. J. Biol. Chem. *286*, 5717–5726.

Yokoyama, A., Lin, M., Naresh, A., Kitabayashi, I., and Cleary, M.L. (2010). A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. Cancer Cell *17*, 198–212.

Zhang, W., Xia, X., Reisenauer, M.R., Hemenway, C.S., and Kone, B.C. (2006). Dot1a-AF9 complex mediates histone H3 Lys-79 hypermethylation and repression of ENaCalpha in an aldosterone-sensitive manner. J. Biol. Chem. *281*, 18059–18068.

Zweckstetter, M., and Bax, A. (2000). Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. J. Am. Chem. Soc. *122*, 3791–3792.

Chapter 3. Degree of recruitment of DOT1L to MLL-AF9 defines hematopoietic transformation potential and level of H3K79 methylation on target genes

3.1. Introduction

As we were completing our structural studies, papers from other groups began to be published which showed that DOT1L plays a critical role in MLL-AF9 leukemogenesis (Biswas et al., 2011; Chang et al., 2010; Nguyen and Zhang, 2011; Nguyen et al., 2011; Shen et al., 2013; Yokoyama et al., 2010). However, none of these papers addressed the role of the direct recruitment of DOT1L to MLL-AF9. Thus, our structural studies led to the question: what biological roles do the multiple DOT1L-AF9 interactions play in MLL-AF9 leukemogenesis?

To answer this, we first developed and biochemically characterized structurally guided point mutations in both MLL-AF9 and DOT1L. We created several sets of mutations that either partially or completely disrupted this interaction. Colleagues in Dr. Nancy Zeleznik-Le's lab at the Oncology Institute of Loyola University Chicago introduced these point mutations into hematopoietic progenitor cells isolated from mouse bone marrow and conducted serial replating assays. We observed a graded decrease in serial replating capability contingent on the number of DOT1L and MLL-AF9 sites affected by the point mutations.

To further understand the biological impact of disrupting this protein-protein interaction, Jeremy Thorpe, from Dr. Mazhar Adli's lab, and I ran ChIP-Seq experiments. We pulled down H3K79me2 and H3K79me3 on the wild-type and mutant hematopoietic progenitor cells sent to us by the Zeleznik-Le lab and

analyzed and interpreted the genomic data with help from Ritambhara Singh, also in the Adli lab and Dr. Stefan Bekiranov. Strikingly, only a small subset of genes shows significant changes in the H3K79me2 or H3K79me3 marks when blocking the MLL-AF9 interaction with DOT1L.

This chapter contains three major sections:

1) The development and biochemical characterization of point mutations.
2) Serial replating assays with the characterized mutants.
3) Chip-Seq experiments and analyses.

3.2. Point mutations in AF9 and DOT1L differentially attenuate the multiple DOT1L-AF9 interactions

The DOT1L-AF9 structure was used as a basis to rationally design point mutations to disrupt this protein-protein interaction. Here we took a two pronged approach and developed mutations both in AF9 and in DOT1L.

3.2.1. Identification and biochemical characterization of AF9 mutations[3]

We identified an AF9 residue, D546, which is part of the AF9 β-hairpin (β2) and is in position to make an electrostatic interaction with DOT1L K878 (**Figure 3-1A**). We mutated this residue to arginine (denoted as D546R) to create charge repulsion. The D546R AF9 mutant significantly reduces binding to both DOT1L high affinity sites (**Figures 3-1B and 3-1C**), cannot pull down a DOT1L construct comprised of the high affinity DOT1L binding sites (**Figures 3-2A and 3-2B**), and has significantly decreased binding to full length DOT1L (**Figure 3-3A**). Interestingly, the AF9 D546R mutant protein is still capable of immunoprecipitating the low affinity Dot1L (Site 1) (**Figure 3-3A**).

We also employed a previously described AF9 D544R mutation. We previously used this mutation in structure-function studies to show that AF4 recruitment to AF9 is essential for MLL-AF9 leukemogenesis (Leach et al., 2013; Lokken et al., 2014). D544 is also a part of the AF9 β-hairpin (β2) and makes direct contacts with DOT1L S882 (**Figure 3-1A**). Interestingly, the D544R mutation shows a pronounced effect on binding to DOT1L Site 3 and a limited

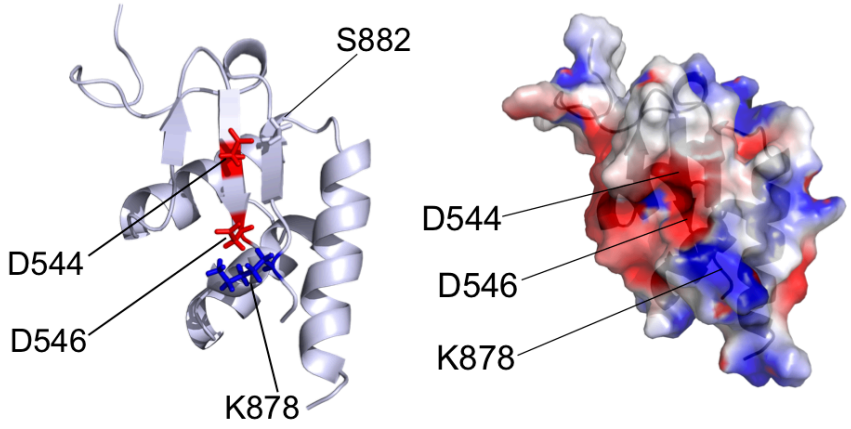
[3] Coimmunoprecipitation data in this section were provided by Nicholas Achille and Alyson Lokken in Dr. Nancy Zeleznik-Le's laboratory.

effect on binding to DOT1L Site 2 (**Figures 3-1B and 3-1C**). Consistent with this, the D544R mutant protein has been shown to still pull down DOT1L (Site 1) (Lokken et al., 2014), has a slightly diminished ability to pull down DOT1L (Sites 2 and 3) (**Figures 3-2A and 3-2B**), and still interacts with full length Dot1L (**Figure 3-3B**). As the binding of wildtype and mutant AF9 proteins with the DOT1L (Site 1) is extremely weak, we were unable to quantify any differences in binding to this site using our FP assays (**Figure 3-1C**).

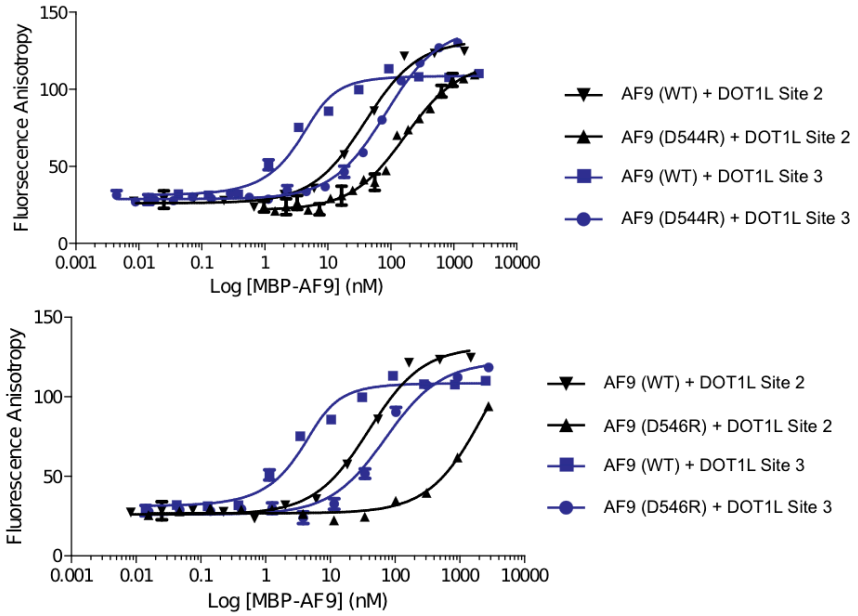
The structure of the DOT1L-AF9 complex that we have determined, and our previously described AF4-AF9 structure (Leach et al., 2013), show that both AF4 and DOT1L bind in the same site on AF9, consistent with previous biochemical studies showing that the binding of the two is mutually exclusive (Biswas et al., 2011; Yokoyama et al., 2010). Based on the similarity of the NMR spectra of AF9 complexes with BCoR and CBX8 (Leach et al., 2013), it is likely these partners also bind in the same site. Due to this, it is challenging to identify point mutations in AF9 that can selectively inhibit the binding of specific partner proteins. Both of our AF9 mutants indeed have effects on the binding of other AF9 interacting partners (**Figure 3-4**). As CBX8 peptide binding is too weak to measure, we used a CBX8 (Valine Mutant) (Leach et al., 2013) to gain insights into CBX8 binding to MBP-AF9 (WT) and the mutants. Substitution of an alanine within the β-strand of CBX8 (see **Chapter 4**) to a Valine leads to a dramatic increase in binding affinity with MLL-AF9 (WT) (**Figure 3-4A**). Notably we see an increase in full-length CBX8 binding with our D546R mutant (**Figure 3-4B** see Chapter 4 for further discussion). As binding to DOT1L and AF4 are presumed to

be most critical for gene activation, it is important to note that the effects of the D544R and D546R mutations on AF4 binding are similar, so a comparison of the biological effects of these two mutations in the context of MLL-AF9 should give meaningful insights into the role of recruitment of DOT1L in particular.

A



B



C

Binding Partner (Kd in nM)	MBP-AF9 (WT)	MBP-AF9(D544R)	MBP-AF9(D546R)
DOT1L Site 1	> 2000	> 2000	> 2000
DOT1L Site 2	26.7 ± 6.7	139 ± 10	> 2000
DOT1L Site 3	1.56 ± .02	55.2 ± 8.2	137 ± 27

Figure 3-1. Identification of mutations that disrupt the DOT1L-AF9 interactions.

A) Left: Cartoon representation of the DOT1L-AF9 structure showing D544 and D546, located on the AF9 β hairpin, that make direct interactions with DOT1L. Right: Electrostatic surface representation depicting that D546 (AF9) and K878 (DOT1L) make a charge-charge interaction.

B) Results of fluorescence polarization assays for determination of the K_d values for binding of MBP-AF9 AHD (WT), MBP-AF9 AHD (D544R) (Top), and MBP-AF9 AHD (D546R) (Bottom), to DOT1L binding motifs 2 and 3.

C) Table of K_d values for binding of MBP-AF9 AHD (WT), MBP-AF9 AHD (D544R), and MBP-AF9 AHD (D546R) to the DOT1L binding motifs. MBP-AF9 (D546R) significantly affects binding of both DOT1L high binding sites 2 and 3, whereas MBP-AF9 (D544R) has a more significant effect only on Site 3. DOT1L (Site 1) peptide had very weak binding with the wildtype protein and both of the AF9 mutations.

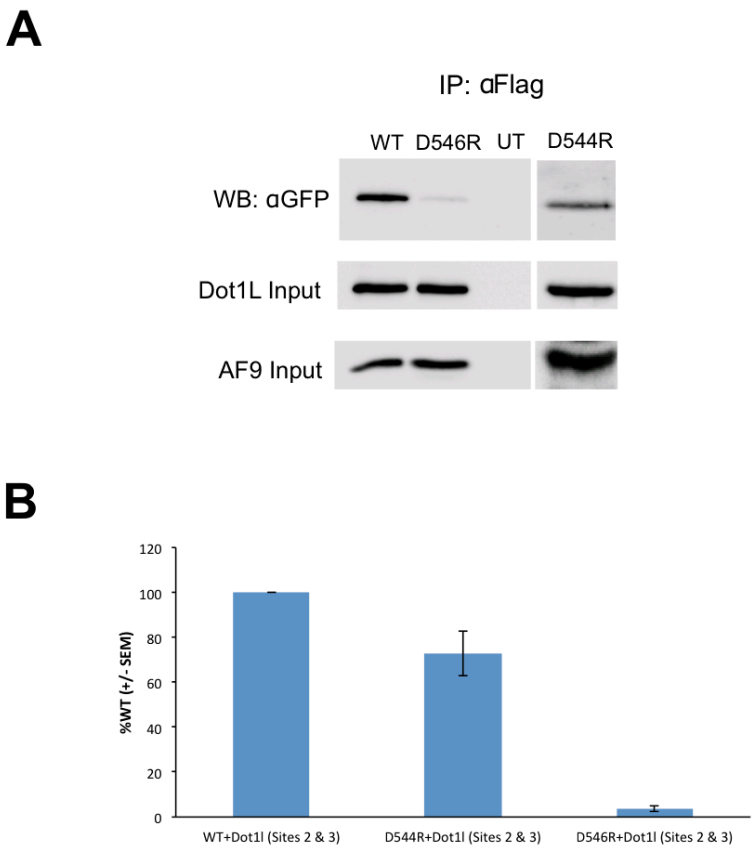


Figure 3-2. Co-IP data of the D546R and D544R AF9 mutants with the high affinity sites of DOT1l (Sites 2 and 3).

HEK293 cells were cotransfected with FLAG-AF9 (WT or mutant) and GFP-DOT1l construct containing the high affinity binding sites (828-1095). Agarose-conjugated FLAG antibodies were used to immunoprecipitate FLAG-AF9 and FLAG-purified proteins were immunoblotted using an anti-GFP antibody. **A)** The D546R mutant significantly reduces binding to the high affinity DOT1l sites. **B)** The pixel intensity of the immunoprecipitated GFP-tagged protein was normalized to FLAG-AF9 input, and set relative to wild type AF9.

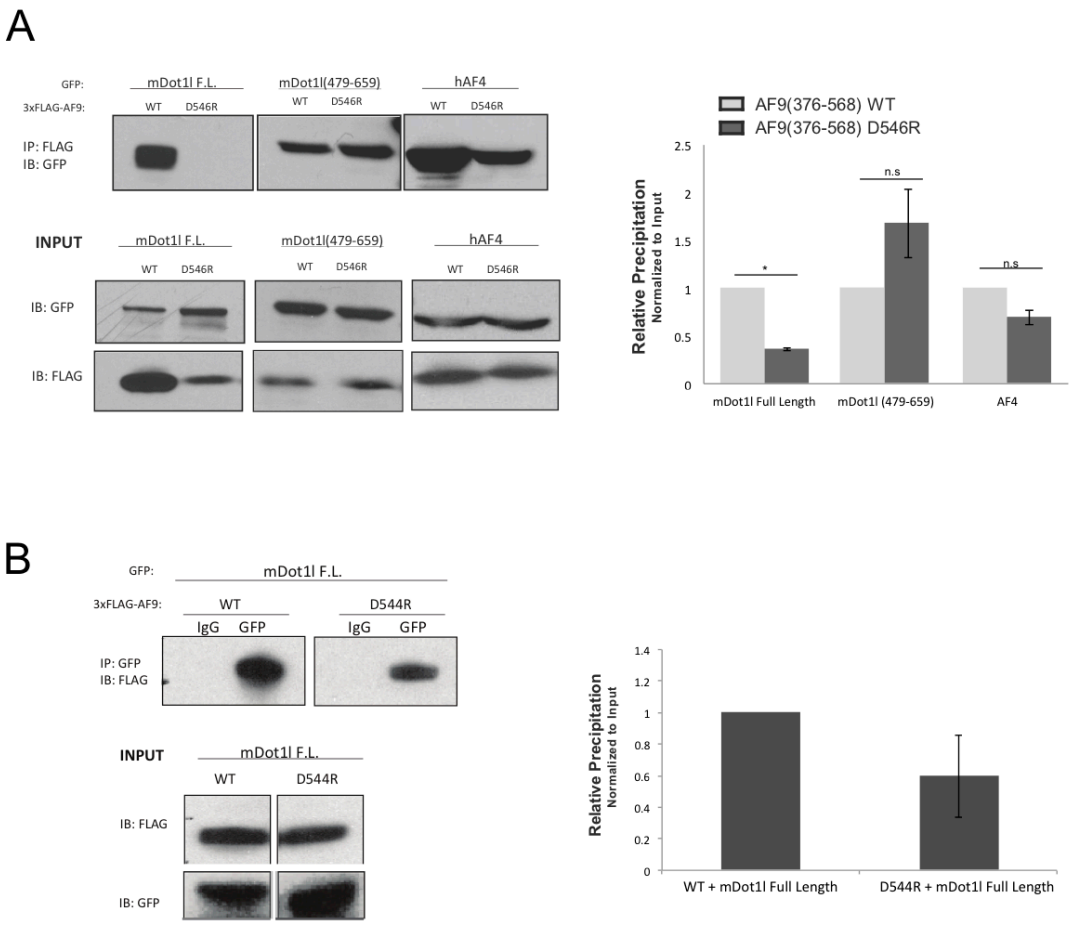


Figure 3-3. Co-IP data of D546R and D544R AF9 mutants with full length DOT1L.

A) Coimmunoprecipitation data showing that the AF9 (D546R) mutant has significantly decreased binding to full length Dot1l. D546R still binds to Dot1l (Site 1) denoted as mDot1l(479-659) and AF4.

B) Coimmunoprecipitation data showing that the AF9 (D544R) mutant still binds to full length Dot1l.

A

Binding Partner (Kd in nM)	MBP-AF9 (WT)	MBP-AF9(D544R)	MBP-AF9(D546R)
AF4	0.17 ± 0.05	9.6 ± 1.9	4.7 ± 3.5
CBX8	> 900	> 2000	> 2000
CBX8 (Valine Mutant)	11 ± 1	50.8 ± 5.7	57.4 ± 8.2
BCoR	24 ± 8	> 3000	> 3000

B

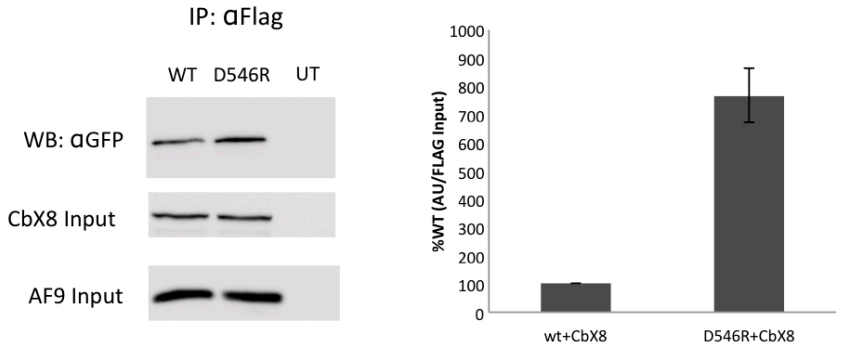


Figure 3-4. D544R and D546R AF9 mutations are not specific for disrupting DOT1L binding.

A) Fluorescently labeled peptides of AF4, CBX8, and BCoR, were titrated with MBP-AF9 AHD (WT) and MBP-AF9 AHD (D544R), MBP-AF9 AHD (D546R) mutants. AF4 is similarly affected with both of the MBP-AF9 mutants. The CBX8 (Valine Mutant) appears to show only a 5-fold decrease in binding with both mutants.

B) HEK293 cells were cotransfected with FLAG-AF9(WT or D546R mutant) and GFP-CBX8 constructs. Coimmunoprecipitation data shows increased binding of D546R to full-length CBX8 compared to WT.

3.2.2. Identification and biochemical characterization of DOT1L mutations

In order to complement the biological readouts we obtained with the AF9 D544R and D546R mutations, we have also characterized mutations in DOT1L that selectively inhibit binding to each of the three AF9 binding motifs. Mutation of two DOT1L hydrophobic residues, V881 and I883, that are buried within the protein-protein interface (**Figure 3-5A**), to alanine (**Figure 3-5B**), denoted as V881A, I883A, completely abrogates binding of DOT1L (Site 3) to AF9 (**Figure 3-5C**). To disrupt the binding to each of the remaining DOT1L binding sites we similarly made mutations of the corresponding DOT1L hydrophobic residues at Site 1 (L640A, I642A) and Site 2 (I867A, I869A). With each of these individual sets of mutations, we were unable to co-purify the separate DOT1L complexes with AF9 (data not shown). Consequently, we can selectively disrupt binding of each of the individual DOT1L binding sites to AF9 and preserve the interaction of the unaffected binding sites (**Figures 3-5D** and **3-5E**). To this end, we created a series of DOT1L mutants that disrupted the three DOT1L-AF9 interactions alone as well as in various combinations.

