Characterization of the Protein-Protein Interactions between Histone Methyltransferase DOT1L and MLL Fusion Proteins towards Developing Small Molecule Inhibitors

by

Chenxi Shen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
In The University of Michigan
2013

Doctoral Committee:

Assistant Professor Zaneta Nikolovska-Coleska, Chair

Associate Professor Jason E. Gestwicki, University of California, San Francisco

Professor Jay L. Hess

Assistant Professor Jeanne A. Stuckey

Associate Professor Raymond C. Trievel

DEDICATION

To my dearest parents

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Nikolovska-Coleska, for providing me the opportunity to learn and grown as a young scientists, for inspiring me to strive for my personal best in research, for encouraging me to think independently and creatively, and for guiding me to develop and pursue my scientific interests persistently. I would also like to thank my committee members, Drs. Jay L. Hess, Raymond C. Trievel, Jeanne Stuckey and Jason E. Gestwicki, for generously offering their time and invaluable insights during the past years.

I want to acknowledge all the current and previous Nikolovska-Coleska lab members, including Garrett Gibbons, Fardokht Abulwerdi, Ahmed Mady, Meilan Liu, Dr. Lei Miao, Dr. Chenzhong Liao, Dr. Julie Di Bernardo and, Dr. Naval Bajwa. It is a pleasant experience to work along such a group of fun, intelligent, kind and hard-working scientists. I thank Drs. Tomasz Cierpicki, Jolanta Grembecka, Andrew Muntean, Yali Dou and their labs for sharing their input at lab meetings. I thank Dr. Shaomeng Wang for allowing me access to the peptide synthesizer in his lab. I thank Xu Ran in Dr. Wang's lab for the help of peptide synthesis. I thank Martha Larsen, Steven Swaney, and Dr. Paul Kirchhoff for their help with high-throughput screening and data triage. Especially thank Martha for

her great expertise and patience for the HTS design and optimization. I thank Dr. Venkatesha Basrur for the mass spectrometry data analysis. I thank Dr. Jennifer Meagher and Krishnapriya Chinnaswamy for the help of protein purification. I thank George Lund for the help of setting up NMR experiments. I thank Dr. Bill Milne for editing this dissertation.

I would like thank my cohorts, especially Ningkun, Brian and Hugo, helped me in many ways when I first come to the United States. I would like to thank the Chemical Biology Doctoral Program for offering me the opportunities to start my graduate studies.

To my friends, both the old ones in China, and the new ones I've made in the last five years, thanks for all of your help and supports. I especially would like to thank Zhonglan Gao a great friend who helped me a lot in research and graduate school application. I also like to thank Lyra, Yipei, Xiao, Wenjing, Jing, Xin, Bo, Yaru, Liuling and Jingjing for being so nice and supportive to me.

Last and most important, I want thank my dearest dad and mom for being such supportive and understanding parents. Your love, patience, and wisdom guide me through every stage of my life. Thank you for everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AF4	ALL1-fused gene from Chromosome 4
ABI-1	Abelson interactor 1
AF10	ALL1-fused gene from chromosome 10
AF17	ALL1-fused gene from chromosome 17
AF6	ALL-1 fused gene from chromosome 6
AF9	ALL1-fused gene from chromosome 9
AUC	Analytical ultracentrifugation
BCoR	BCL-6 Co-repressor
Bpa	Benzophenone alanine
CBP	CREB-binding protein
CBX8	Chromobox protein homolog 8
CD	Circular dichroism
DCM	Dichloromethane
DIEA	*N,N*-Diisopropylethylamine
DOT1L	Disrupter of telomere silencing 1-like
EDC	*N*-(3'-Dimethylaminopropyl)-*N'*-ethylcarbodiimide
ELISA	Enzyme-linked immunosorbent assay
ENL	Eleven nineteen leukemia