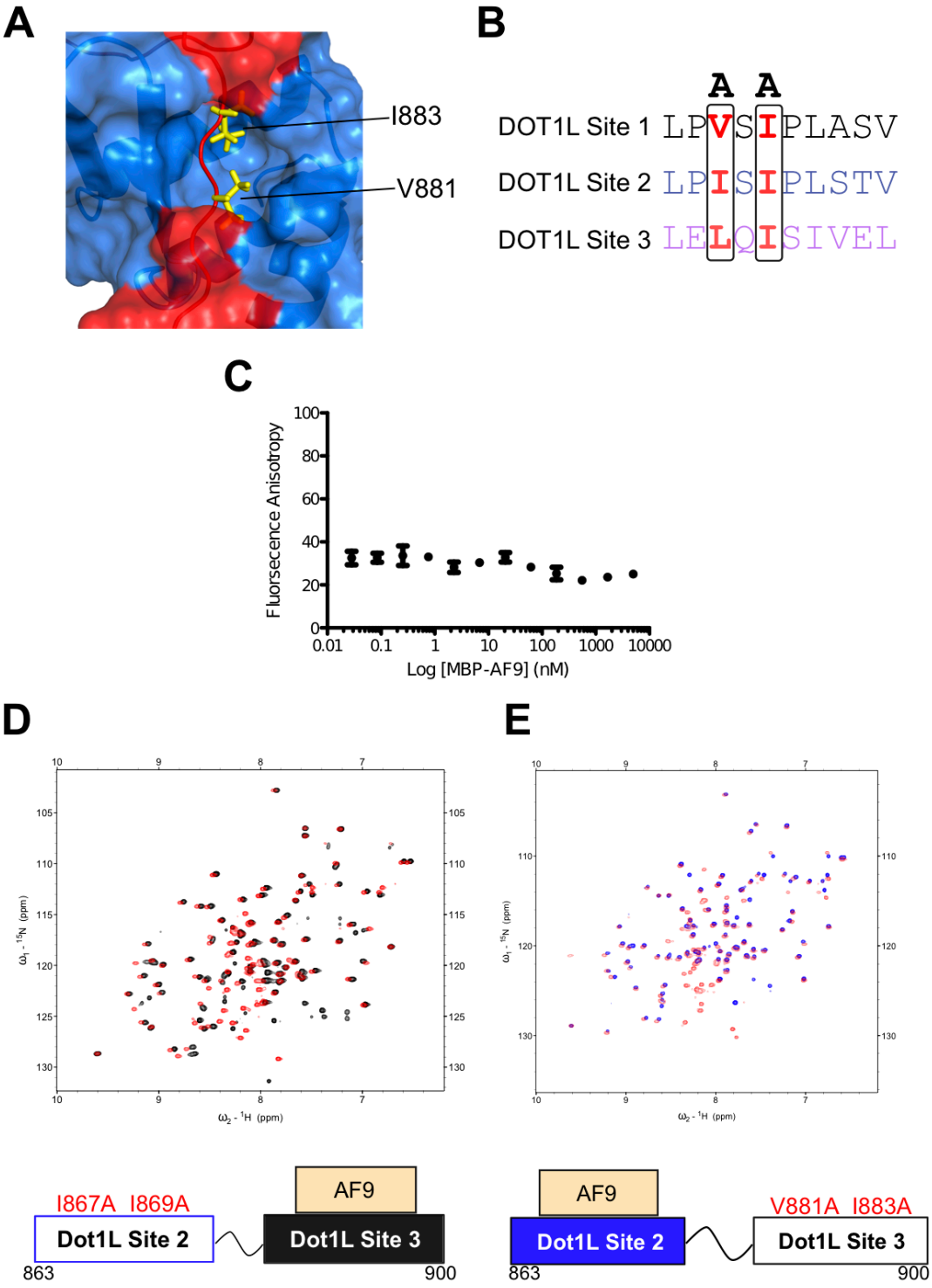

A
I883
V881
B
A A
DOT1L Site 1 LPVSIPLASV
DOT1L Site 2 LPISIPLSTV
DOT1L Site 3 LELQISIVEL
C
Fluorescence Anisotropy
Log [MBP-AF9] (nM)
D
E
I867A I869A
Dot1L Site 2
AF9
Dot1L Site 3
863
900
AF9
Dot1L Site 2
V881A I883A
Dot1L Site 3
863
900

Figure 3-5. Alanine mutants can specifically disrupt each of the DOT1L binding sites.

A) Surface representation of the DOT1L (red) – AF9 (blue) complex. Sidechains of two buried hydrophobic residues form DOT1L, V881 and I883, are shown in yellow.

B) Two alanine mutations of similarly positioned hydrophobic residues within each DOT1L binding site used to disrupt AF9 binding.

C) Results of fluorescence polarization assay for MBP-AF9 AHD titrated into fluorescently tagged DOT1L peptide with (V881A, I883A), showing no binding.

D) A complex of the entire repeat motif of DOT1L (aa. 863-900), with the alanine mutants I867A, I869A located in DOT1L (Site 2), was co-expressed with AF9 and purified. ^{15}N-^{1}H HSQC of the DOT1L repeat motif with these alanine mutations (red) overlaid with ^{15}N-^{1}H HSQC of DOT1L (Site 3) – AF9 (black). As the I867A, I869A mutations are located within DOT1L Site 2 of the repeat motif, only DOT1L binding Site 3 in the repeat motif can bind to AF9. The ^{15}N-^{1}H HSQC of the mutant repeat motif is almost identical to that of DOT1L (Site 3) – AF9.

E) Similarly, a complex of the entire repeat motif of DOT1L (aa. 863-900) with the alanine mutants V881A, I883A located in DOT1L (Site 3), was co-expressed with AF9 and purified. ^{15}N-^{1}H HSQC of the DOT1L repeat motif with these alanine mutations (red) overlaid with the ^{15}N-^{1}H HSQC of DOT1L (Site 2) – AF9 (blue). As the V881A, I883A mutations are located within DOT1L Site 3 of the repeat motif, only DOT1L binding Site 2 in the repeat motif can bind to AF9. Thus, the ^{15}N-^{1}H HSQC of this mutant repeat motif is almost identical to that of DOT1L (Site 2) – AF9.

3.3. Disruption of DOT1L recruitment via high affinity binding sites to MLL-AF9 leads to dramatic losses in hematopoietic transformation

In vitro serial replating colony formation assays have been used extensively to determine the leukemogenic transformation potential of hematopoietic cells. When wildtype bone marrow progenitor cells isolated from mice are cultured in semi-solid media containing the appropriate cytokines, they generally form colonies within the first week of plating, then differentiate and die within a period of several weeks. However, bone marrow progenitor cells that are retrovirally infected with MLL fusion proteins have an ability to proliferate extensively and can be serially replated in methylcellulose colony assays (**Figure 3-6A**). To this end, we used serial replating colony formation assays to assess the biological impact of differentially disrupting the DOT1L interactions with MLL-AF9.

3.3.1. Serial replating assays with MLL-AF9 (D544R) and MLL-AF9 (D546R)[4]

Briefly, hematopoietic progenitor cells isolated from mouse bone marrow were transduced with retrovirus expressing either wildtype MLL-AF9 (WT), or the mutants, MLL-AF9 (D544R) and MLL-AF9 (D546R). Cells were then serially replated on a weekly basis over a period of four weeks. Colony forming ability shows a gradient between cells expressing MLL-AF9 (WT), which efficiently replated for four weeks, the MLL-AF9 (D544R) mutant, which exhibited reduced colony formation in agreement with our recent publication (Lokken et al., 2014), and MLL-AF9 (D546R), which showed a more dramatic loss in colony formation

[4] Data in this section were provided by Nicholas Achille and Alyson Lokken in Dr. Nancy Zeleznik-Le's laboratory.

(**Figure 3-6B**). Both of the mutants showed more diffuse colonies comprised of fewer cells compared to the dense, compact colonies formed by MLL-AF9 (WT) cells (**Figure 3-6C**). Cytospin and Wright-Giemsa staining indicate that MLL-AF9 (WT) expressing cells exhibit a blast-like morphology and cells expressing MLL-AF9 (D546R) appear to be differentiated (**Figure 3-6C**).

3.3.2. Serial replating assays with DOT1L Alanine Mutants[5]

A complementary experiment was performed to determine whether there were functional differences between the individual AF9-binding sites of DOT1L using our DOT1L alanine mutants. Bone marrow progenitor cells isolated from conditional *Dot1l* deletion mice transformed with MLL-AF9 (Chang et al., 2010) were co-transduced with retroviruses expressing GFP-wildtype DOT1L or GFP-mutant DOT1L plus either mCherry-Cre or mCherry alone. Cre-mediated deletion creates a *Dot1l* null allele, abolishing Dot1l methyltransferase activity and its function. Subsequently, GFP/mCherry double positive cells were sorted and assessed for colony forming ability (**Figure 3-7A**).

Deletion of endogenous Dot1l significantly decreased colony forming ability compared to addback of WT Dot1l (**Figure 3-7B**), corroborating results that the Zeleznik-Le lab has shown previously (Chang et al., 2010). Intriguingly, the remaining colonies from the KO result from the expansion of cells that escaped Cre-mediated deletion (**Figure 3-7C**, MLL-AF9 +Cre: Lane 4). Because there is such strong selective pressure for MLL-AF9-transformed cells to retain

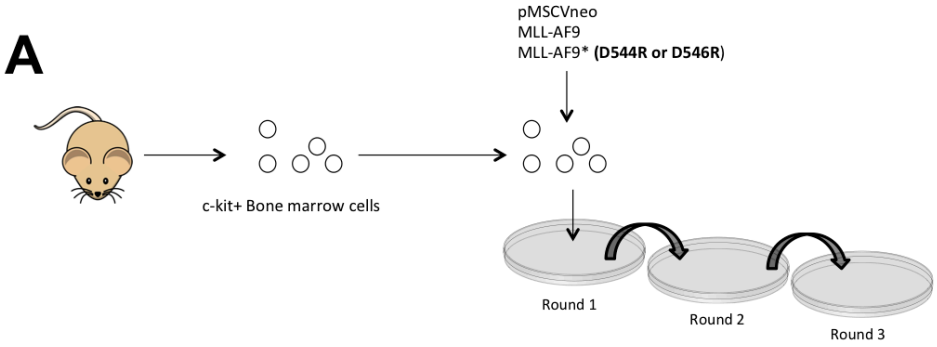
[5] Data in this section were provided by Nicholas Achille in Dr. Nancy Zeleznik-Le's laboratory.

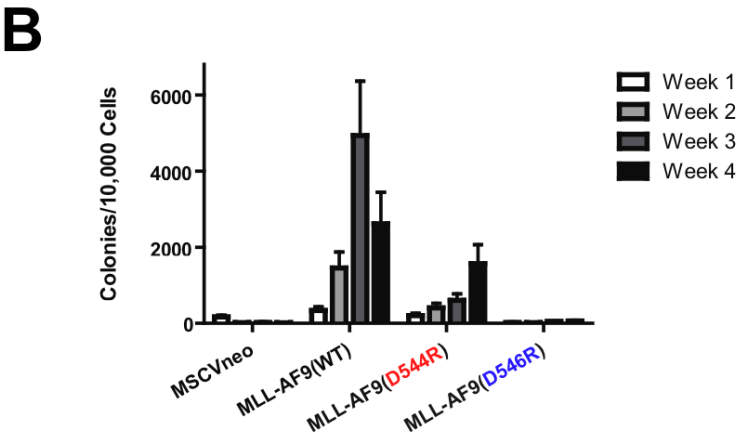
functional Dot1l expression, unless exogenous functional DOT1L is provided (**Figure 3-7C**, MLL-AF9 +Cre: Lane 5), only cells that retain endogenous Dot1l grow and expand, as demonstrated by the presence of the undeleted *Dot1l* allele (**Figure 3-7C**). This dependence was not true for E2A-PBX-transformed cells, another leukemia associated oncogene, as the Zeleznik-Le lab previously showed that the E2A-PBX fusion is not dependent on Dot1l. (**Figures 3-7C** and **3-7D**).

To further determine the functional significance of each AF9-binding site in DOT1L, mutant versions of DOT1L were exogenously provided in combination with deletion of the endogenous *Dot1l*. Mutations that disrupt each AF9-binding site in DOT1L individually cause reduced colony formation, but only the mutant DOT1L (Site 2) rises to statistical significance (**Figure 3-8A**). With each single site mutant, there is no selection bias for cells retaining the endogenous *Dot1l* allele (**Figure 3-8B**, lanes 1 through 3).

Simultaneous mutation of two sites proved to be particularly interesting. Mutation of Sites 1+2 showed no decrease in colony forming ability compared to wildtype DOT1L, but rather an increase (**Figure 3-8C**). Mutation of Sites 1+3 was no different than when no exogenous DOT1L was added (**Figure 3-8C**). Although mutation of Sites 2+3 did not demonstrate a statistically different colony-forming ability, remaining colonies were due to expansion of cells retaining the endogenous *Dot1l* allele (**Figure 3-8B**, lane 6). Thus, DOT1L with both Sites 2+3 mutated does not confer colony-forming capacity to MLL-AF9-transformed cells. Simultaneously blocking all three DOT1L binding sites with

alanine mutations shows colonies with differentiated morphology (**Figure 3-8D**) and a loss in colony formation similar to that when no exogenous DOT1L was provided (**Figure 3-8C**). This demonstrates that high affinity DOT1L binding to AF9 is necessary for the colony forming ability of MLL-AF9, and that the multiple binding sites act in concert with one another.


A
pMSCVneo
MLL-AF9
MLL-AF9* (D544R or D546R)
c-kit+ Bone marrow cells
Round 1
Round 2
Round 3


B
Colonies/10,000 Cells
6000
4000
2000
0
Week 1
Week 2
Week 3
Week 4
MSCVneo
MLL-AF9(WT)
MLL-AF9(D544R)
MLL-AF9(D546R)

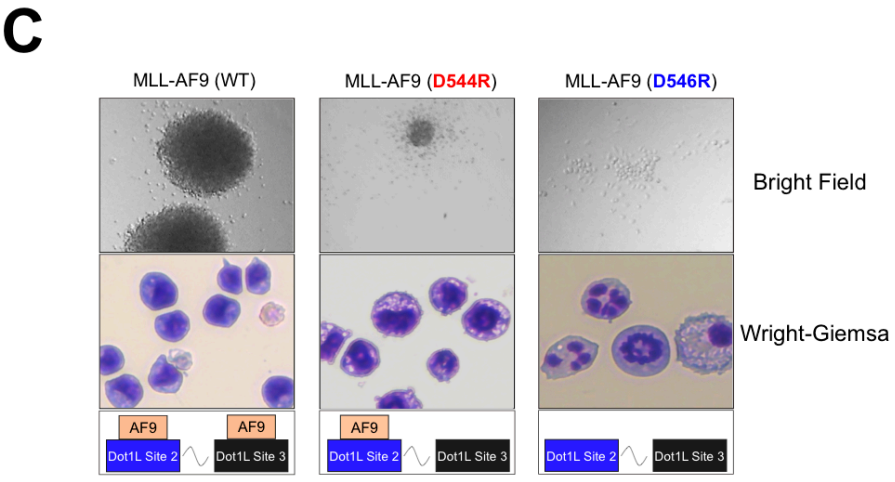
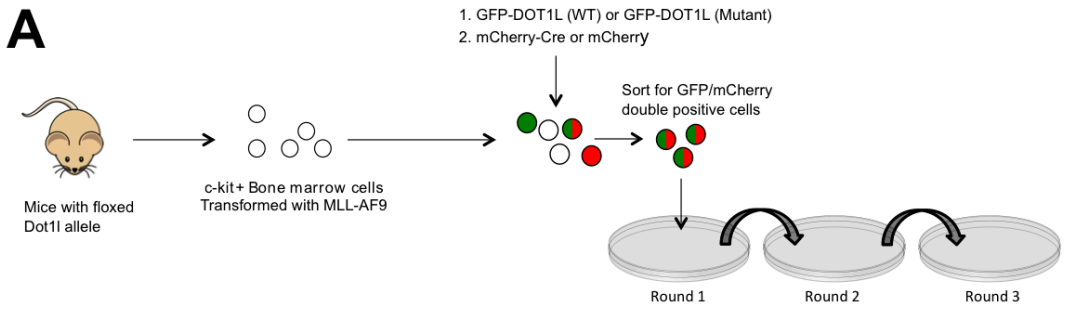

C
MLL-AF9 (WT)
MLL-AF9 (D544R)
MLL-AF9 (D546R)
Bright Field
Wright-Giemsa
AF9
AF9
Dot1L Site 2
Dot1L Site 3
AF9
Dot1L Site 2
Dot1L Site 3
Dot1L Site 2
Dot1L Site 3

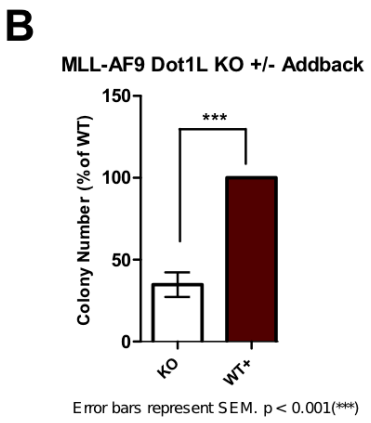
Figure 3-6. Mutants in MLL-AF9 show that the level of DOT1L recruitment to MLL-AF9 defines the degree of serial replating capability.

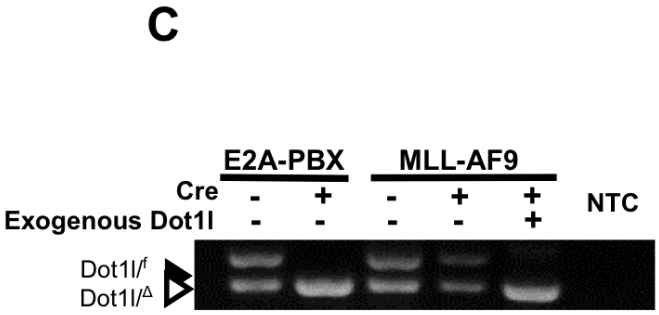
A) Diagram of the serial replating assay.

B) Results of serial replating assays for MLL-AF9 (WT), MLL-AF9 (D544R), and MLL-AF9 (D546R). Cells expressing MLL-AF9 (WT) consistently replated over a period of 4 weeks. Cells expressing MLL-AF9 (D544R) show a reduction in serial replating ability. Cells expressing MLL-AF9 (D546R) show a complete abrogation of serial replating ability

C) Bright field and Wright-Giemsa images of MLL-AF9 (WT) colonies show large, tight colonies (top) containing mostly cells with a blast-like morphology (bottom), whereas MLL- AF9(D544R) have more dispersed colonies, some with tight centers, and more differentiated cells. Strikingly, MLL-AF9(D546R) colonies are completely diffuse, with almost exclusively differentiated cells.


A
1. GFP-DOT1L (WT) or GFP-DOT1L (Mutant)
2. mCherry-Cre or mCherry
Sort for GFP/mCherry double positive cells
Mice with floxed Dot1l allele
c-kit+ Bone marrow cells Transformed with MLL-AF9
Round 1
Round 2
Round 3


B
MLL-AF9 Dot1L KO +/- Addback
Colony Number (% of WT)
150
100
50
0
***
KO
WT+
Error bars represent SEM. p < 0.001(***)


C
E2A-PBX
MLL-AF9
Cre
- + - + +
Exogenous Dot1l
- - - - +
NTC
Dot1l/f
Dot1l/Δ

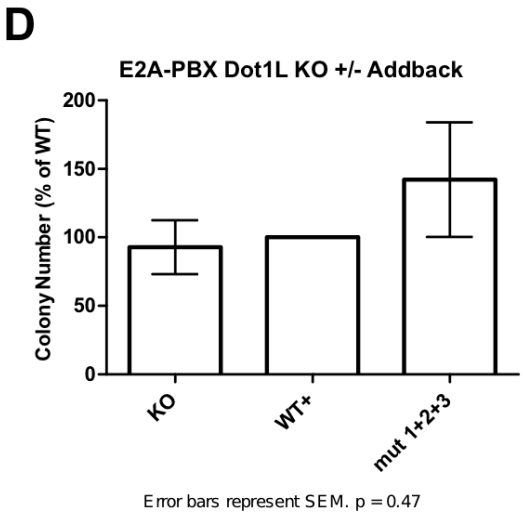

D
E2A-PBX Dot1L KO +/- Addback
Colony Number (% of WT)
200
150
100
50
0
KO
WT+
mut 1+2+3
Error bars represent SEM. p = 0.47

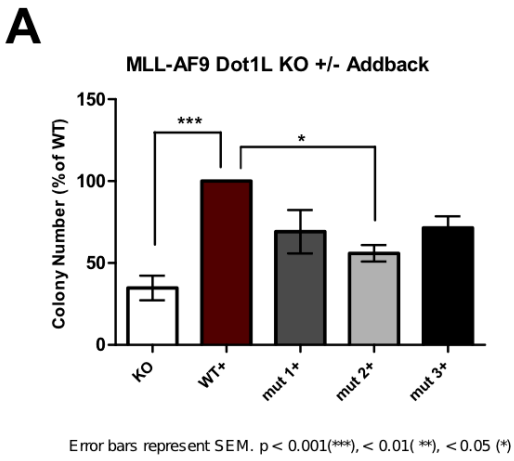
Figure 3-7. Deletion of endogenous Dot1l significantly decreased colony forming ability in MLL-AF9.

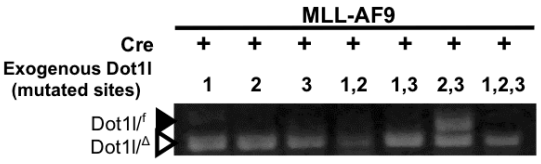
A) Diagram of serial replating assays with DOT1L mutants.

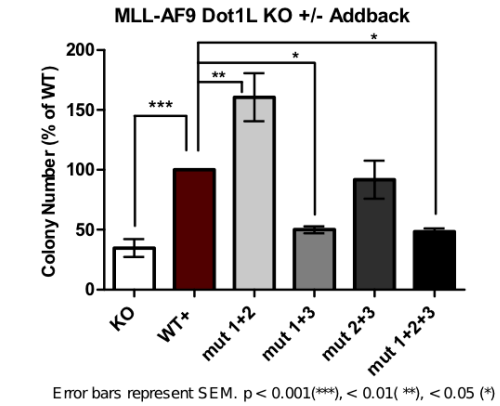
B) Results of serial replating assays show that deletion of endogenous Dot1l significantly decreased colony forming ability compared to addback of WT Dot1l.

C) *Dot1l* genomic status was examined by PCR at day 7 after methylcellulose culture of MLL-AF9- or E2A-PBX-transformed *Dot1l*$^{fl/\Delta}$ cells, expressing the indicated exogenous proteins. Dark arrowhead, floxed allele (510 bp); open arrowhead, deleted allele (378 bp). NTC, non-template control.

D) Results of methylcellulose colony assays from E2A-PBX-transformed *Dot1l*$^{fl/\Delta}$ cells, co-transduced with wild type or mutant Dot1l, plus Cre recombinase. E2A- PBX fusion is not dependent on Dot1l and shows no significant differences in colony number with Dot1l KO, WT or the three site Dot1l mutant (Sites 1 + 2 + 3). Thus, our Dot1l three site mutant does not have any unintended effects.


A
MLL-AF9 Dot1L KO +/- Addback
Colony Number (% of WT)
150
100
50
0
***
*
KO
WT+
mut 1+
mut 2+
mut 3+
Error bars represent SEM. p < 0.001(***), < 0.01(**), < 0.05 (*)


B
MLL-AF9
Cre
+ + + + + + +
Exogenous Dot1l
(mutated sites)
1 2 3 1,2 1,3 2,3 1,2,3
Dot1l/f
Dot1l/Δ


C
MLL-AF9 Dot1L KO +/- Addback
Colony Number (% of WT)
200
150
100
50
0
***
**
*
*
KO
WT+
mut 1+2
mut 1+3
mut 2+3
mut 1+2+3
Error bars represent SEM. p < 0.001(***), < 0.01(**), < 0.05 (*)

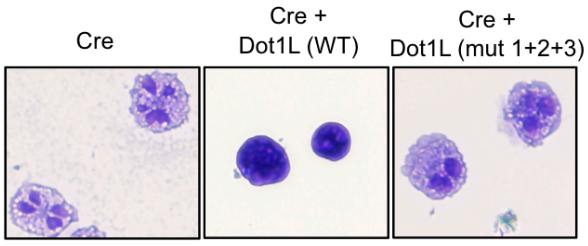

D
Cre
Cre +
Dot1L (WT)
Cre +
Dot1L (mut 1+2+3)

Figure 3-8. Mutants in DOT1L show that the level of DOT1L recruitment to MLL-AF9 defines the degree of serial replating capability.

A and C) Results of methylcellulose colony assays from MLL-AF9-transformed *Dot1l*$^{fl/D}$ cells, co-transduced with wild type or mutant Dot1l, either with or without Cre recombinase.

B) *Dot1l* genomic status was examined by PCR at day 7 after methylcellulose culture of MLL-AF9 transformed *Dot1l*$^{fl/\Delta}$ cells, expressing the indicated exogenous proteins. Dark arrowhead, floxed allele (510 bp); open arrowhead, deleted allele (378 bp). NTC, non-template control.

D) Wright-Giemsa images of cells from Dot1l complementation methylcellulose colony experiment, showing ability of exogenous Dot1l (WT), but not Dot1l (mut 1+2+3) to rescue blast-like morphology of MLL-AF9-transformed *Dot1l*$^{fl/\Delta}$ cells.

3.4. MLL-AF9 (D544R) and MLL-AF9 (D546R) display distinct patterns of loss of H3K79me2 and H3K79me3 on a select set of genes[6] [7]

To assess the epigenetic effects of disrupting the multiple DOT1L and MLL-AF9 interactions, we used chromatin immunoprecipitation followed by next-generation sequencing (ChIP-Seq) to identify the genome-wide localization of H3K79me2 and H3K79me3 in primary hematopoietic progenitor cells expressing wildtype or mutant MLL-AF9. We analyzed H3K79me2 and H3K79me3 profiles and compared both MLL-AF9 (D544R) and MLL-AF9 (D546R) mutants to wildtype MLL-AF9 (WT). There were modest changes in individual genes for H3K79me2 and H3K79me3 marks when comparing the MLL-AF9 (D544R) mutant to MLL-AF9 (WT), but none rising to the level of significance (FDR < 0.1). In contrast, our MLL-AF9 (D546R) data show that with complete disruption of the high affinity DOT1L interaction to MLL-AF9, 44 genes display a significant loss of the H3K79me2 mark, and 42 genes show a reduction of the H3K79me3 mark (FDR < 0.1) compared to MLL-AF9 (WT) data (**Figure 3-9**). A majority (31 out of 44 for H3K79me2 and 30 out of 42 for H3K79me3) of these genes are direct targets of MLL-AF9, as defined by a previous MLL-AF9 ChIP-Seq study (**Figure 3-9A**) (Bernt et al., 2011). A number of the identified genes with both decreased H3K79me2 and H3K79me3 have been shown to play a role in MLL-

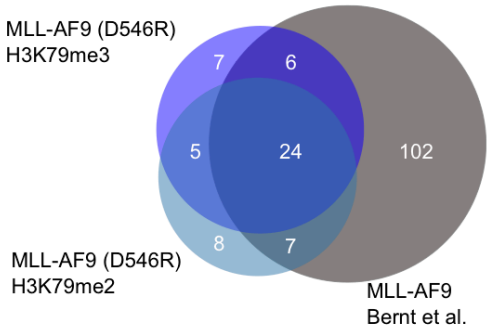
[6]Chip-Seq experiments were conducted with Jeremy Thrope in Dr. Mazhar Adli's laboratory.

[7]Ritambhara Singh, also in Dr. Adli's laboratory, processed the raw ChIP-seq data to generate the read count tables that I could subsequently use for downstream analyses.

rearranged leukemia such as *Hoxa9*, *Meis1*, *Runx2,* and *Jmjd1C* (**Figure 3-10 and Table 3-1**). There are not many genes with differentially affected changes in only H3K79me2 or H3K79me3, but not both, with the MLL-AF9(D546R) mutant. Of note is *Cdk6*, which shows a significant decrease in the H3K79me3 mark, but not H3K79me2, and has recently been shown to be important for MLL-rearranged leukemia (**Figure 3-10F**) (Placke et al., 2014). In contrast, *Eya1* shows decreased H3K79me2 without significant change in H3K79me3, to the level of detection, (**Figure 3-10E**), is overexpressed in MLL leukemia (Wang et al., 2011).

Mapping the genome-wide distribution of H3K79me2 and H3K79me3 marks across the set of genes that we found to be changed as a result of disrupting the DOT1L interaction with MLL-AF9, demonstrated no significant decrease in H3K79me2 levels for MLL-AF9 (D544R), but a substantial reduction for MLL-AF9 (D546R) (**Figure 3-11A**). Interestingly, a different pattern of effects on H3K79me3 are observed. The H3K79me3 data show that MLL-AF9 (D544R) levels significantly decrease in the H3K79me3 profile plot compared to wildtype (Wilcoxon test p-value = 0.02) (**Figure 3-11B**) and a complete loss of H3K79me3 is observed for MLL-AF9 (D546R). Thus, there is a differential effect of the loss of binding of MLL-AF9 to one or two high affinity DOT1L binding motifs on H3K79me2 and H3K79me3 levels at target genes.

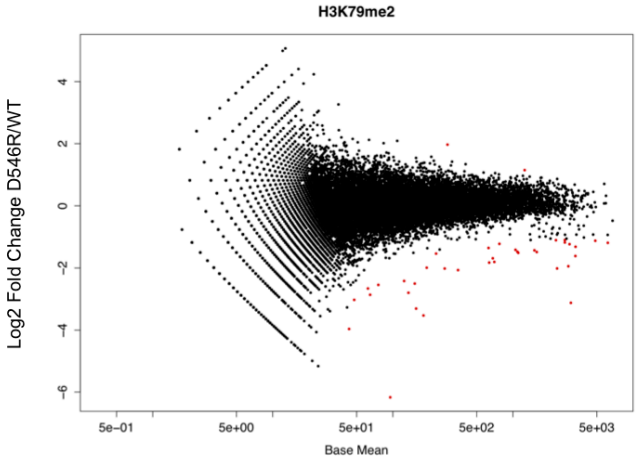
A



24 Common Genes

AI314180	PDE3B
ARID2	PHIP
CDK13	RDX
DCUN1D1	RUNX2
ERBB2IP	SENP6
FCHO2	SGK3
HOXA10	SIX1
HOXA3	TNPO1
HOXA7	TSC22D2
HOXA9	TSHZ1
JMJD1C	ZBTB25
MEIS1	ZFP800

B



C

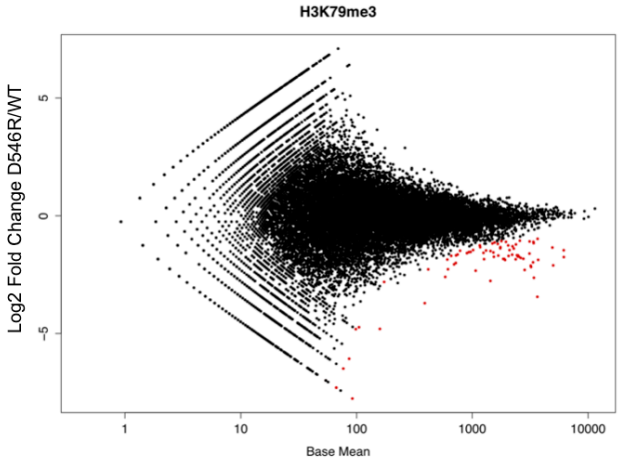


Figure 3-9. MLL-AF9 (D544R) and MLL-AF9 (D546R) display distinct patterns of loss of H3K79me2 and H3K79me3 on a select set of genes.

A) Venn diagram overlaying gene sets showing significant loss of H3K79me2 and H3K79me3 marks upon comparison of MLL-AF9 (D546R) with MLL-AF9 (WT) (FDR < 0.1). Additionally overlaid is the set of genes identified previously by ChIP-Seq as direct MLL-AF9 targets (Bernt et al., 2011). Listed to the right are genes that overlap between the H3K79me2, H3K79me3, and MLL-AF9 direct target datasets.

B and C) MAplots showing the identification of changes in individual genes for H3K79me2 and H3K79me3 marks when comparing the MLL-AF9 (D546R) to MLL-AF9 (WT) (FDR < 0.1, labeled in red). We see significant differences in only a small subset of genes.

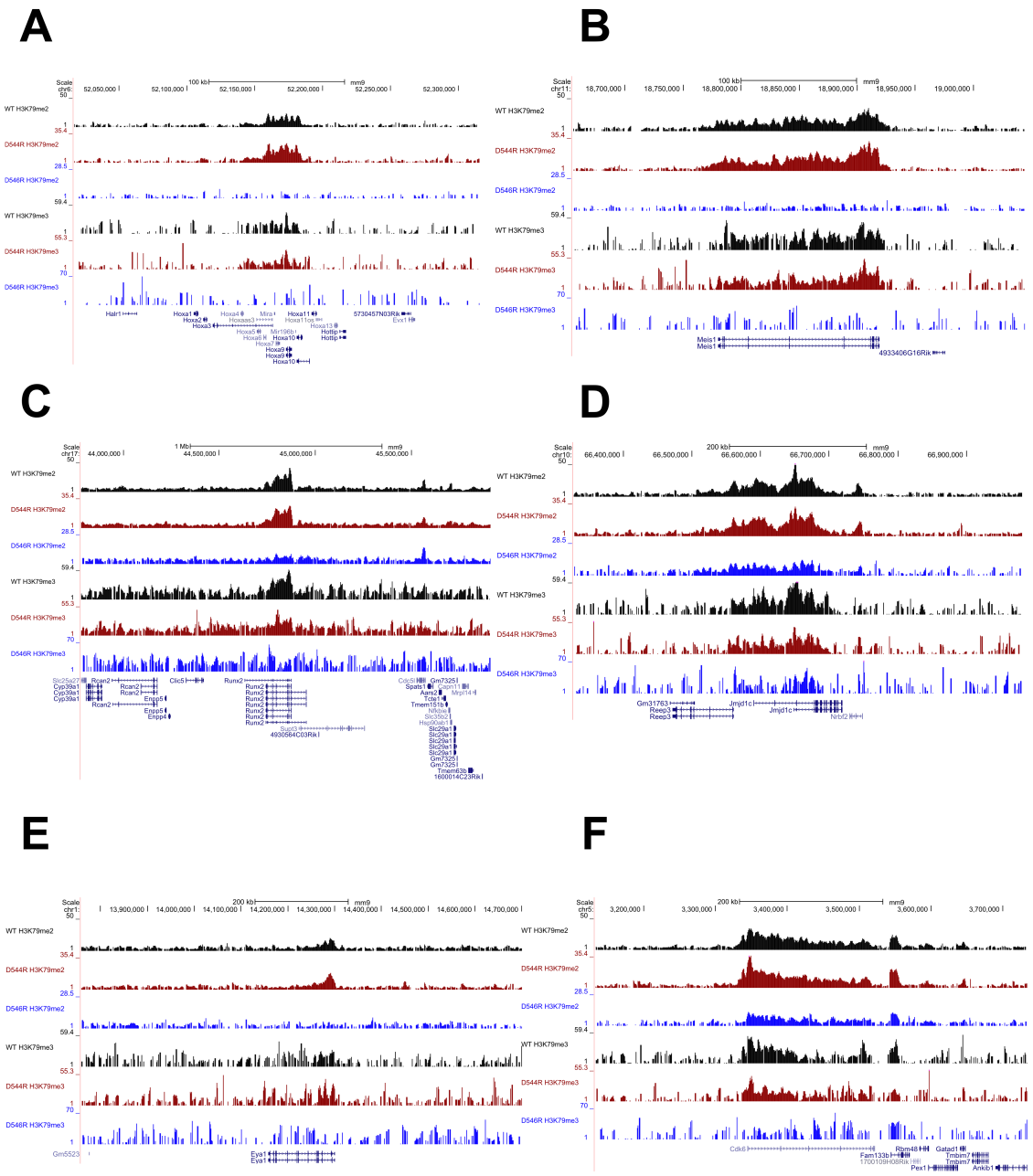


Figure 3-10. Representative ChIP-Seq track marks:

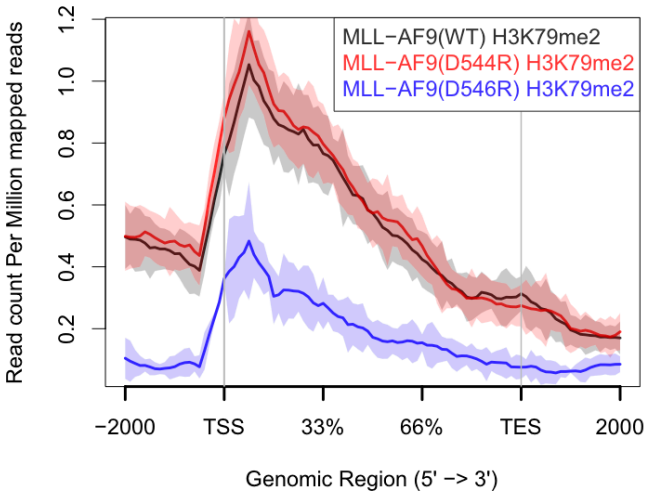
(Red Tracks – D544Rme2 and me3, Blue Tracks – D546Rme2 and me3)

A) *Hoxa9* **B)** *Meis1* **C)** *Runx2* **D)** *Jmjd1c* **E)** *Eya1* **F)** *Cdk6*

Genes showing a loss of H3K79me2 marks that are MLL-AF9 targets:
ARID1B
BAZ2B
CENPV
EYA1
MBNL1
PDS5A
RNF220
Genes showing a loss of H3K79me3 marks that are MLL-AF9 targets:
A630072M18RIK
ANKIB1
CDK6
NIPBL
SMC4
TOP2B
Genes showing a loss of both H3K79me2 and H3K79me3 that are not MLL-AF9 targerts:
CHD9
SKAP2
SMIM14
TMEM126A
ZEB2
Genes uniquely showing losses in H3K79me2:
2700086A05RIK
BCOR
CHIL3
ELF1
HSPA2
MEF2C
ST7
ZC3H12C
Genes uniquely showing losses in H3K79me3:
FMR1
MCM6
PELI1
QK
RAP2B
SLC16A1
ZEB2OS

Table 3.1 Complete lists of genes identified to have significant losses in H3K79me2 or H3K79me3 with the MLL-AF9 (D546R) mutant.

A


MLL−AF9(WT) H3K79me2
MLL−AF9(D544R) H3K79me2
MLL−AF9(D546R) H3K79me2
Read count Per Million mapped reads
0.2
0.4
0.6
0.8
1.0
1.2
−2000
TSS
33%
66%
TES
2000
Genomic Region (5' −> 3')

B

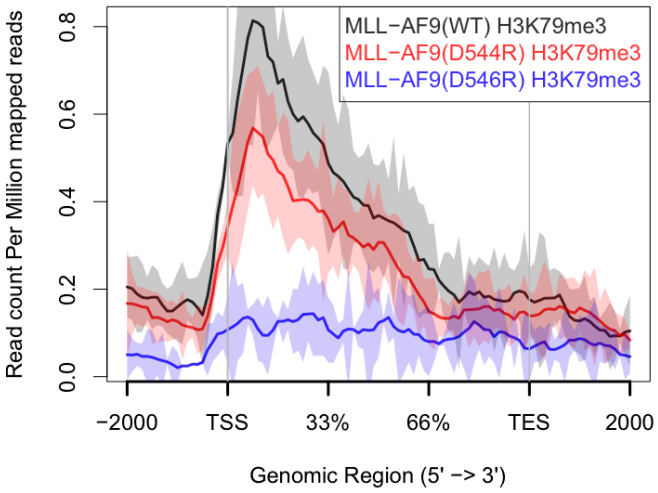

MLL−AF9(WT) H3K79me3
MLL−AF9(D544R) H3K79me3
MLL−AF9(D546R) H3K79me3
Read count Per Million mapped reads
0.0
0.2
0.4
0.6
0.8
−2000
TSS
33%
66%
TES
2000
Genomic Region (5' −> 3')

Figure 3-11. Genomic profiles of D546R H3K79me2 and H3K79me3 at the identified target genes.

A) H3K79me2 genomic profile at genes we identified to be significant in Figure 5A shows that MLL-AF9 (D546R) significantly reduces H3K79me2 marks whereas there is no difference in the profile between MLL-AF9 (D544R) and MLL-AF9 (WT).

B) H3K79me3 genomic profile at genes we identified to be significant in Figure 5A shows that MLL-AF9 (D546R) reduces H3K79me3 to background levels, even lower than observed for the H3K79me2 profile of this same mutant. There is also a significant difference in MLL-AF9 (D544R) and MLL-AF9 (WT) profiles (Wilcoxon test p value = 0.02).

3.5. Selected H3K79me2 and H3K79me3 genes display distinct patterns of gene expression in MLL-AF9 (D544R) and MLL-AF9 (D546R) mutant cells [8]

We next investigated whether the differential recruitment of Dot1l to MLL-AF9 and observed changes in H3K79 di- and tri-methylation correlated to changes in gene expression levels in selected genes. To assess this, gene expression levels were determined by quantitative RT-PCR (qRT-PCR) in primary bone marrow cells expressing MLL-AF9, and the MLL-AF9 mutants. We selected genes that represented a decrease in both H3K79me2 and me3 (*Hoxa9* and *Meis1*), a decrease in only H3K79me2 or H3K79me3 (*Eya1* and *Cdk6*, respectively), or no change in either (*Myb* and *Gfi1*), with our MLL-AF9 (D546R) mutant (**Figure 3-12A**). All of these genes are MLL-AF9 targets (Bernt et al., 2011) and have previously been shown to be involved in leukemogenesis (Khandanpour et al., 2013; Placke et al., 2014; Wang et al., 2011).

Interestingly, we observe different patterns of gene expression for these different selected gene classes (**Figure 3-12**). *Hoxa9* and *Meis1* both show losses in H3K79me2 and H3K79me3 (**Figure 3-10A and B**) and we observe a graded decrease in expression going from wildtype MLL-AF9 to MLL-AF9(D544R) to MLL-AF9(D546R) (**Figures 3-12B and C**). *Gfi1* and *Myb* show no change in H3K79me2 and H3K79me3, however they show different patterns of expression with the mutations. *Gfi1* shows no effect with D544R but decreased expression with D546R (**Figure 3-12D**) whereas *Myb* shows similar reductions in

[8] qRT-PCR experiments were conducted by Nicholas Achille in Dr. Zeleznik-Le's Laboratory

expression with both mutations (**Figure 3-12E**), suggesting additional mechanisms of regulation at these genes. *Cdk6* only shows losses in H3K79me3 and shows a similar decrease in expression for both D544R and D546R mutations (**Figure 3-12F**), consistent with a role for only H3K79me3 at this gene. Interestingly, *Eya1* shows only decreased H3K79me2 but displays the same graded reduction in expression with the mutations as observed for *Hoxa9* and *Meis1* (**Figure 3-12G**).