FP	Fluorescence polarization
FRET	Fluorescence resonance energy transfer
5-FAM	5-Carboxyfluorescein
GEM	Group epitope mapping
HATU	O-(7-azabenzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate
HOAT	1-Hydroxy-7-azabenzotriazole
Hox	Homeobox
HSQC	Heteronuclear Single Quantum Coherence
HTS	High-throughput screening
ITC	Isothermal Calorimetry
LIC	Ligation independent cloning
MLL	Mixed lineage leukemia
mP	millipolarization
Mtt	4-methyltrityl
NHS	*N*-hydroxysuccinimmide
NMP	*N*-Methyl-2-pyrrolidone
NMR	Nuclear magnetic resonance
Pc3	Polycomb 3 homolog
PDB	Protein Data Bank
PPI	Protein-protein interaction
pTEFb	The positive transcription elongation factor
pTEFb	positive elongation factor b
SAM	S-adenosyl methionine

SET	Su(var)3-9 (a suppressor of variegation 3-9), E(z) (enhancer of zeste), and trithorax
SPR	Suface plasmon resonance
STD	Saturation Transfer Difference
TFA	Trifluoroacetic acid
YEATS	Yaf9, ENL, AF9, Taf14, Sas5

ABSTRACT

Characterization of the Protein-Protein Interactions between Histone Methyltransferase DOT1L and MLL Fusion Proteins towards Developing Small Molecule Inhibitors

by

Chenxi Shen

Chair: Zaneta Nikolovska-Coleska

The MLL fusion proteins activate target genes in part via recruitment of DOT1L (disruptor of telomeric silencing 1-like), a unique histone H3 lysine 79 (H3K79) methyltransferase. The resulting hypermethylation of H3K79 at the Hox and MEIS1 loci is a pivotal event for leukemogenesis in acute leukemia, suggesting that the protein-protein interactions (PPI) between DOT1L and MLL oncogenic fusion proteins represent a potential therapeutic target. We have initiated efforts in this direction by detailed biochemical and functional characterization of the PPI between DOT1L and the MLL fusion proteins, AF9 and ENL. We determined that full length DOT1L protein binds to AF9 and ENL with K_d of 33nM and 206nM respectively. The binding site in DOT1L was mapped and found that only 10

amino acids in DOT1L are involved in the interaction with AF9 or ENL. The identified DOT1L 10mer peptide directly binds to the AF9/ENL C-terminal domain. Alanine scanning mutagenesis studies of DOT1L 10mer peptide demonstrated that four conserved hydrophobic residues are important for the binding to AF9/ENL. We also found that phosphorylation of Serine 868 in human DOT1L abolished AF9/ENL-DOT1L interaction significantly. To examine the critical role of the DOT1L-AF9 interaction in transformation by MLL fusion protein, we performed a colony forming unit (CFU) assay using MLL-AF9 transformed cells lacking endogenous DOT1L. While introduction of the wild type DOT1L construct can restore CFU formation, the DOT1L construct lacking the 10 amino acids was unable to rescue the colony forming potential of the MLL-immortalized cells. These results demonstrate that PPI of DOT1L with AF9/ENL is required for transformation of MLL-AF9 and suggest that disruption of this PPI is a promising therapeutic strategy for MLL-fusion protein associated leukemia. Based on these studies, we employed high-throughput screening (HTS) approach for the discovery of inhibitors targeting DOT1L recruitment by MLL-fusion proteins. For this purpose we developed and optimized a quantitative, reliable fluorescence polarization (FP) based binding assay (Z' factor of 0.72) using AF9 protein and fluorescein-labeled DOT1L 10mer peptide. With the optimized FP assay, HTS of approximately 101,000 compounds was performed for identification of small molecule inhibitors and 39 compounds demonstrated dose-dependent inhibition. The analysis of the chemical structures of these 39 compounds showed that they can be classified into five clusters with different chemical scaffolds and the rest of

the compounds have unique structures (Figure 3.10). Using our FP assay along with surface plasmon resonance (SPR) and STD NMR based secondary assays, four compounds were confirmed with IC_{50} values ranging from 10 – 50 μM. In summary, we confirmed the requirement for DOT1L recruitment by MLL fusion proteins in leukemogenesis, mapped the AF9 and ENL binding site in DOT1L protein necessary for the PPI and demonstrated the feasibility of development of small molecules targeting the C-terminal hydrophobic domain in AF9 and ENL, indicating it is a “druggable” target. Our work has laid the foundation for further development of both peptidomimetics and small-molecule inhibitors of the DOT1L/MLL fusion protein interaction. Such inhibitors can serve as both chemical tools for investigation of the role of this interaction in leukemogenesis and normal hematopoiesis and as lead compounds for the development of a new class of potential anticancer therapeutics.