We additionally measured expression in the same genes for the set of DOT1L Alanine mutants (**Section 3.3.2**) as we did with the MLL-AF9 mutants. We thought that we would observe a graded decrease in expression contingent on the number of DOT1L binding sites blocked from binding to MLL-AF9. To our surprise we did not find any distinct patterns of gene expression (**Figure 3-13**). Only *Meis1* showed losses in expression with the DOT1L triple mutant, while all of the other genes showed an increase. These results were unexpected as we surmised that gene expression from the DOT1L triple Alanine mutants would be similar to the MLL-AF9 (D546R) mutants. Indeed, we cannot make any conclusive statements regarding gene expression with the single or double DOT1L mutants across all of the genes.

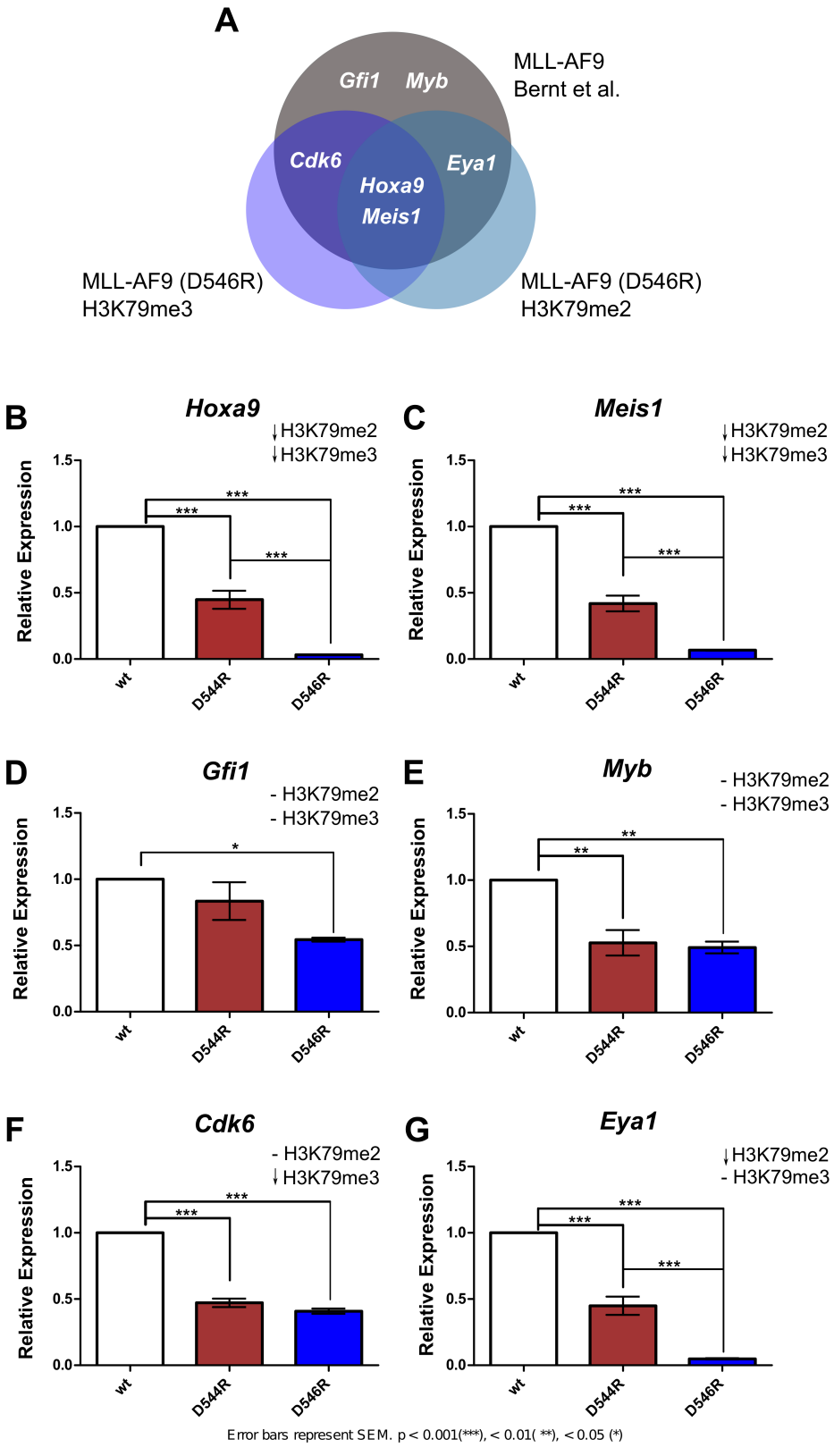

A
Gfi1
Myb
Cdk6
Eya1
Hoxa9
Meis1
MLL-AF9
Bernt et al.
MLL-AF9 (D546R)
H3K79me3
MLL-AF9 (D546R)
H3K79me2
B
Hoxa9
↓H3K79me2
↓H3K79me3
C
Meis1
↓H3K79me2
↓H3K79me3
D
Gfi1
- H3K79me2
- H3K79me3
E
Myb
- H3K79me2
- H3K79me3
F
Cdk6
- H3K79me2
↓H3K79me3
G
Eya1
↓H3K79me2
- H3K79me3
Relative Expression
wt
D544R
D546R
Error bars represent SEM. p < 0.001(***), < 0.01(**), < 0.05 (*)

Figure 3-12. Gene expression data using MLL-AF9 (WT), MLL-AF9 (D544R), and MLL-AF9 (D546R).

A) Venn diagram depicting genes that were selected for gene expression analyses. Gene expression data of **B)** *Hoxa9* **C)** *Meis1* **D)** *Gfi1* **E)** *Myb* **F)** *Cdk6* **G)** *Eya1* from murine bone marrow progenitor cells expressing MLL-AF9 (WT), MLL-AF9(D544R) or MLL-AF9(D546R) after one week in methylcellulose culture.

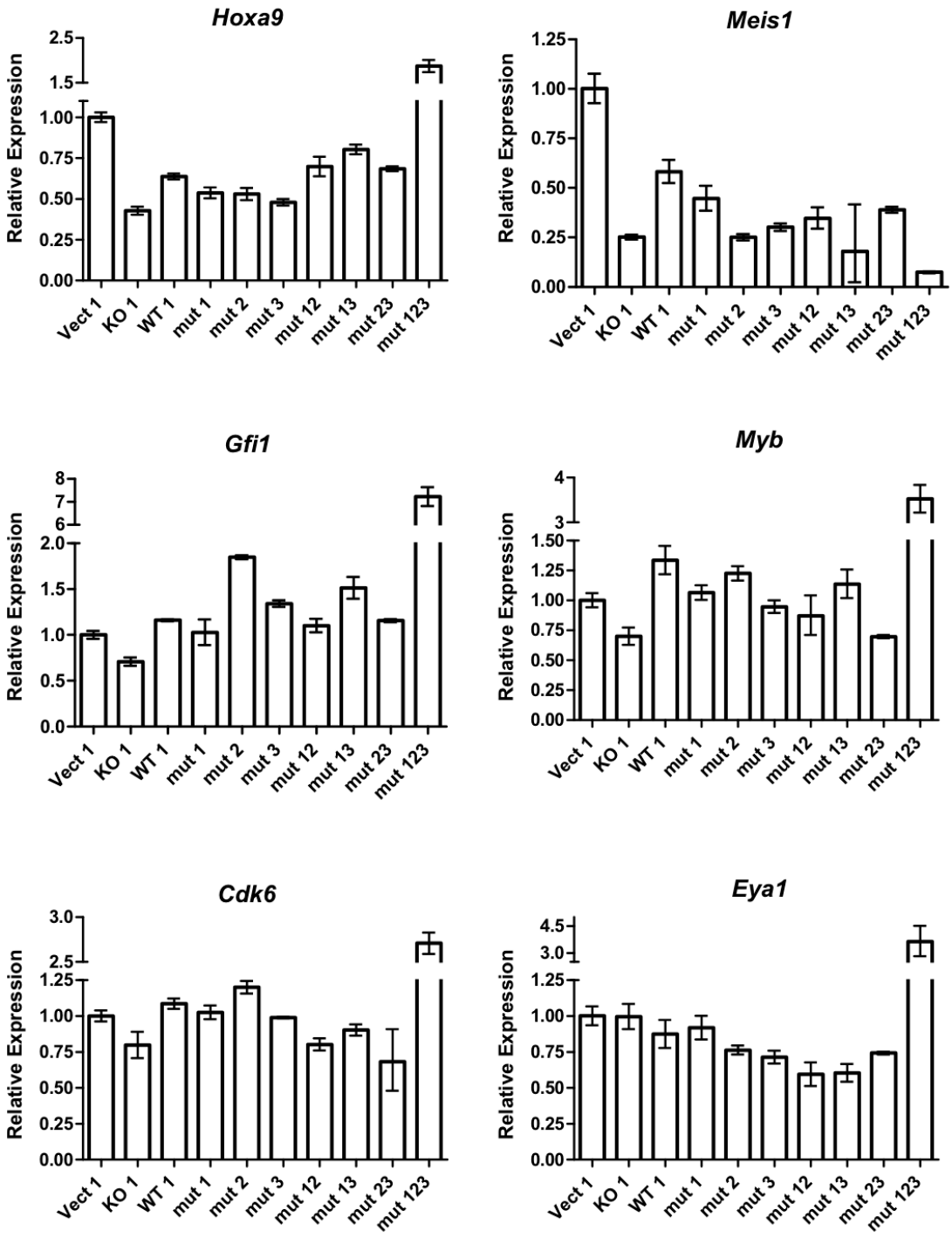


Figure 3-13. Gene expression data using the DOT1L Alanine Mutants

3.6. Discussion

3.6.1. Our AF9 mutations and those previously published do not show specificity for a particular binding partner

We used structure-function studies to show that the level of DOT1L recruitment to the MLL-AF9 protein correlates to the degree of leukemogenic transformation potential. While we were successful in identifying two separate AF9 mutations with differential effects on the DOT1L interactions with AF9 (D544R and D546R), neither is completely specific for the DOT1L-AF9 interaction and both have varying effects on other AF9 binding partners (**Figure 3-4**). As the structures of the different AF9 complexes are very similar, it is challenging to make such specific mutations (**Figure 2-7D**).

Measurements of the binding affinities of these AF9 mutant proteins with binding partners show that these and other previously described AF9 (or ENL) mutations (Biswas et al., 2011; Lokken et al., 2014; Maethner et al., 2013; Tan et al., 2011) are not completely specific for a single protein partner (**Figure 3-14**).

Biswas, et al. created Proline mutants in AF9 and claimed that their AF9 mutants prevented the AF4 protein from interacting with AF9, yet still retains the DOT1L-AF9 interaction. They used these mutations in hematopoietic transformation assays to show that the AF4-AF9 interaction is necessary and called into question the importance of the DOT1L-AF9 interaction in leukemogenesis (Biswas et al., 2011). We could not copurify the AF9 protein with either of these complexes using the double mutants identified in this paper, (L504P, D505P) and (D505P, E506P) (data not shown), both of which are located within an AF9 α1 helix (**Figure 3-14A**). While we do not disagree about

the importance of the AF4-AF9 interaction, it is difficult to make any claims based on their double proline mutations.

It is important to note that co-immunoprecipitation experiments alone do not necessarily provide sufficient information to determine lack of effect of a particular mutation on protein interaction. While it is definitely the case that the full-length proteins may behave differently, our results point out the importance of having the appropriate quantitative binding and structural data to meaningfully interpret the biological effects of these mutations.

3.6.2. Degree of direct DOT1L recruitment to MLL-AF9 affects transformation potential

We clearly observe that the transforming properties of MLL-AF9, as measured by colony formation, are affected by the degree of direct recruitment of DOT1L to MLL-AF9. The D544R and D546R AF9 mutations we have employed have similar effects on the binding of AF4, whereas their effects on DOT1L high affinity site binding differ significantly (D544R partial disruption, D546R complete disruption). As DOT1L and AF4 are the two known AF9 partners involved in positive regulation of transcription (see discussion of CBX8 in **Chapter 4**), this makes it possible to compare the two mutants to assess the effects of a graded reduction in DOT1L binding. The D544R mutation (partial disruption of high affinity DOT1L binding) reduces colony formation substantially but we still observe colony formation. However, the D546R mutation (complete disruption of high affinity DOT1L binding) results in an even greater decrease in colony

formation ability, clearly delineating the importance of direct recruitment of DOT1L through its high affinity binding sites.

As the point mutations in MLL-AF9 can affect interaction with other binding partners as well, we have performed complementary experiments to confirm our MLL-AF9 mutation data. We introduced point mutations into DOT1L that disrupt AF9 binding and probed for the effect on MLL-AF9 transformation properties in the absence of endogenous wildtype DOT1L. Individually blocking each DOT1L binding site shows similar losses in colony formation suggesting that individual DOT1L binding sites can have distinct functional roles. However, we see the most dramatic reduction in serial replating ability upon simultaneously blocking multiple DOT1L interactions with MLL-AF9. Mutation of either all three sites (1+2+3), sites (2+3) or sites (1+3) function similarly to cells without any Dot1l. This suggests an additive function of the binding sites. Indeed, morphological data show that both our MLL-AF9 (D546R) and our Dot1L Site (1+2+3) mutants lead to differentiation of hematopoietic progenitors. Interestingly, mutation of sites (1+2), retaining only the highest affinity interaction site intact, functions similar to wildtype DOT1L. This is an unexpected result, as we obtain different results by mutating sites (1+2) and sites (1+3), both of which have one high affinity binding intact. In this context, it is important to remember that the mutations in DOT1L will affect not only binding to MLL-AF9 but also binding to wildtype AF9 and ENL, amongst other proteins. Thus, the phenotypic output will be the sum of effects on multiple targets, i.e. it is not possible to ascribe the observed effects only to the MLL-AF9 interaction.

3.6.3. Graded disruption of the DOT1L and MLL-AF9 interaction leads to losses of H3K79me2 and H3K79me3 at only a subset of genes

DOT1L is a non-processive, or distributive, enzyme, meaning that at most one round of methylation can take place in each encounter with its substrate before the enzyme must dissociate and re-associate to achieve subsequent rounds of methylation. As a result of this distributive nature of DOT1L, it has been suggested that there could be a functional redundancy between the H3K79 methylation marks (Frederiks et al., 2008). In contrast, several studies have suggested that different methylation states at H3K79 may have different functions in gene regulation (Nguyen and Zhang, 2011). H3K79me2 and H3K79me3 in yeast do not have overlapping chromatin patterns, and unlike H3K79me3 levels, which do not vary over the cell cycle, H3K79me2 levels change (Schulze et al., 2009b). While it is unclear if these different methylation marks play separate roles, it is firmly established that H3K79me2 is linked to transcriptional elongation and that aberrant DOT1L-mediated H3K79 dimethylation is observed in MLL-AF9 leukemias (Bernt et al., 2011).

We show that blocking the high affinity DOT1L interactions with MLL-AF9, via our MLL-AF9 (D546R) mutant, results in significant losses of H3K79me2 and H3K79me3 at only a select number of genes, many of which have already been defined as MLL-AF9 targets. Additionally, we observe significant losses in the methylation pattern across the gene body at these genes, and the losses are more pronounced for the H3K79me3 mark. With the MLL-AF9 (D544R) mutation, we do observe losses in the H3K79me3 mark that are significant across the gene

body despite changes in individual genes not rising to the level of statistical significance, a decrease not seen in the H3K79me2 mark.

We hypothesize that losses of the H3K79me3 mark are directly linked to the presence of multiple sites for binding of AF9 to DOT1L, particularly the high affinity repeat motif in DOT1L. Binding of two (or three) sites on DOT1L to MLL-AF9 would significantly increase the residence time of DOT1L at a specific site on the chromatin. As the enzyme is distributive, a longer residence time increases the probability of proceeding all the way to the H3K79 tri-methylated state (**Figure 3-15**). With a reduction in contacts between MLL-AF9 and DOT1L (D544R mutant), the residence time would be decreased and the level of H3K79 tri-methylation would be reduced. This may also explain the more pronounced effect of the D546R mutation on H3K79me3 versus H3K79me2 across the gene body. Thus, there is a plausible mechanism for the cell to selectively di- or tri-methylate genomic sites by using mono- versus di- (or tri-) valent interaction of DOT1L with AF9 (or ENL) at particular sites in the genome. As DOT1L is the only known H3K79 methyltransferase, H3K79 methylation levels reflect DOT1L occupancy, however our model will likely require further experimental validation using techniques such as cross-linking kinetic analyses (Poorey et al., 2013) to directly assess DOT1L residence time at specific sites in the genome.

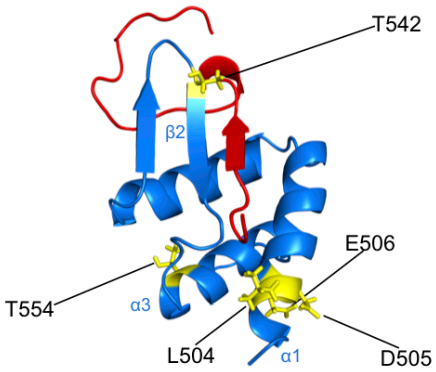
3.6.4. Role of the AF10 and DOT1L interaction in MLL-AF9 leukemogenesis

We observe different patterns of changes in H3K79 methylation marks and in gene expression for various MLL-AF9 target genes, suggesting an underlying complexity in the regulation, which has not been fully elucidated.

Indeed, a recent study by Armstrong and co-workers has shown that the DOT1L mediated conversion of H3K79me1 to H3K79me2 is regulated by AF10 (Deshpande et al., 2014). They show that knockout of AF10 impairs the transformation capability of MLL-AF9, leading to increased myeloid differentiation. Additionally, AF10 knockout leads to a profound reduction in H3K79me3 levels and to lesser but significant decreases in H3K79me2 at specific *Hoxa* genes, with concomitant changes in gene expression. Directly plotting these changes in H3K79me2 and H3K79me3 from Deshpande et al, as a result of AF10 deletion, and highlighting the genes that share losses in H3K79me2 and me3 that are MLL-AF9 targets (**Section 3.4**), show that the *Hoxa* genes have the most significant losses in these methylation marks. Similarly, genes such as *Jmjd1c,* and *Dcun1d1*, amongst others, show large losses in both marks, whereas other genes that we identified such as *Runx2* have only modest changes (**Figures 3-16A and 3-16C**). Upon the examination of those genes in which gene expression analyses were conducted (**Section 3.5**), we observe some expected results such as losses in *Hoxa9* and *Cdk6* di- and tri-methylation. However, we see relatively modest changes in *Meis1* compared to large changes we see in our expression data and surprisingly, increases in H3K79me2 and me3 in *Eya1* upon deletion of AF10 (**Figure 3-16B**). This suggests that some genes, especially the *Hoxa* cluster, are more sensitive towards AF10 mediated deletion and may also suggest differentially mediated AF10 regulation of DOT1L at certain genes.

Interestingly, it has been reported that the octapeptide-leucine zipper motif (OM-LZ) in AF10 (aa. 714-779) directly binds to DOT1L (aa. 615-656) at a site that is in close proximity to the lowest affinity AF9 binding site (Site 1 aa. 628-653, **see Chapter 2**) (Okada et al., 2005). Adding further complexity, this region of DOT1L has been has been reported to bind to the phosphorylated CTD of RNA polymerase II (DOT1L aa. 618-627) (Kim et al., 2012). Therefore one region of DOT1L encompasses binding of AF10, AF9 and the pCTD of RNA polymerase II. It is either possible that all three can bind to DOT1L simultaneously, that these interactions are mutually exclusive, or that they are context dependent in MLL-AF9 leukemogenesis. Furthermore, the AF10 deficient MLL- AF9 leukemia cells showed an increased sensitivity measured by effect on proliferation to the DOT1L inhibitor EPZ004777. This eliminated H3K79me2, residual H3K79me1 marks, and further reduced the expression of *Hoxa7–10* and *Meis1* (Deshpande et al., 2014). This suggests that targeting both the enzymatic activity of DOT1L and either the AF10-DOT1L or AF9-DOT1L interactions may be a successful therapeutic avenue.

A



B

Binding Partner (Kd in nM)	MBP-AF9 (WT)	MBP-AF9 (T542A)	MBP-AF9 (T554A)
AF4	0.17 ± 0.05	7.8 ± 2.3	0.28 ± 0.13
DOT1L Site 2	26.7 ± 6.7	90.2 ± 6.9	40.1 ± 3.4
DOT1L Site 3	1.56 ± .02	29.7 ± 3.3	10.2 ± 0.9
CBX8 (Valine Mutant)	8.7 ± 0.7	57.6 ± 6.5	24.4 ± 2.6

Figure 3-14. Previously published AF9 point mutations are not specific

A) AF9 amino acids that are mutated in other publications are labeled on the DOT1L-AF9 NMR structure.

B) Binding data of AF9 mutants (T542A and T554A) to the AF9 binding partners.

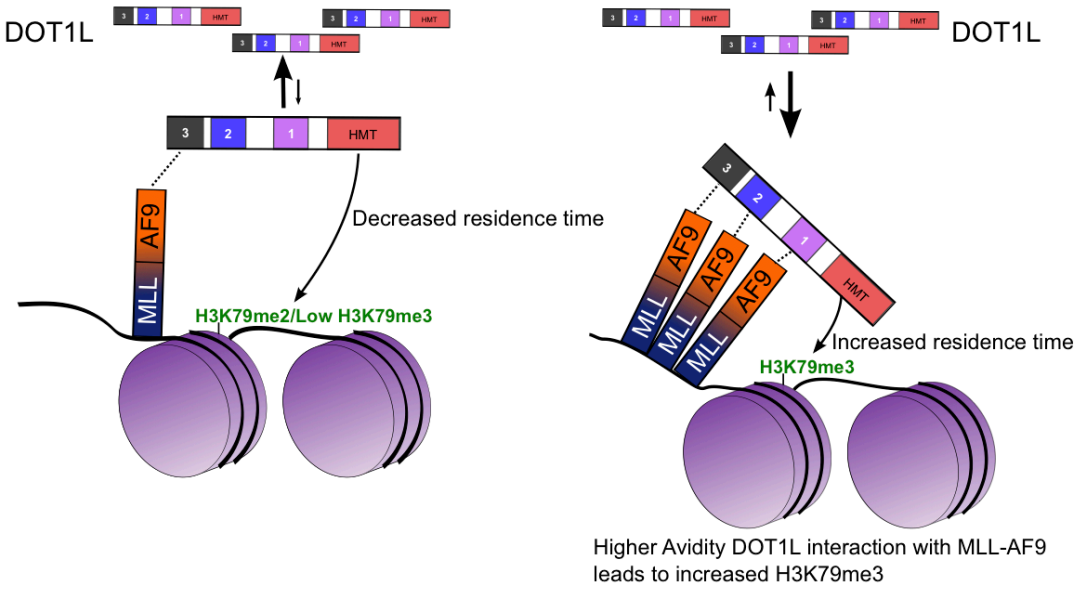
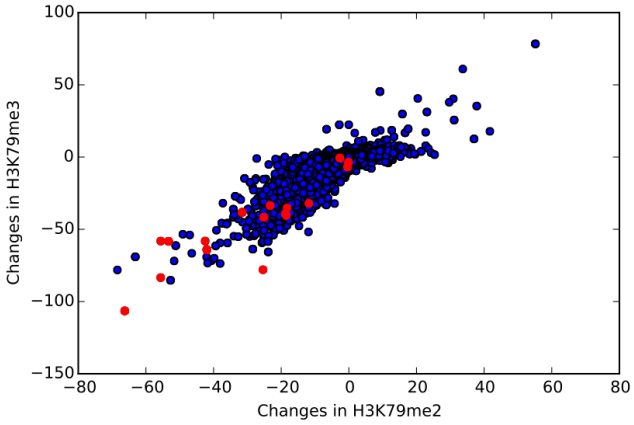


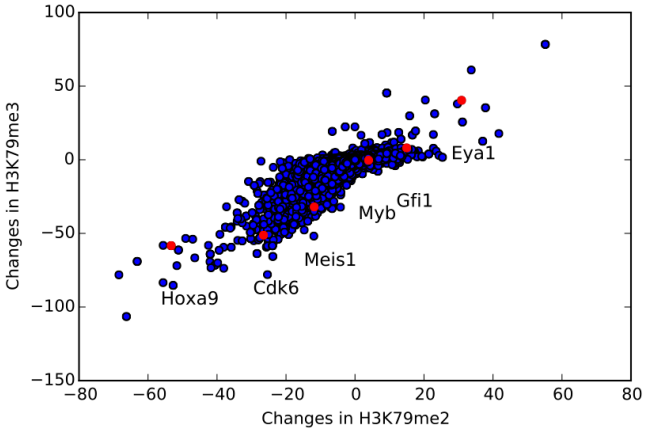
Figure 3-15. Proposed model of protein recruitment to MLL-AF9

One binding event of DOT1L to MLL-AF9 leads to H3K79 di-methylation, but binding of two (or three) sites on DOT1L to MLL-AF9 would significantly increase the residence time of DOT1L at a specific site on the chromatin. As the number of binding events of DOT1L to MLL-AF9 increase from one to two (or three), this increases the residence time of DOT1L at a specific site on the chromatin increasing the probability of proceeding to the H3K79 tri-methylated state.

A



B



C

Description	Differences in H3K79me2	Differences in H3K79me3
Hoxa7	-55.60845248	-58.14870699
Hoxa3	-66.26247651	-106.4800149
Hoxa10	-55.62425915	-83.44102424
Hoxa9	-53.29630493	-58.34059391
Jmjd1c	-25.37303187	-78.0012548
Rdx	-31.54054423	-38.2333751
Erbb2ip	-18.70842535	-39.1441674
Senp6	-18.47563517	-40.39587276
Runx2	-0.379240935	-6.830478895
Zbtb25	-42.47910928	-58.1373536
Arid2	-18.22766572	-35.15900721
Sgk3	-2.64449606	-0.654231615
Tnpo1	-23.26478685	-33.49740168
Dcun1d1	-41.96928677	-64.07008449
Phip	-25.10907297	-41.65135177
Meis1	-11.81823136	-32.05215163
Tshz1	-0.10107279	-3.508198672

Figure 3-16. The dependence of H3K79 di- and tri- methylation on AF10. ChIP-Seq data for H3K79 di- and tri- methylation in MLL-AF9 bone marrow progenitor cells that either contained floxed or deleted AF10. The X-axis shows the differences in H3K79 dimethylation between floxed and deleted AF10. Similarly, the Y-axis consists of changes in H3K79 tri-methylation. Direct correlation between H3K79me2 and me3 were not published in Deshpande et al and these data were kindly provided by Dr. Ani Deshpande (Sanford Burnham) for our data analysis (Deshpande et al., 2014). **A)** All of the genes from Deshpande, et al. data sets are labeled in blue. Indicated in red are all of the MLL-AF9 target genes that show changes in H3K79me2 and me3 patterns as identified in **Figure 3-9**. Raw data indicating changes in methylation for these subset of genes are shown in **C. B)** Genes that we conducted gene expression analysis based on our Chip-Seq data as indicated in **Figure 3-12** are labeled in red.

3.7. Experimental procedures

Sections 3.72, 3.7.3, 3.7.4, 3.78 were provided by Dr. Nancy Zeleznik-Le's laboratory. **Sections 3.7.5, 3.7.6, and 3.7.7** were provided by Dr. Mazhar Adli's laboratory.

3.7.1. Fluorescence Polarization (FP) Binding Measurements

MBP-AF9 (D544R) and MBP-AF9 (D546R) mutants were generated by site-directed mutagenesis, the proteins were purified and binding experiments were conducted in the same manner as described in Chapter 2. To measure binding of the DOT1L (Site 3) alanine mutant, the following peptide was ordered from Biosyn:

DOT1L Site 3 Alanine Mutant: FITC-AHX- NKLP**A**S**A**PLASVVLPSRAERARST

3.7.2. Co-Immunoprecipitation and Western Blots

After 48 hours, cells were lysed and FLAG-tagged proteins were immunoprecipitated with anti-FLAG M2 antibody-conjugated agarose beads (Sigma-Aldrich), or GFP-tagged protein immunoprecipitated with anti-GFP antibody (Life Technologies) or IgG control plus Protein A beads. Western blot of FLAG- or GFP-purified proteins was performed using anti-GFP antibody (Life Technologies) or anti-FLAG antibody (Sigma-Aldrich), respectively. Either 0.5% or 1% of the total cell lysate was used to detect input protein levels. Pixel analysis was performed using MultiGauge v3.0 (FUJIFILM) software. Relative precipitation was calculated by normalizing the immunoprecipitated protein to the FLAG input, and set relative to wild type AF9. Error bars indicate the standard

error of the mean. Statistical significance was assessed using Student's t-test with * = 0.05, ** = $p<0.01$.

3.7.3. MLL-AF9 wildtype and MLL-AF9 mutant serial replating assays

MLL-AF9 wildtype and MLL-AF9 mutant serial replating assays were conducted with murine bone marrow c-kit+ cells were transduced with MSCVneo, MSCVneo-MLL-AF9 (WT), MSCVneo-MLL-AF9 (D544R) or MSCVneo-MLL-AF9 (D546R) retroviruses. Cells were plated in methylcellulose medium with cytokines and G418 as we have previously described (Cierpicki et al., 2010). Colonies were enumerated and cells were serially replated after 7 days for each of four weeks. Colony assays were conducted in duplicate and repeated n=10 or n=8 for MLL-AF9(D544R) and MLL-AF9(D546R), respectively.

3.7.4. DOT1L mutant serial replating assays

For DOT1L mutant serial replating assays, full length DOT1L, either wild type or containing alanine mutations, were cloned into an MSCV-mCherry retroviral vector (kindly provided by Dr. Jiwang Zhang, Loyola University Chicago) and verified by sequencing. MLL-AF9-transformed bone marrow progenitor cells from *Dot1*$^{fl/\Delta}$ mice, which we previously described (Chang et al., 2010), were co-infected with retroviruses expressing mCherry-DOT1L (wild type or mutant) plus either MIGR1 or MIGR1-Cre-expressing retroviruses. After 20-24 hours, mCherry/GFP double-positive cells were sorted by FACS and cultured in methylcellulose as described above. After one week, colonies were counted and cells processed for cytospin and for genomic DNA. PCR was used to determine status of the endogenous *Dot1l* floxed allele using three primers that can

distinguish the deleted from the floxed allele, as we have done previously (Chang et al., 2010). Primer sequences used: flox forward: 5'- GCT CTG CTT TAG GGG CAT CCT G – 3'; cre forward: 5' – CGA GCC CTT CTC CCC TGA GGT GTA TG – 3'; Reverse: 5' – GTG TGC ATG TGT CAG TGT CTG AAC CAA GTG – 3'. All primers were from IDT.

3.7.5. ChIP-Seq Experiments

Briefly, approximately ~500,000 MLL-AF9 wildtype and mutant progenitor cells isolated after one week culture in methylcellulose colony assay were crosslinked with 1% formaldehyde for 10 min and neutralized with final 0.125 M glycine for 5 min at 37°C. After lysing with SDS lysis buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl, pH 8.1) for 20 min on ice, the chromatin was sonicated using a Branson digital sonifier (40% amplitude with 0.7 s "on" and 1.3 s "off" pulse cycles). After sonication, chromatin was diluted 10X with ChiP-dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA and 16.7 mM Tris-HCl, pH 8.1) and incubated with ~1.5 &g antibodies overnight at 4°C. Then, mixture of protein A-G magnetic beads (Dynabeads, Life Technologies) were used to pull down antibody- chromatin complexes which were then washed serially with low-salt immune complex wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl (pH 8.1) and 150 mM NaCl); LiCl wash buffer (0.25 M LiCl, 1% NP40, 1% deoxycholate, 1 mM EDTA and 10 mM Tris-HCl (pH 8.1), and finally with TE buffer (10 mM Tris-HCl and 1 mM EDTA (pH 8.0). The chromatin was recovered from the beads using elution buffer (0.2% SDS and 0.1 M $NaHCO_3$ supplemented with fresh 5 mM DTT) at 65°C for 10 min. ChIP-DNA was

extracted with ethanol precipitation after reverse crosslinking, proteinase K and RNAse digestion. DNA was quantified via Qubit Fluorometer.

3.7.6. ChIP DNA sequencing and data processing

For DNA sequencing, approximately 0.5-3 ng ChIP-DNA was processed for whole genome library preparation using Illumina TruSeq Chip Library Kit according to the manufacturer's protocol. The sequencing reads from MiSeq were aligned to mouse mm9 genome using BOWTIE alignment tool (Langmead et al., 2009). These aligned reads were then processed and converted into BigWig (http://genome.ucsc.edu/goldenPath/help/bigWig.html) format, which were loaded in the UCSC Genome browser (http://genome.ucsc.edu) for visualization. BEDTOOLS (Quinlan and Hall, 2010) suite commands were used for format conversions. Finally, the reads were converted to BED format (http://genome.ucsc.edu/FAQ/FAQformat#format1) for further data processing such as peak calling.

3.7.7. Identifying differentially enriched genes.

To identify genes enriched for *H3K79me2* and *H3K79me3*, we calculated the total number of reads in the gene body regions for each gene. To calculate the enrichments, Transcription Start Site (TSS) and Transcription End Site (TES) coordinates were downloaded from UCSC Genome browser for the mouse (mm9) gene list. BED file of regions of interest were created and defined as coordinates from TSS(-2000bp) to TES, for each gene. Peak finding was performed using MACS14 (Zhang et al., 2008) with ChIP-seq BED files of the samples as input and default parameters (pvalue=1e-6). Once peaks were

obtained for all the tracks, we selected regions of interest for the genes that overlapped with these peaks, with the BEDTOOLS *intersect* command. This was done in order to filter out genes whose regions of interest did not show significant enrichment in the ChIP-seq data. Differentially enriched genes were identified using the DEseq package from Bioconductor (Anders and Huber, 2010). Genomic profiles were generated using Ngsplot (Shen et al., 2014).

3.7.8. Quantitative RT-PCR

Cells were harvested from methylcellulose colony assays after one week, RNA was isolated and cDNA was synthesized. Quantitative Real-Time PCR was performed using TaqMan probes for *Hoxa9*, *Meis1*, *Gfi1, Myb, Cdk6 and Eya1* (Applied Biosystems), and data were analyzed using the $2^{-\Delta\Delta}$ Ct method. Expression was normalized to *Gapdh* expression and was performed in triplicate.

3.8. References for Chapter 3

Bernt, K.M., Zhu, N., Sinha, A.U., Vempati, S., Faber, J., Krivtsov, A. V, Feng, Z., Punt, N., Daigle, A., Bullinger, L., et al. (2011). MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. Cancer Cell *20*, 66–78.

Biswas, D., Milne, T. a, Basrur, V., Kim, J., Elenitoba-Johnson, K.S.J., Allis, C.D., and Roeder, R.G. (2011). Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. Proc. Natl. Acad. Sci. U. S. A. *108*, 15751–15756.

Chang, M.-J., Wu, H., Achille, N.J., Reisenauer, M.R., Chou, C.-W., Zeleznik-Le, N.J., Hemenway, C.S., and Zhang, W. (2010). Histone H3 lysine 79 methyltransferase Dot1 is required for immortalization by MLL oncogenes. Cancer Res. *70*, 10234–10242.

Deshpande, A.J., Deshpande, A., Sinha, A.U., Chen, L., Chang, J., Cihan, A., Fazio, M., Chen, C., Zhu, N., Koche, R., et al. (2014). AF10 Regulates Progressive H3K79 Methylation and HOX Gene Expression in Diverse AML Subtypes. Cancer Cell.

Frederiks, F., Tzouros, M., Oudgenoeg, G., van Welsem, T., Fornerod, M., Krijgsveld, J., and van Leeuwen, F. (2008). Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states. Nat. Struct. Mol. Biol. *15*, 550–557.

Khandanpour, C., Phelan, J.D., Vassen, L., Schütte, J., Chen, R., Horman, S.R., Gaudreau, M.-C., Krongold, J., Zhu, J., Paul, W.E., et al. (2013). Growth factor independence 1 antagonizes a p53-induced DNA damage response pathway in lymphoblastic leukemia. Cancer Cell *23*, 200–214.

Kim, S.-K., Jung, I., Lee, H., Kang, K., Kim, M., Jeong, K., Kwon, C.S., Han, Y.-M., Kim, Y.S., Kim, D., et al. (2012). Human histone H3K79 methyltransferase DOT1L protein [corrected] binds actively transcribing RNA polymerase II to regulate gene expression. J. Biol. Chem. *287*, 39698–39709.

Leach, B.I., Kuntimaddi, A., Schmidt, C.R., Cierpicki, T., Johnson, S. a, and Bushweller, J.H. (2013). Leukemia fusion target AF9 is an intrinsically disordered transcriptional regulator that recruits multiple partners via coupled folding and binding. Structure *21*, 176–183.

Lokken, A.A., Achille, N.J., Chang, M., Lin, J.J., Kuntimaddi, A., Leach, B.I., Malik, B., Nesbit, J.B., Zhang, S., Bushweller, J.H., et al. (2014). Importance of a Specific Amino Acid Pairing for Murine MLL Leukemias Driven by MLLT1/3 or AFF1/4. Leuk. Res.

Maethner, E., Garcia-Cuellar, M.-P., Breitinger, C., Takacova, S., Divoky, V., Hess, J.L., and Slany, R.K. (2013). MLL-ENL inhibits polycomb repressive complex 1 to achieve efficient transformation of hematopoietic cells. Cell Rep. *3*, 1553–1566.

Nguyen, A.T., and Zhang, Y. (2011). The diverse functions of Dot1 and H3K79 methylation. Genes Dev. *25*, 1345–1358.

Nguyen, A.T., Taranova, O., He, J., and Zhang, Y. (2011). DOT1L, the H3K79 methyltransferase, is required for MLL-AF9-mediated leukemogenesis. Blood *117*, 6912–6922.

Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V.M., Su, L., Xu, G., and Zhang, Y. (2005). hDOT1L links histone methylation to leukemogenesis. Cell *121*, 167–178.

Placke, T., Faber, K., Nonami, A., Putwain, S.L., Salih, H.R., Heidel, F.H., Krämer, A., Root, D.E., Barbie, D.A., Krivtsov, A. V, et al. (2014). Requirement for CDK6 in MLL-rearranged acute myeloid leukemia. Blood *124*, 13–23.

Poorey, K., Viswanathan, R., Carver, M.N., Karpova, T.S., Cirimotich, S.M., McNally, J.G., Bekiranov, S., and Auble, D.T. (2013). Measuring chromatin interaction dynamics on the second time scale at single-copy genes. Science *342*, 369–372.

Schulze, J.M., Jackson, J., Nakanishi, S., Gardner, J.M., Hentrich, T., Haug, J., Johnston, M., Jaspersen, S.L., Kobor, M.S., and Shilatifard, A. (2009). Linking cell cycle to histone modifications: SBF and H2B monoubiquitination machinery and cell-cycle regulation of H3K79 dimethylation. Mol. Cell *35*, 626–641.

Shen, C., Jo, S.Y., Liao, C., Hess, J.L., and Nikolovska-Coleska, Z. (2013). Targeting recruitment of disruptor of telomeric silencing 1-like (DOT1L): characterizing the interactions between DOT1L and mixed lineage leukemia (MLL) fusion proteins. J. Biol. Chem. *288*, 30585–30596.

Tan, J., Jones, M., Koseki, H., Nakayama, M., Muntean, A.G., Maillard, I., and Hess, J.L. (2011). CBX8, a Polycomb Group Protein, Is Essential for MLL-AF9-Induced Leukemogenesis. Cancer Cell *20*, 563–575.

Wang, Q.-F., Wu, G., Mi, S., He, F., Wu, J., Dong, J., Luo, R.T., Mattison, R., Kaberlein, J.J., Prabhakar, S., et al. (2011). MLL fusion proteins preferentially regulate a subset of wild-type MLL target genes in the leukemic genome. Blood *117*, 6895–6905.

Yokoyama, A., Lin, M., Naresh, A., Kitabayashi, I., and Cleary, M.L. (2010). A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. Cancer Cell *17*, 198–212.

Chapter 4. Biochemical and Structural characterization of the CBX8-AF9 interaction

4.1. Introduction

Polycomb group (PcG) proteins were identified as playing an important role in development in *Drosophila melanogaster* as PcG mutations result in body segment transformations (Lewis, 1978). Mice with Polycomb group gene knockout are embryonic lethal. These PcG proteins act in concert with the Trithorax group (TrxG) proteins to maintain either repressive or active transcriptional states of important genes. As both the PcG and TrxG proteins are involved in chromatin modifications, the ensuing marks that these proteins deposit are heritable and a part of cellular or epigenetic memory (Reviewed in Ringrose and Paro, 2004 and in Simon and Kingston, 2009). The PcG proteins form two separate complexes, PRC1 and PRC2, and are thought to function in a sequential manner to repress transcription through chromatin modifications. PRC2 is composed of three components, Ezh1/Ezh2, Eed, Suz12. The SET domains of Ezh1/Ezh2 contribute to gene repression through the tri-methylation of Histone H3 Lysine 27 (H3K27me3), or Histone H3 Lysine 9 (H3K9me3) (Cao et al., 2002; Kuzmichev, 2002). The PRC1 complex contains Ring1 proteins that monoubiqutinate histone H2A Lysine 119 (H2AK119ub) which leads to chromatin compaction and the pausing of RNA polymerase II (Francis et al., 2004; de Napoles et al., 2004; Simon and Kingston, 2009). The composition of the PRC1 complex is more variable but contains Phc, ubiquitin ligases Ring1a and 1b, Bmi1 and mutually exclusive CBX proteins (Gao et al., 2012).