Chapter 1

Introduction

1.1 Epigenetics and histone methylation

1.1.1 Epigenetics

Epigenetics is the study of the mechanisms underlying inheritable phenotypic variations caused by sequence-independent DNA alterations. These changes include DNA methylation, small, non-coding RNAs and histone modifications [1]. Post-translational modification of histones provides an important regulatory platform for processes such as gene transcription and repair of DNA damage. Different covalent post-translational histone modifications generate a myriad combination of histone marks, which collectively are termed the 'histone code' [2, 3] and include "marks" such as phosphorylation (P), acetylation (Ac), methylation (Me) and ubiquitination (Ub) [4-6].

Post-translational modification of histones is a dynamic and reversible process regulated by various histone-modifying enzymes, known as 'writers' and 'erasers'. The writers are enzymes that catalyze chemical modifications of histones in a residue-specific manner. Examples include histone methyltransferase and histone acetyltransferase. Erasers are enzymes that

remove chemical modifications from histones, for example, histone demethylase (HDMT) which removes methylation marks and histone deacetylase (HDAC) which removes histone acetylation modifications [2, 3].

The ‘reader’ proteins bind specifically to a certain type or combination of histone modifications and translate the histone code into a meaningful biological outcome, whether transcriptional activation, silencing or other cellular response [7]. Histone readers possess motifs that are capable of specifically interacting with the modified histones. For example, the reader bromodomain can recognize acetylated lysine and plant homeodomain (PHD) finger and chromodomain bind methylated lysine.

Emerging evidence indicates that misregulation of histone modification, caused by deregulation of these histone ‘writers’, ‘erasers’ or ‘readers’, contributes to the initiation and progression of human cancers [8] and these epigenetic players might be potential targets for therapeutic intervention. Clinical success of histone deacetylase inhibitors in cutaneous T cell lymphoma and DNA demethylating agents in myelodysplastic syndrome offers a compelling argument for therapeutic intervention of epigenetic players in the treatment of cancer. The pharmacological doses of an HDAC inhibitor, all-*trans*-retinoic acid, induces relatively specific effects on histone H3 acetylation in PML-RAR α-fusion-positive acute promyelocytic leukemia cells (PML), and the H3 acetylation of the histone underlies both differentiation therapy and epigenetic therapy of PML [9, 10]

1.1.2 Histone methytransferases

Histone methylation is a chemical modification involving the addition of one (mono), two (di) or three (tri) methyl groups to lysine or arginine residues in a histone protein [4] and it is one of the most studied post transcription modifications. Within histones, methylation has been observed at multiple lysine sites, including K4 of histone H3 (H3K4), and H3K9, H3K27, H3K36, H3K79, and H4K20. In general, methylation at H3K9, H3K27, and H4K20 correlates with repression of transcription, while methylation at H3K4, H3K36, and H3K79 correlates with activation of gene transcription [11-13]. Histone methyltransferases (HMTase), the writers responsible for adding the methyl marks to histones, include lysine methyltransferases and arginine methyltransferases. Lysine methyltransferase can be classified based on the presence or absence of the evolutionarily conserved SET domain, named after *Drosophila* Su(var)3-9 (Suppressor of Variegation 3-9), E(z) (Enhancer of Zeste), and Trithorax [14]. The mixed lineage leukemia (MLL) family of histone methyltransferases are typical SET domain lysine methyltransferases, which methylate H3K4 [15], while Disruptor of telomeric silencing 1 like (DOT1L) is an non-SET domain lysine methyltransferase, responsible for H3K79 methylation [16].

Histone methyltransferases are known to play important roles in many biological processes, including cell-cycle regulation, DNA damage [17] and stress response, development and differentiation [18, 19]. Establishment of an appropriate pattern of histone methylation is not only crucial for normal

development and differentiation, but is also intimately associated with tumor initiation and development [4, 20].