Due to the importance of PcG proteins in cellular regulation, both PRC1 and PRC2 components have been found to play roles in oncogenesis. Increased expression of EZH2 has been observed in a wide range of cancers. Inhibitors of EZH2 have shown efficacy in treating non-Hodgkin lymphoma mouse models (Knutson et al., 2014). Additionally, global reduction of PRC2 function by conditional knockout of Ezh2 impaired MLL-AF9-induced leukemogenesis in vivo (Neff et al., 2012; Tanaka et al., 2012). BMI1 and RING1B have elevated expression levels in gastrointestinal tumors and lymphomas (Sánchez-Beato et al., 2006).

As the PRCs are involved in the maintenance of epigenetic memory, their regulation should be fine-tuned and specific. However, the mechanisms underlying the PRC recruitment to target genes remain unclear. One model suggests that PRC2 is recruited to chromatin through long non-coding RNAs such as Xist (Zhao et al., 2008) and HOTAIR (Rinn et al., 2007), or transcription factors such as Snail (Herranz et al., 2008). A second model suggests that PRC2 recruitment is based on the marks already present on chromatin. Methylation of histone marks correlated with active transcription has been shown to prevent the deposition of H3K27me3 by PRC2 (Schmitges et al., 2011). A recent paper has shown that there is crosstalk between Jarid2 and the PRC2 complex, and that methylation of Jarid2 by PRC2 is necessary for the deposition of H3K27me3 (Sanulli et al., 2015). It is also widely unknown how the transition from a repressed state modulated by PRCs proceeds to an active state.

Five different Chromobox proteins (CBX2, CBX4, CBX6, CBX7 and CBX8) have been found in mammals. CBX proteins are made up of a conserved N-terminal chromodomain, which binds to methylated histones: H3K27me3 or H3K9me2/me3 with unequal affinities (Kaustov et al., 2011). Both of these histone marks correlate with silent genomic regions. Therefore, the prevailing model is that PRC2 complex trimethylates H3K27 and this mark is read by CBX proteins in the PRC1 complex, which leads to H2AK119 ubiquitination (Simon and Kingston, 2009). CBX proteins differ in a number of their domains suggesting that they could provide for different PRC1 functionalities. CBX proteins show distinct expression patterns during hematopoietic stages suggesting the formation of different PRC1 complexes during hematopoietic differentiation (Senthilkumar and Mishra, 2009). Indeed, the composition of the PRC1 complex changes during differentiation of ES and hematopoietic stem cells (Klauke et al., 2013). A recent paper suggests that PRC1 complexes that contain CBX8 replace those that contain CBX7 in order to repress developmental genes allowing for differentiation (Creppe et al., 2014). Some CBX proteins can associate with protein complexes other than PRC1 and may play a non-PRC1 dependent role in transcriptional regulation (Kerppola, 2009). It is unclear whether CBX8 is able to interact with protein complexes outside of PRC1.

CBX8 (Polycomb 3/Pc3/hPC3/CBX8) has been shown to be present in complexes that are involved in MLL fusions. As CBX8 has been well characterized to play a role in transcriptional repression, it is not clear why CBX8 would be associated with MLL-AF9 fusions. This raises the question as to the

role that the direct recruitment of CBX8 to MLL-AF9 plays in MLL-AF9 leukemogenesis.

4.2. Characterization of the CBX8-AF9 interaction[9]

Previous studies have shown that CBX8 (aa. 250-349) directly interacts with both AF9 and ENL (Hemenway et al., 2001; Maethner et al., 2013). This stretch of residues within CBX8 is not conserved between other CBX proteins suggesting that the CBX8 interaction with AF9 is specific. We co-expressed these CBX8 residues (aa. 250-349) with the C-terminal domain of AF9 (aa. 499-568) and were able to detect a robust protein-protein interaction via size exclusion chromatography. However, this initial co-expression led to a low quality NMR spectrum not amenable for structural studies. We predicted that CBX8 binds to the same regions of the C-terminal domain of AF9 (aa. 499-568) as AF4 and DOT1L. Thus, we assumed if we narrowed down the CBX8 construct we would be able to acquire a high-resolution 2D NMR spectrum. We used secondary structure predictions such as JPRED and disorder predictors such as PONDR to potentially identify regions of secondary structure and detect flexible regions. Similar to our studies with DOT1L-AF9, we created a series of deletion constructs and analyzed them using heteronuclear triple resonance NMR to assign resonances and $\{^{15}N\}$-^{1}H heteronuclear NOE measurements to assess the dynamic behavior of the residues. These experiments enabled us to identify which CBX8 residues were flexible, not involved in the protein-protein interaction,

[9] Initial experiments to identify the minimal interacting regions between CBX8 and AF9 were conducted by Ben Leach.

and remove them. This was carried out in an iterative fashion and we successfully identified the minimal interacting regions of this protein-protein interaction: CBX8 (aa. 326-349) and AF9 (aa. 499-568) (**Figure 4-1A**). We were able to make backbone assignments of all AF9 and CBX8 residues in this complex, with the exception of Leucine 343 N-H. However, we were able to observe all other L343 resonances in our triple resonance experiments. Similar to AF4 and DOT1L, the CBX8 portion of this interaction is predicted to be disordered (**Figure 4-1B**), meaning that this region of CBX8 undergoes coupled folding and binding to form a structured complex with AF9. Unlike DOT1L, the CBX8 peptide motif, LIARIPVARI, or a similar motif was not found at any other place in the full length CBX8 protein, meaning that this this stretch of amino acids (aa. 326-349) constitutes the only interaction between these two proteins.


A
10
9
8
7
6
105
110
115
120
125
130
135
341ArgH,N
536PheH,N
550LeuH,N
522IleH,N
519GluH,N
517LeuH,N
336ArgH,N
B
PONDR Score
1.0
0.8
0.6
0.5
0.4
0.2
0.0
0
50
100
150
200
250
300
350
Residue Number
Disorder
Order
VL-XT

Figure 4-1. Identification and characterization of the CBX8-AF9 interaction.

A) Assigned ^{15}N-^{1}H HSQC NMR spectra of CBX8 in complex with AF9. Both proteins are labeled and assigned.

B) PONDR plot of the full length CBX8 protein. The CBX8 (aa. 326-349) interaction site with AF9 is highlighted in blue and is predicted to be disordered.

4.3. Solution NMR Structure of the CBX8-AF9 complex

We solved the high resolution NMR structure of CBX8 in complex with AF9 (in preparation for PDB deposition). We initially had problems assigning our ^{13}C Aliphatic NOEs using CYANA generated automatic assignments due to limited dispersion and severe peak overlap. Similar to our NMR experiments with DOT1L-AF9, we initially ran our ^{13}C aliphatic NOE assignments on the 600MHz magnet with a cryoprobe. We switched over to manual assignments and reran all our NOE experiments using the 800MHz magnet with a cryoprobe, and additionally sent out a ^{13}C ^{15}N double labeled NMR sample to run experiments on the 900MHz at The National Magnetic Resonance Facility at Madisoin (NMRFAM). Due to malfunctioning of the magnet at NMRFAM, we were unable to make full use of the NOE experiments or use any of the NMR experiments for structure calculation purposes, but we were able to use ^{13}C HSQC and HCCH-TOCSY experiments run on the 900MHz instrument to help guide us in our NOE assignments. To this end, we used a combination of dihedral angles, NOEs and chemical shifts to calculate and refine the structure without any significant constraint violations (**Table 4-1**). The NMR ensemble of the 10 lowest energy structures of CBX8-AF9 shows a well-formed complex, backbone RMSD = 0.37Å (**Figure 4-2A**). The CBX8-AF9 complex forms a mixed alpha-beta structure. AF9 forms three helices around the CBX8 peptide (**Figure 4-2B**) and a β hairpin, which forms a three-stranded antiparallel β sheet with the β strand from CBX8. CBX8 residues (aa. 333-339) form a β strand followed by a β turn (aa. 340-344).

The following CBX8 C-terminal amino acids, 345-349, are unstructured (**Figure 4-5 HetNOE**).

Comparison of AF9 in complexes with its binding partners shows the formation of similar mixed alpha-beta structures (**Figure 4-3A**). Superposition of the backbone residues of CBX8-AF9 onto DOT1L-AF9 yields an RMSD of 1.3Å (**Figure 4-3B**) and an RMSD of 1.6Å with the AF4-AF9 structure (**Figure 4-3C**). The main differences between the AF9 complex structures lie in the dynamic β hairpin on AF9 (aa. 539-542) and in the β-turn followed by the extended strand of each of the binding partner proteins.

Table 4.1 Statistics for NMR Data Collection and Structure Calculations of the CBX8-AF9 complex	Value
NMR Distance and Dihedral Constraints	
Distance constraints	
Total NOE	1882
Intraresidue	960
Interresidue	922
Sequential ([|i-j| = 1)	401
Medium range (|i-j| <= 4)	242
Long range (|i-j| > 5)	279
Intermolecular	132
Total dihedral angle restraints	143
ϕ	71
ψ	72
Structure Statistics	
Violations (mean and SD)	
Distance constraints (Å)	.016 +/- .001
Dihedral angle constraints (°)	.248 +/- 0.098
Maximum dihedral angle violation (°)	2.837
Maximum distance constraint violation (Å)	0.303
Deviations from idealized geometry	
Bond lengths (Å)	.001 +/- 0.000
Bond angles (°)	.366 +/- 0.004
Impropers (°)	.219 +/- 0.005
Average pairwise rmsd (Å)	
Heavy	0.89
Backbone	0.37

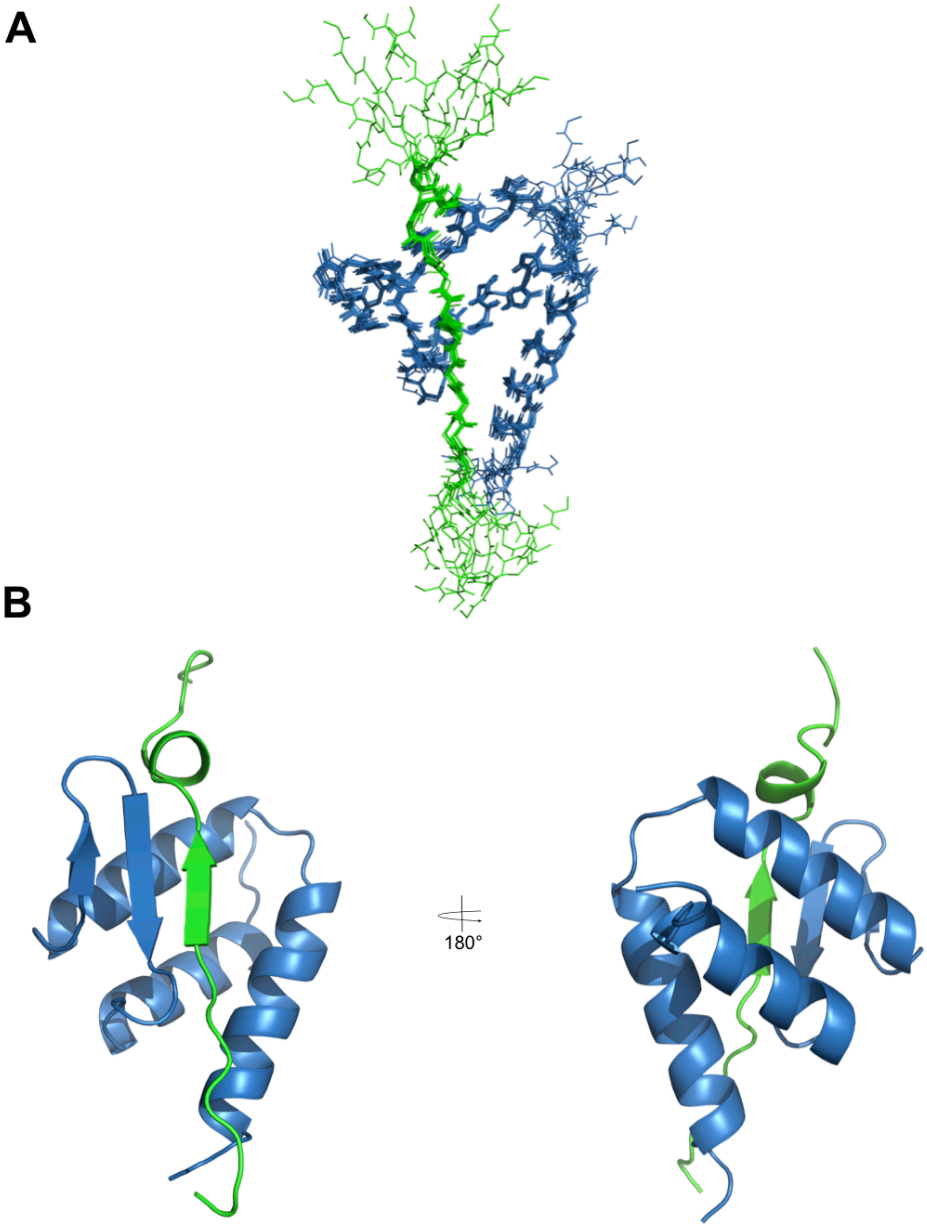


Figure 4-2. Structure of the CBX8-AF9 complex.

A) Ensemble of the 10 lowest energy conformers. CBX8 is shown in green and AF9 in blue.

B) Cartoon representation of the lowest energy conformer. CBX8 (green) forms a β strand followed by a β turn along with the β hairpin from AF9 (blue). AF9 additionally forms three α helices.

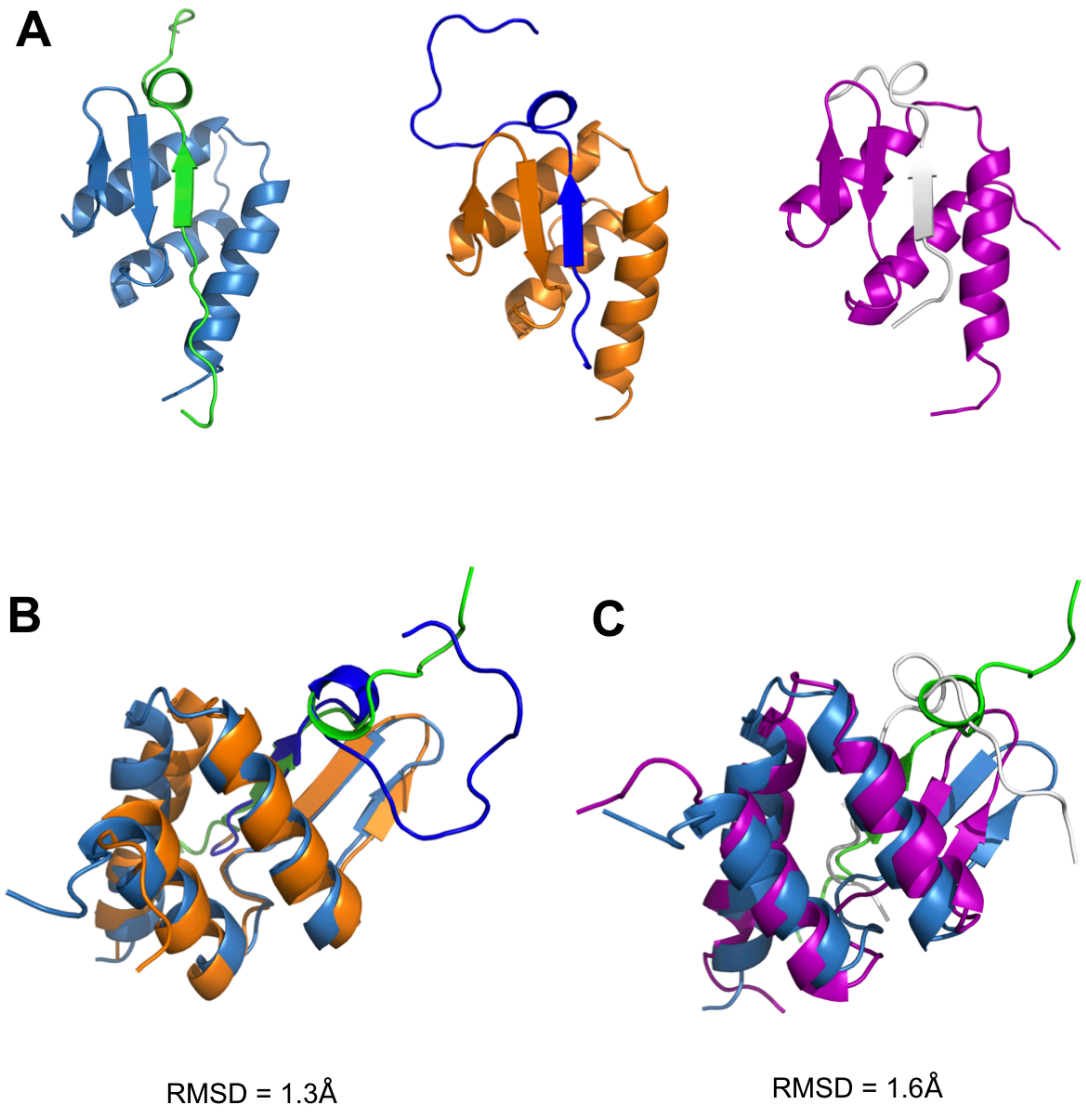


Figure 4-3. The CBX8-AF9 complex is structurally similar to the other AF9 complexes.

A) Cartoon representation of the AF9 complexes. From left to right: CBX8-AF9 (blue and green), DOT1L-AF9 (orange and blue) and AF4-AF9 (purple and white).

B) Superposition of CBX8-AF9 onto DOT1L-AF9 yields an RMSD of 1.3 Å.

C) Superposition of CBX8-AF9 onto AF4-AF9 yields an RMSD of 1.6Å.

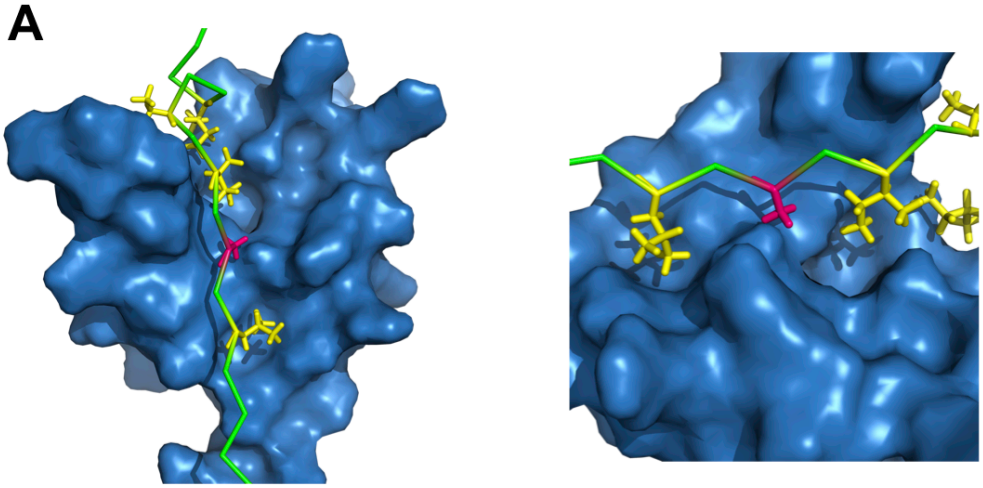
4.4. Binding properties and Dynamics of the CBX8-AF9 complex

Similar to the other AF9 complexes, the interface between the CBX8 and AF9 proteins is hydrophobic, as CBX8 L333, I337, and V339 are critical hydrophobic residues that are buried within the CBX8-AF9 interface. Intriguingly, the CBX8 interaction site with AF9 is a little different from the hydrophobic motifs that we observe with AF4 and DOT1L by its inclusion of an Alanine within the third position of the peptide motif (**Figure 4-4A and B**). This is opposite of what we observe with DOT1L Site1 where the inclusion of a bulky Leucine within the binding motif presumably causes weaker binding to AF9 due to the spatial constraints within the DOT1L-AF9 binding pocket. Here we presumed that the lack of interactions between the CBX8 Alanine (aa. 335) and AF9 residues does not allow for the CBX8 β-strand to become buried into the AF9 interface. Indeed, the binding of CBX8 to AF9 is weaker than DOT1L or AF4 (Kd > 900nM). We presumed that artificial replacement of this Alanine with a Valine would increase binding. Indeed, this is the case as CBX8 (Valine mutant) binds with high affinity to MBP-AF9 (K_d = 11nM). This CBX8 (Valine mutant) provides us with an artificial means of quantifying the binding of AF9 mutations with CBX8 (**Figure 3-4A**).

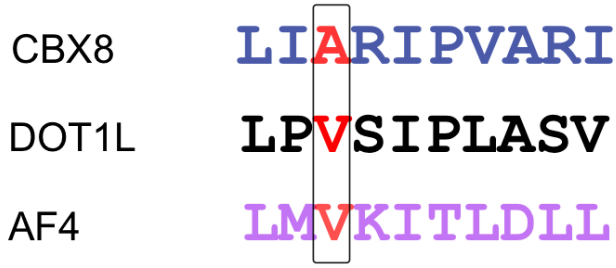
The CBX8-AF9 complex exhibits a wide range of backbone dynamics. Similar to our DOT1L-AF9 complexes the majority of the AF9 residues were ordered but we observe generally elevated $R_1 * R_2$ values near the loop regions on AF9 (aa. 516-521, 532-534, 540-543, 549-554) (**Figure 4-5**). Additionally, we observe that the AF9 β strand adjacent to the CBX8 peptide exhibits both fast

and slow timescale motions, different from what we observed in the DOT1L-AF9 complexes and similar to what we observed in an AF4-AF9 complex (Leach et al., 2013).