1.2 Disruptor of telomeric silencing 1 like (DOT1L), a histone methyltransferase

1.2.1 DOT1L is a unique histone methyltransferase

DOT1L is a novel histone methyltransferase that was first identified in yeast by its ability to dysregulate gene silencing in the vicinity of telomeres [21]. It is the mammalian counterpart of yeast Dot1 and currently the only identified histone 3 lysine 79 (H3K79) methyltransferase [16, 22]. It is distinct from other histone lysine methyltransferases in several ways: 1) DOT1L does not contain the evolutionarily conserved SET domain [23] and is structurally more similar to arginine methyltransferases; 2) it methylates histone H3 lysine 79 (H3K79), which is in a globular domain located on the nucleosome surface and is accessible to the chromatin binding factors; 3) it is capable of introducing mono-,

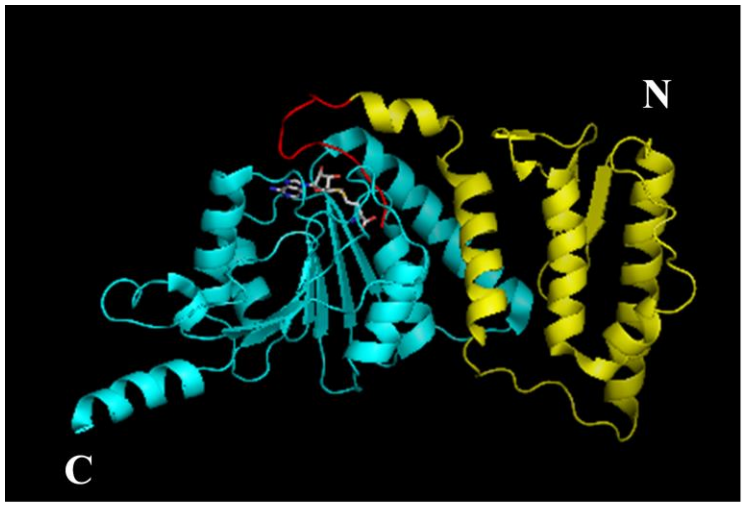


Figure 1.1 Structure of catalytic domain of DOT1L protein (PDB ID: 1WN3). The N-terminal region (aa 1-126) is colored yellow, the open α/β structure (aa 141-332) is shown in cyan, and the loop L-EF connecting the two regions is shown in red. The bound SAM molecule is shown in a ball-and-stick model (carbon, gray; nitrogen, blue; oxygen, red; sulfur, yellow).

di-, and tri-methylation of histone H3K79 in a nonprocessive manner, which means that after each methylation, DOT1L must dissociate and re-associate with K79 to allow cofactor exchange and formation of higher levels of methylation [16, 24]. The enzymatic domain of DOT1L is located on the N-terminus of the protein and the crystal structure of the catalytic domain has been solved [16] (Figure 1.1). Generally, the structure of the catalytic domain is composed of α helical domain of an open α/β structure, including a 7-stranded β sheet. DOT1L belongs to the class I S-adenosyl methionine (SAM)-dependent methyltransferases. The catalytic domain of DOT1L and has resemblance with a variety of classical methyltransferase, for example, the catechol O-methyltransferase (COMPT) (PDB 1VID) [16, 25] In the catalytic domain, there is SAM-binding pocket in proximity to a potential lysine-binding channel and a positively charged, flexible region which is required for nucleosome binding and enzymatic activity of the catalytic domain (Figure 1.1). The C-terminus of DOT1L outside the catalytic domain is involved in protein-protein interactions and it has been reported that DOT1L interacts with AF9/ENL, AF10 and RNA pol II through the C-terminus (Figure. 2.2). These interactions are important for the role of DOT1L in transcription activation and elongation [26-29], which are in details discussed in Chapter 2.

1.2.2 Role of DOT1L in diverse biological processes

DOT1L has diverse biological functions and plays critical roles in multiple biological processes. In addition to its participation in telomere silencing, genome-wide localization studies revealed that H3K79 methylation is enriched in