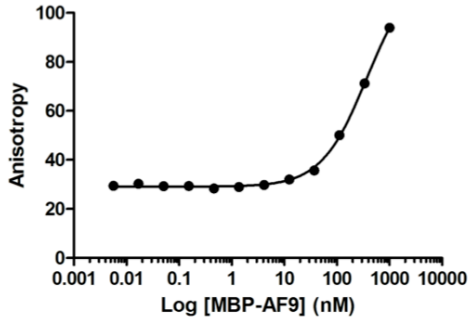
A



B


CBX8
LIARIPVARI
DOT1L
LPVSIPLASV
AF4
LMVKITLDLL

C


Anisotropy
100
80
60
40
20
0
0.001 0.01 0.1 1 10 100 1000 10000
Log [MBP-AF9] (nM)

D

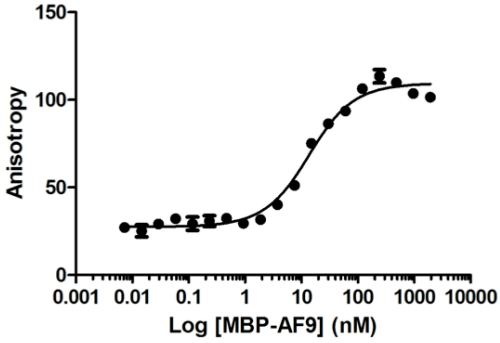

Anisotropy
150
100
50
0
0.001 0.01 0.1 1 10 100 1000 10000
Log [MBP-AF9] (nM)

Figure 4-4. Valine mutations within CBX8 "restores" binding affinity with AF9.

A) Surface representation of the CBX8-AF9 complex. AF9 is labeled in blue and labeled in yellow are hydrophobic side chains of CBX8 that are buried within the protein complex. Labeled in pink is the alanine of CBX8.

B) Alignment of the binding motifs of CBX8, DOT1L and AF4. Labeled in red is third position where CBX8 has an Alanine instead of a Valine.

C) Wildtype CBX8 binds weakly to MBP-AF9

D) CBX8 Valine mutant binds to MBP-AF9 tightly, K_d = 11nM.

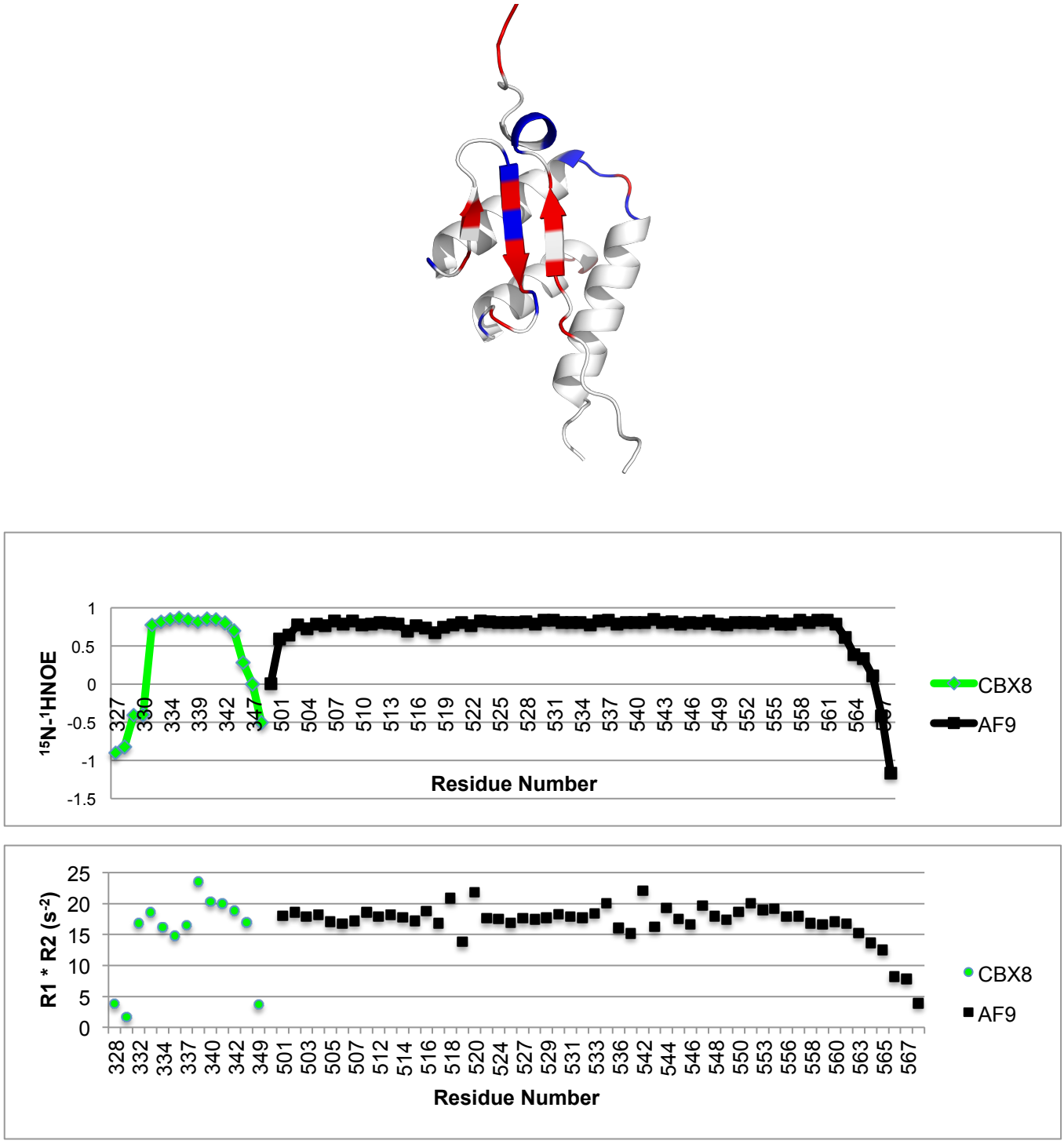


Figure 4-5. Dynamics of the CBX8-AF9 Complex.

CBX8-AF9 structure where Red indicates depressed R1 * R2 and Blue indicates elevated R1 * R2 motions. Graphs of R1 * R2 and ^{15}N-^{1}H NOE of the complex are shown below.

4.5. Discussion

It has been reported that direct recruitment of CBX8 is essential for MLL-AF9 and MLL-ENL leukemogenesis (Maethner et al., 2013; Tan et al., 2011). Tan, et al. and Maethner, el al. used two separate mutations in AF9 and ENL (T542A and T554A) to argue that the CBX8-AF9 and the CBX8-ENL interactions are critical for leukemogenesis. Using co-immunoprecipitation data, they both claim that their mutations are specific for the AF9 (or ENL) interactions. Our structural data show that the T554A mutation on AF9 is not located near the peptide binding site, is on AF9 α3 helix, and has minimal effect on the binding to the AF9 interacting partners (**Figure 3-14**). On the other hand, the T542A AF9 mutation is located on the β2 strand and makes contacts with the peptide (**Figure 3-14A**). As expected, T542A affects binding of all of the AF9 partners (**Figure 3-14B**). Thus, the AF9 point mutations used in these studies (T542A and T554A), that they claim are specific for blocking CBX8 binding to AF9, not only are not specific, but have additional impact on DOT1L and AF4 binding (**Figure 3-14B**). Thus, further detailed studies are needed to determine the mechanism of CBX8-dependence in MLL-AF9 and MLL-ENL leukemogenesis.

Additionally, several studies have conducted in-depth biochemical experiments to show that AF9 and ENL bind mutually exclusively to their partners (Biswas et al., 2011; Maethner et al., 2013). Here, we provide a structural basis to their claims as CBX8, AF4, and the multiple binding sites within DOT1L form nearly identical structures with the C-terminal domain of AF9. CBX8 is unique amongst all of the AF9 binding partners due to its weak binding affinity. We

showed that this was due to the alanine residue in the third position of the AF9 binding motif. Substitution of this alanine with a valine leads to high binding affinity comparable to that of the other AF9 binding partners. It is possible that this particular alanine residue is the reason for the increased dynamics of the CBX8-AF9 complex compared to the backbone dynamics of the other AF9 binding partners. Thus, future R_1 * R_2 experiments may reveal that backbone dynamics are quenched as a result of the CBX8 valine substitution.

We have not yet conducted any biological studies to confirm the role of the direct recruitment of CBX8 to MLL-AF9. Based on our NMR solution structures, it is unlikely that we will be able to find a point mutation on AF9 that is specific for this CBX8-AF9 interaction. Also, we cannot take advantage of a multivalent interaction for the CBX8-AF9 interaction as we did with DOT1L. The introduction of the D546R mutation into MLL-AF9 results in an increase in CBX8 binding, as measured by co-immunoprecipitation (**Figure 3-4**). Structurally, we cannot provide a reason for the significant increase in binding of the full length CBX8 protein to AF9. Apparently, loss of binding of the co-activators, particularly DOT1L, results in increased repressor (CBX8) binding. While the exact mechanism is not yet clear, we would speculate that CBX8 binding to MLL-AF9 (or AF9) leading to stable transcriptional silencing is the default state in the absence of activators; the binding of DOT1L and AF4, in some sequence, relieves this default repressive state leading to activation of gene expression (**Figure 4-6**). To elucidate the role of this interaction we can make use the AF9 (D546R) mutation that increases CBX8 binding to AF9 and probe for the role of

CBX8 and H3K27me3 through ChIP-Seq and ChIP-qPCR experiments. Unlike with the DOT1L protein, there is a ChIP-validated antibody towards the native CBX8 protein that has been used in published genomic studies (Klauke et al., 2013). Thus, we can conduct genome wide experiments with H3K27me3 and CBX8 to understand the role of increased binding to AF9 when DOT1L and AF4 do not bind to this AF9 (D546R) mutant. Furthermore, we can make use of a conditional Cbx8 knock out mouse model (Tan et al., 2011) to knock down native Cbx8 expression followed by transduction of either Cbx8 (WT) or Cbx8 (Valine) mutations to hematopoietic progenitors along with a tagged MLL-AF9. Following the sorting of double positive Cbx8 and MLL-AF9 cells, we can conduct serial replating assays, as we did with our DOT1L mutations (**Section 3.3.2**), and probe for levels of gene expression.

We would ideally expect that the levels of H3K27me3 would be high at the subset of target genes that we have identified (**Section 3.4**), due to the losses in H3K79me2 and H3K79me3 marks. Deshpande, et al. have shown that AF10 mediated knockdown does not increase the levels of H3K27me3 at every gene, even though there are losses in the H3K79 methylation marks. They identify that the majority of the increases in H3K27me3 occur at the promoter regions of *Hoxa7-10* genes. Therefore, some genes appear to be more responsive towards AF10 deletion compared to others (Deshpande et al., 2014). As expected none of their data show genes that have increases in both H3K79 methylation and H3K27 methylation. Based on their data, we highlighted both the genes that we conducted gene expression analysis, and the genes that we found were MLL-

AF9 and H3K79me2 and me3 targets (**Figure 4-7**). As our D546R mutant leads to losses in gene expression, we assumed that those genes in which we conducted gene expression analysis on, would display increases in H3K27me3 levels. The data do show that *Hoxa9* has increases in the H3K27me3 mark. Yet, while we do see some losses in H3K79me2 and me3 levels, at some genes such as *Meis1*, and *Cdk6*, we cannot identify any particular trends in losses of H3K79me2/me3 or increases in H3K27me3 based on AF10 deletion. Unusually, there are genes that show increases in H3K79 methylation and losses in H3K27 upon deletion of AF10, such as *Eya1*.

There is such a high degree of complexity associated with this system that is only starting to become appreciated. The roles of the proteins involved in MLL-AF9 leukemogenesis are only beginning to be understood, and this underscores the importance of fully clarifying the role of the direct recruitment of CBX8 to MLL-AF9.

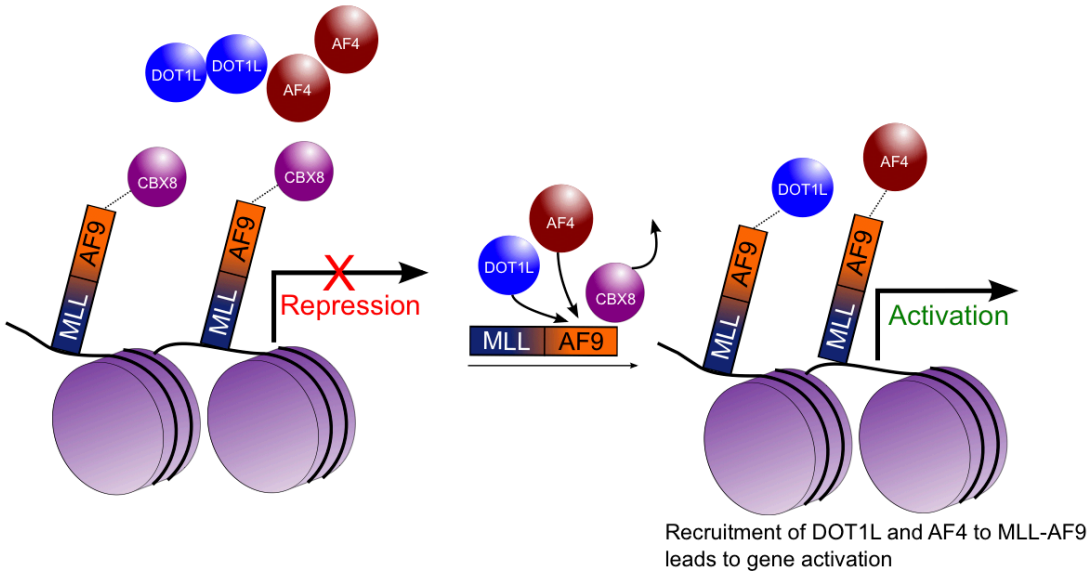


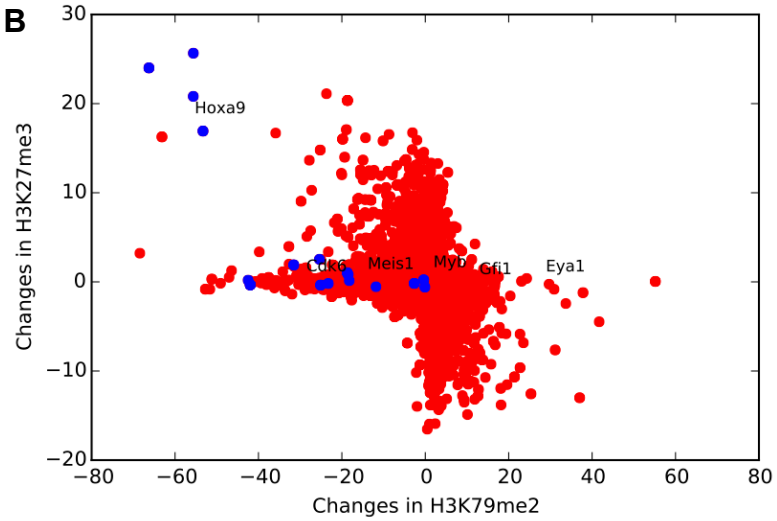
Figure 4-6. Model of protein recruitment to MLL-AF9.

CBX8 is initially bound to MLL-AF9 leading to a default transcriptionally repressed state. Recruitment of activators such as DOT1L or AF4 leads to the displacement of CBX8 from MLL-AF9 leading to gene activation.

A

Description	Differences in H3K27me3	Differences in H3K79me2	Differences in H3K79me3
Hoxa7	25.65935351	-55.60845248	-58.14870699
Hoxa3	24.0059816	-66.26247651	-106.4800149
Hoxa10	20.8145406	-55.62425915	-83.44102424
Hoxa9	16.9204035	-53.29630493	-58.34059391
Jmjd1c	2.53490927	-25.37303187	-78.0012548
Rdx	1.902016825	-31.54054423	-38.2333751
Erbb2ip	1.024415211	-18.70842535	-39.1441674
Senp6	0.805330795	-18.47563517	-40.39587276
Runx2	0.223421914	-0.379240935	-6.830478895
Zbtb25	0.188006132	-42.47910928	-58.1373536
Arid2	0.1317019	-18.22766572	-35.15900721
Sgk3	-0.173138079	-2.64449606	-0.654231615
Tnpo1	-0.184754772	-23.26478685	-33.49740168
Dcun1d1	-0.33359329	-41.96928677	-64.07008449
Phip	-0.361069596	-25.10907297	-41.65135177
Meis1	-0.544590555	-11.81823136	-32.05215163
Tshz1	-0.585395264	-0.10107279	-3.508198672

B



C

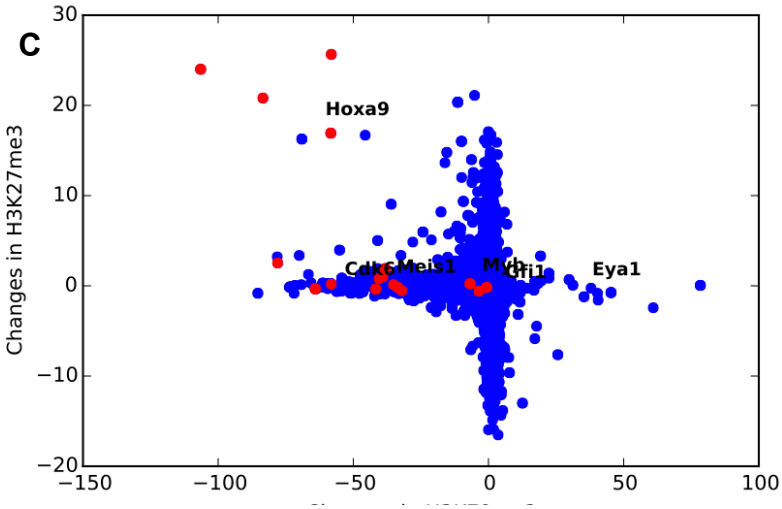


Figure 4-7. Changes in H3K27me3 and H3K79me2 & me3 as a result of AF10 deletion in MLL-AF9 leukemia.

A) Raw data indicating MLL-AF9 target genes, as identified in **Figure 3-9**, that reflect changes in H3K79me2 and me3 patterns and H3K27me3 between floxed and deleted AF10. **B)** The X-axis shows the differences in H3K79 dimethylation between floxed and deleted AF10. Similarly, the Y-axis consists of changes in H3K27 tri-methylation. MLL-AF9 target genes that we identified (blue) and labeled are genes that we conducted gene expression analysis. **C)** Same graph as in **B** but H3K79me3 changes are plotted on the X axis.
These raw data were kindly provided by Dr. Ani Deshpande (Sanford Burnham) for our data analysis (Deshpande et al., 2014).

4.6. Experimental procedures

4.6.1. Protein Cloning, Expression, and Purification

We made use of the pETDuet-1 vector (EMD Millipore) as before (**Section 2.8.1**). CBX8 was cloned into in the first cloning site and AF9 was cloned in the second cloning site of pETDuet-1. Proteins were co-expressed as before in Rosetta 2(DE3) cells (EMD Millipore) in European Molecular Biology Laboratory media. The CBX8-AF9 protein complex was purified using Ni-NTA chromatography and size exclusion chromatography with Superdex 75 column (GE Life Sciences).

4.6.2. Optimization and CBX8-AF9 Complex NMR Resonance Assignments

All protein samples were exchanged into the same buffer employed for NMR data collection containing 25mM Bis-Tris/MES (pH 7.0), 100mM NaCl, 1mM DTT. We increased the pH from 6.0 as we used in other AF9 complexes as this pH allowed for greater solubility. NMR experiments were conducted using a Bruker 600 MHz magnet, a Bruker 800 MHz magnet or a Bruker 900 MHz magnet (samples sent to NMRFAM) equipped with cryogenically cooled probes. All NMR data were processed and analyzed using NMRPipe (Delaglio et al., 1995) and CCPNMR (Vranken et al., 2005). {1H}15N heteronuclear NOE experiments were recorded with 3 second relaxation delays. Residues identified as flexible based on depressed heteronuclear NOE values were removed from the construct and the structured residues were cloned into pETDuet-1 to generate the final CBX8-AF9 construct: CBX8 residues: (aa. 326-349) in the first

pETDuet-1 cloning site, and AF9 residues (aa. 499-568) in the second cloning site.

All NMR experiments for structure determination, double labeled (^{15}N ^{13}C) CBX8- AF9 complex at a concentration of 750 μM was used. The similar suite of NMR experiments was conducted as described in (**Section 2.8.3**). We used the ^{13}C NMR spectra from the 900 MHz magnet at NMRFAM to help guide us with NOE assignments. However, none of the raw data from NMRFAM were used in structure determination.

4.6.3. CBX8-AF9 Complex Structure Determination

First, we calculated preliminary structures using the XPLOR-NIH simulated annealing protocol with restraints derived from dihedral angles and NOE distances (Schwieters et al., 2003). Φ and ψ dihedral angle restraints were generated by TALOS+ based on Cα, Cβ, C′, and N chemical shifts (Shen et al., 2009). NOE assignments were manually assigned. Distance restraints were then generated based on NOE cross-peak intensities and placed into four categories: 1.8−2.8, 1.8−3.3, 1.8−4.2, and 1.8−5.5 Å for structure calculations. NOE violations greater than 0.2 Å were analyzed and appropriately corrected. The lowest energy structure conformation was then selected for a second stage of refinement in XPLOR-NIH. Here, the temperature annealing step was lowered to 4°C. NMR structures were viewed using MOLMOL (Koradi et al., 1996). We did not record RDCs here due to the internal dynamics of the CBX8-AF9 complex. The 10 lowest total energy structures calculated out of 100 were used to represent the ensemble conformation. The structures were displayed using

PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

4.6.4. NMR Relaxation Experiments

T1 and T2 NMR relaxation experiments were conducted to study protein dynamics using relaxation delays of 10, 180, 300, 500, 900, 1300 and 1800ms (T1) and 10, 30, 50, 90, 130, 170 and 230ms (T2). Experiments were conducted on the Bruker 600 MHz magnet.

4.6.5. Fluorescence Polarization (FP) Binding Measurements

FP binding measurements were conducted as before (**Section 2.8.6**). The pure MBP-AF9 protein was titrated into the 5nM CBX8 synthesized peptides. We used the following peptides purchased from Biosyn:

CBX8: FITC-AHX- RPSLIARIPVARILGDPEEE

CBX8 Valine Mutant (A335V): FITC-AHX- RPSLIVRIPVARILGDPEEE

4.7. References for Chapter 4

Biswas, D., Milne, T. a, Basrur, V., Kim, J., Elenitoba-Johnson, K.S.J., Allis, C.D., and Roeder, R.G. (2011). Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. Proc. Natl. Acad. Sci. U. S. A. *108*, 15751–15756.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039–1043.

Creppe, C., Palau, A., Malinverni, R., Valero, V., and Buschbeck, M. (2014). A Cbx8-containing polycomb complex facilitates the transition to gene activation during ES cell differentiation. PLoS Genet. *10*, e1004851.

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*.

Deshpande, A.J., Deshpande, A., Sinha, A.U., Chen, L., Chang, J., Cihan, A., Fazio, M., Chen, C., Zhu, N., Koche, R., et al. (2014). AF10 Regulates Progressive H3K79 Methylation and HOX Gene Expression in Diverse AML Subtypes. Cancer Cell.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574–1577.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol. Cell *45*, 344–356.

Hemenway, C.S., de Erkenez, A.C., and Gould, G.C. (2001). The polycomb protein MPc3 interacts with AF9, an MLL fusion partner in t(9;11)(p22;q23) acute leukemias. Oncogene *20*, 3798–3805.

Herranz, N., Pasini, D., Díaz, V.M., Francí, C., Gutierrez, A., Dave, N., Escrivà, M., Hernandez-Muñoz, I., Di Croce, L., Helin, K., et al. (2008). Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. Mol. Cell. Biol. *28*, 4772–4781.

Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G.A., Li, Z., Vedadi, M., Schapira, M., et al. (2011). Recognition and specificity determinants of the human cbx chromodomains. J. Biol. Chem. *286*, 521–529.

Kerppola, T.K. (2009). Polycomb group complexes – many combinations, many functions. Trends Cell Biol. *19*, 692–704.

Klauke, K., Radulović, V., Broekhuis, M., Weersing, E., Zwart, E., Olthof, S., Ritsema, M., Bruggeman, S., Wu, X., Helin, K., et al. (2013). Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. Nat. Cell Biol. *15*, 353–362.

Knutson, S.K., Kawano, S., Minoshima, Y., Warholic, N.M., Huang, K.-C., Xiao, Y., Kadowaki, T., Uesugi, M., Kuznetsov, G., Kumar, N., et al. (2014). Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity in EZH2-mutant non-Hodgkin lymphoma. Mol. Cancer Ther. *13*, 842–854.

Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. J. Mol. Graph. *14*, 51–55.

Kuzmichev, A. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. *16*, 2893–2905.

Leach, B.I., Kuntimaddi, A., Schmidt, C.R., Cierpicki, T., Johnson, S. a, and Bushweller, J.H. (2013). Leukemia fusion target AF9 is an intrinsically disordered transcriptional regulator that recruits multiple partners via coupled folding and binding. Structure *21*, 176–183.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Maethner, E., Garcia-Cuellar, M.-P., Breitinger, C., Takacova, S., Divoky, V., Hess, J.L., and Slany, R.K. (2013). MLL-ENL inhibits polycomb repressive complex 1 to achieve efficient transformation of hematopoietic cells. Cell Rep. *3*, 1553–1566.

De Napoles, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M., et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev. Cell *7*, 663–676.

Neff, T., Sinha, A.U., Kluk, M.J., Zhu, N., Khattab, M.H., Stein, L., Xie, H., Orkin, S.H., and Armstrong, S.A. (2012). Polycomb repressive complex 2 is required for MLL-AF9 leukemia. Proc. Natl. Acad. Sci. *109*, 5028–5033.

Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu. Rev. Genet. *38*, 413–443.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Sánchez-Beato, M., Sánchez, E., González-Carreró, J., Morente, M., Díez, A., Sánchez-Verde, L., Martín, M.C., Cigudosa, J.C., Vidal, M., and Piris, M.A. (2006). Variability in the expression of polycomb proteins in different normal and tumoral tissues. A pilot study using tissue microarrays. Mod. Pathol. *19*, 684–694.

Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., Michaud, A., Lombard, B., da Rocha, S.T., Offer, J., et al. (2015). Jarid2

Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. Mol. Cell *57*, 769–783.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. Mol. Cell *42*, 330–341.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Marius Clore, G. (2003). The Xplor-NIH NMR molecular structure determination package. J. Magn. Reson. *160*, 65–73.

Senthilkumar, R., and Mishra, R.K. (2009). Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. BMC Genomics *10*, 549.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR *44*, 213–223.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat. Rev. Mol. Cell Biol. *10*, 697–708.

Tan, J., Jones, M., Koseki, H., Nakayama, M., Muntean, A.G., Maillard, I., and Hess, J.L. (2011). CBX8, a Polycomb Group Protein, Is Essential for MLL-AF9-Induced Leukemogenesis. Cancer Cell *20*, 563–575.

Tanaka, S., Miyagi, S., Sashida, G., Chiba, T., Yuan, J., Mochizuki-Kashio, M., Suzuki, Y., Sugano, S., Nakaseko, C., Yokote, K., et al. (2012). Ezh2 augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia. Blood *120*, 1107–1117.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687–696.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750–756.

Chapter 5. Perspectives & Future Directions

We have shown that the degree of DOT1L recruitment to MLL-AF9 defines the level of hematopoietic transformation. Previous work has shown that DOT1L is necessary for MLL-AF9 driven leukemia, but has not addressed whether this was due to direct recruitment to MLL-AF9 or to more global effects of loss of DOT1L. In (**Chapter 2**) we found that there were three separate DOT1L interacting sites with AF9 with varying binding affinities. We showed that these separate complexes fold into similar structures and present the first NMR solution structure of the DOT1L-AF9 complex. In (**Chapter 3**) we use our structural studies to make point mutations in both AF9 and DOT1L to show that transformation potential is dependent on the number of DOT1L and AF9 interacting sites. Furthermore, only small subsets of genes, the majority of which are MLL-AF9 targets, are affected upon blocking this protein-protein interaction. Partial loss of DOT1L binding was shown to impact levels of H3K79me3 but not H3K79me2 in the gene body of affected genes. More complete loss of binding led to reduction in H3K79me2 and complete loss of H3K79me3. Our results suggest that the valency of interaction between MLL-AF9 and DOT1L impacts the level of methylation at target genes, with H3K79me3 being more sensitive than H3K79me2 to this. We suggest that, as DOT1L is a distributive enzyme, this is due to increased residence time at sites with higher valency interactions versus sites with lower valency interactions. As DOT1L is the only known H3K79 methyltransferase, H3K79 methylation levels reflect DOT1L occupancy, however our model will likely require further experimental validation using techniques such

as cross-linking kinetic analyses (Poorey et al., 2013) to directly assess DOT1L residence time at specific sites in the genome. However, the lack of a ChIP validated antibody for native DOT1L will likely require the use of tagged DOT1L for such experiments. Based on the ChIP-Seq results and gene expression data, we present evidence for selective roles for H3K79me2 or H3K79me3 at specific genes. Specific roles for these two marks have been shown in lower organisms but not in mammalian cells.

It is possible to envision a mechanism whereby the multivalency of the DOT1L-AF9 interactions may directly facilitate H3K79 methylation through direct binding of the YEATS domain to the nucleosome (**Figure 5-1**). The YEATS domain of AF9 binds to H3K9ac. As the two H3 tails face in the same direction, this would allow for the binding of two AF9 AHD domains to two separate DOT1L sites within the same DOT1L protein. It is important, however, to remember that the YEATS domain is lost in MLL-AF9 fusions.

Lastly, in (**Chapter 4**) we present the protein complex structure of the CBX8-AF9 complex and propose biological studies to elucidate the role of this interaction in MLL-AF9 leukemogenesis. With the exception of CBX8, the other AF9 binding partners have tight binding affinities; this poses a conundrum as to how the binding and exchange events occur. It is possible that due to the disordered nature of this protein, kinetic advantages lie in the displacement of one binding partner with another, meaning that coupled folding and binding increases the rate of exchange. In this mechanism, proposed by Dyson and Wright, dissociation of a partner bound to AF9 would be dependent on the

binding of a competitor (Dyson and Wright, 2005). In this mechanism, there would be an intermediate exchange state that would have all three proteins bound at the same time. NMR experiments such as protein titration, relaxation dispersion and ZZ exchange could allow for the capture of this intermediate state, potentially due to the time regime of this event. We aimed to obtain kinetic data through the use of Surface Plasmon Resonance experiments to not only gain insights into the on/off rates of these protein interactions, but to attempt to measure whether a third protein binding partner would significantly increase the kinetic off rate, as proposed by Dyson and Wright. However we were unsuccessful, largely due to the intractability of working with our peptides. Future efforts should lie in optimizing the binding of our peptides to the SPR chip, possibly through amine coupling versus using an affinity tag. Another approach would be to use single-molecule fluorescence resonance energy transfer (smFRET), a technique that has already been used to study the coupled folding and binding of IDPs and proteins can be used at extremely low concentrations (Ferreon et al., 2013).

As AF9 binds both activators (AF4 and DOT1L) and repressors (CBX8 and BCoR) of gene expression, it suggests that it functions as a signaling hub that provides different outputs depending on the binding partner. It is not clear, however, how this “dance” of binding partners is orchestrated. Understanding the biological order of events perhaps through using a technique such as time-resolved ChIP may be an avenue to explore. Additionally, experiments such as ChIP-Seq or ChIP-qPCR on H3K79me2/3 and H3K27me3 with our AF9 mutants

as well as with the DOT1L alanine mutants are necessary to fully understand the readouts from our biological experiments. As it has been recently shown that H3K79me1 levels increase during differentiation of hematopoietic progenitors (Deshpande et al., 2014), it is possible that we would observe an increase in H3K79me1 levels and not H3K27me3 levels as a result of our AF9 mutations. Thus future genomic experiments including this H3K79me1 mark may lead to valuable insights into this system. Lastly, it would be interesting to understand the role of H4K16ac in this system as it has shown to play a role in H3K79 trimethylation.

Further studies are necessary to understand similar interactions in other MLL-fusion leukemias, but it is reasonable to expect that MLL-ENL fusions will behave in a similar manner. MLL-AF9 and MLL-ENL leukemias are highly aggressive and patients often suffer from early relapse after treatment (Krivtsov and Armstrong, 2007). The most promising current therapeutic for MLL-fusion leukemias, EPZ-5676, targets the enzymatic activity of DOT1L. Not surprisingly, due to its genome wide role in regulation of transcription, DOT1L inhibition has an effect on many genes, the long-term effects of which is not yet clear (Basavapathruni et al., 2014; Daigle et al., 2011, 2013; Klaus et al., 2014). We propose that the DOT1L interactions with either AF9 or ENL would be excellent therapeutic targets for MLL-AF9 and MLL-ENL leukemias. However, from a drug design perspective this poses a problem: How do you target a protein that is unstructured? Even though we may find “hits” from fragment libraries, it still may be difficult to design structure-activity relationship studies with this system.

Nonetheless, there have been several recent success targeting IDPs, from designing small molecule inhibitors for multiple independent binding sites on c-Myc (Hammoudeh et al., 2009), to targeting the disordered C terminus of PTP1B with an allosteric inhibitor (Krishnan et al., 2014). A more successful approach, perhaps, would be to target the protein-protein interaction *in vivo.* One group has attempted to design molecules that mimic the conformation of intrinsically disordered regions as they are bound to their partner protein (Lao et al., 2014a, 2014b). Blocking this DOT1L-AF9 protein interaction is likely to have distinct advantages over inhibiting DOT1L enzymatic activity, as only a very limited number of MLL-AF9 target genes show significant losses in H3K79 methylation marks, thus the scale of the effects on gene expression would be limited to these genes and those which are targets of wildtype AF9 and ENL regulation.

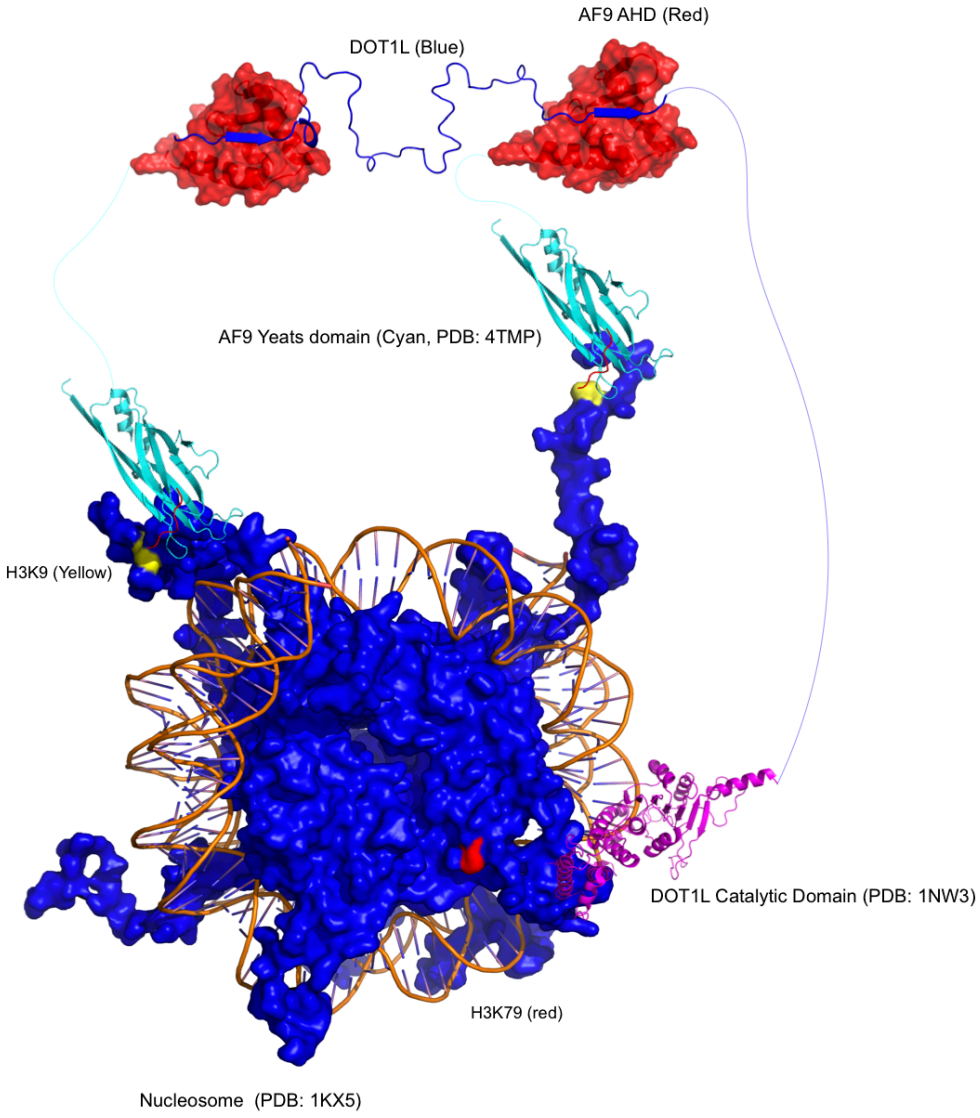

AF9 AHD (Red)
DOT1L (Blue)
AF9 Yeats domain (Cyan, PDB: 4TMP)
H3K9 (Yellow)
DOT1L Catalytic Domain (PDB: 1NW3)
H3K79 (red)
Nucleosome (PDB: 1KX5)

Figure 5-1. Model of AF9 mediated recruitment of DOT1L to the nucleosome.

The AF9 YEATS domain binds to acetylated Histone H3 Lysine 9 (H3K9ac). The two H3 tails face in the same direction, potentially providing a mechanism whereby two AF9 proteins can bind to H3K9ac via the YEATS domain. Subsequently, the two AF9 (AHD) domains can each bind to a single DOT1L binding site, as there are multiple DOT1L binding sites to AF9, which would allow for H3K79 methylation through the catalytic domain. The several hundred amino acid stretch between the DOT1L-AF9 binding sites and the catalytic domain are disordered. Thus, the binding of the YEATS domain to the nucleosome can potentially lead to direct H3K79 methylation via AF9 AHD binding of the DOT1L repeat motifs.

5.1. References for Chapter 5

Basavapathruni, A., Olhava, E.J., Daigle, S.R., Therkelsen, C.A., Jin, L., Boriack-Sjodin, P.A., Allain, C.J., Klaus, C.R., Raimondi, A., Scott, M.P., et al. (2014). Nonclinical pharmacokinetics and metabolism of EPZ-5676, a novel DOT1L histone methyltransferase inhibitor. Biopharm. Drug Dispos. 35, 237–252.

Daigle, S.R., Olhava, E.J., Therkelsen, C.A., Majer, C.R., Sneeringer, C.J., Song, J., Johnston, L.D., Scott, M.P., Smith, J.J., Xiao, Y., et al. (2011). Selective Killing of Mixed Lineage Leukemia Cells by a Potent Small-Molecule DOT1L Inhibitor. Cancer Cell 20, 53–65.

Daigle, S.R., Olhava, E.J., Therkelsen, C.A., Basavapathruni, A., Jin, L., Boriack-Sjodin, P.A., Allain, C.J., Klaus, C.R., Raimondi, A., Scott, M.P., et al. (2013). Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. Blood 122, 1017–1025.

Deshpande, A.J., Deshpande, A., Sinha, A.U., Chen, L., Chang, J., Cihan, A., Fazio, M., Chen, C., Zhu, N., Koche, R., et al. (2014). AF10 Regulates Progressive H3K79 Methylation and HOX Gene Expression in Diverse AML Subtypes. Cancer Cell.

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6, 197–208.

Ferreon, A.C.M., Ferreon, J.C., Wright, P.E., and Deniz, A.A. (2013). Modulation of allostery by protein intrinsic disorder. Nature 498, 390–394.

Hammoudeh, D.I., Follis, A.V., Prochownik, E. V., and Metallo, S.J. (2009). Multiple Independent Binding Sites for Small-Molecule Inhibitors on the Oncoprotein c-Myc.

Klaus, C.R., Iwanowicz, D., Johnston, D., Campbell, C.A., Smith, J.J., Moyer, M.P., Copeland, R.A., Olhava, E.J., Scott, M.P., Pollock, R.M., et al. (2014). DOT1L inhibitor EPZ-5676 displays synergistic antiproliferative activity in combination with standard of care drugs and hypomethylating agents in MLL-rearranged leukemia cells. J. Pharmacol. Exp. Ther. 350, 646–656.

Krishnan, N., Koveal, D., Miller, D.H., Xue, B., Akshinthala, S.D., Kragelj, J., Jensen, M.R., Gauss, C.-M., Page, R., Blackledge, M., et al. (2014). Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. Nat. Chem. Biol. 10, 558–566.

Krivtsov, A. V, and Armstrong, S.A. (2007). MLL translocations, histone modifications and leukaemia stem-cell development. Nat. Rev. Cancer 7, 823–833.

Lao, B.B., Drew, K., Guarracino, D.A., Brewer, T.F., Heindel, D.W., Bonneau, R., and Arora, P.S. (2014a). Rational design of topographical helix mimics as potent inhibitors of protein-protein interactions. J. Am. Chem. Soc. 136, 7877–7888.

Lao, B.B., Grishagin, I., Mesallati, H., Brewer, T.F., Olenyuk, B.Z., and Arora, P.S. (2014b). In vivo modulation of hypoxia-inducible signaling by topographical helix mimetics. Proc. Natl. Acad. Sci. 111, 7531–7536.

Poorey, K., Viswanathan, R., Carver, M.N., Karpova, T.S., Cirimotich, S.M., McNally, J.G., Bekiranov, S., and Auble, D.T. (2013). Measuring chromatin interaction dynamics on the second time scale at single-copy genes. Science 342, 369–372.