gene coding regions [30]. DOT1L preferentially occupies the proximal transcribed region of active genes, and is correlated with enrichment of H3K79 di- and tri-methylation [31, 32], implicating a potential role for DOT1L in regulation of transcription. Involvement of DOT1L in transcriptional elongation is supported by fact that DOT1L is associated with proteins with transcription activation and elongation capabilities. The first DOT1L-associated complex, named the ENL-associated proteins (EAP) complex, was purified by immunoprecipitation using antibodies against ENL [26]. The complex contains DOT1L, p-TEFb, AF4, AF5q31, LAF4, Ring1, CBX8 and BCoR [26, 33]. p-TEFb is a cyclin-dependent kinase complex that phosphorylates Ser2 in the RNA Pol II C-terminal domain (CTD). This phosphorylation event is required for the transition of RNA pol II from the transcription initiation to the transcription elongation phase. The EAP complex possesses both H3K79 methyltransferase and RNA Pol II CTD kinase activities. Knockdown of ENL reduced the genome-wide H3K79 me2 level as well as the global transcriptional elongation activity, suggesting a link between H3K79 methylation and transcriptional elongation. Other evidence indicating the role of DOT1L in transcription is the functional link between DOT1L and phosphorylated RNA Pol II. It has been reported that DOT1L directly interacts with the phosphorylated RNA Pol II CTD [29]. Genome-wide profiling analysis indicates that the occupancy of DOT1L largely overlaps with that of RNA Pol II at actively transcribed genes, especially surrounding transcriptional start sites.

Transcriptional activation by DOT1L-mediated H3K79 methylation is also exploited in leukemia stem cells by chromosomal rearrangements. DOT1L is

involved in the pathobiology of mixed lineage leukemia (MLL), which is characterized by chromosomal translocations involving the *MLL* gene. Misregulation of DOT1L occurs because components of the elongation complexes EAP, SEC and AEP interact with DOT1L and are frequent translocation partners of MLL. These oncogenic fusion proteins recruit DOT1L to ectopic loci whose mis-expression contributes to the transformed phenotype. The important role of DOT1L in MLL will be discussed in detail in the following sections.

Mammalian DOT1L is an essential gene that is required for embryogenesis and hematopoiesis [34, 35]. Dot1L-deficient embryos show multiple developmental abnormalities, including growth impairments, angiogenesis defects in the yolk sac, and cardiac dilation, and die 9.5 - 10.5 days post coitum. Embryonic stem (ES) cells from Dot1L mutant blastocysts show global loss of H3K79 methylation capability and reduced levels of heterochromatic marks (H3K9 di-methylation and H4K20 tri-methylation) at centromeres and telomeres, changes which are accompanied by aneuploidy, telomere elongation, and proliferation defects.

The cardiovascular defect observed in mDOT1L knockout embryos [35] suggested that mDOT1L may play a role in cardiac development, which is consistent with *ex vivo* studies demonstrating a correlation between increased H3K79 methylation and an induced expression of cardiovascular marker genes during myocardiogenesis [36]. A cardiac-specific knockout mouse model was

generated to elucidate the role of mDOT1L in cardiac development and function [37]. This cardiac-specific knockout of mDOT1L caused dilated cardiomyopathy (DCM), which indicated that DOT1L plays an important role in cardiac function.

DOT1L function is also vital for hematopoiesis. In a mouse line, in which a gene trap has been used to abrogate DOT1L activity [38], the embryonic lethality occurred at mid-gestation in *Dot1L*-deficient mice. A defect in early erythropoiesis is a primary contributor to this embryonic lethality. Gene expression in hematopoietic progenitor cells of these embryos indicated that G*ATA2* and *PU.1* are genes critical for early hematopoietic fate decisions. The down-regulation of Gata2 in Dot1L-deficient hematopoietic cells caused de-repression of the myeloid transcription factor PU.1 which blocks erythroid differentiation.

Taken together, all these studies show that DOT1L is vital to development and its role under normal cellular conditions as well as transformed conditions is emerging and attracting more attention.

1.3 Mixed lineage leukemia (MLL)

1.3.1 Wild type MLL and MLL fusion protein

MLL is a particularly aggressive hematopoietic malignancy that occurs predominately in pediatric patients. *MLL* gene rearrangements are present in over 70% of infant leukemias [39] and in general account for approximately 5% of acute lymphoblastic leukemia (ALL), 5-10% of acute myeloid leukemia (AML), and a significant portion of mixed lineage leukemia (MLL) cases [40].

Furthermore, almost all secondary leukemias associated with exposure to topoisomerase II targeting drugs involve chromosome translocation at 11q23 [41].

Translocation of the *MLL* gene on chromosome 11q23 is recognized as one of the typical characteristics of MLL. *MLL* gene rearrangements generate a large variety of oncogenic MLL fusion proteins. In 2010, 66 MLL fusion proteins have been identified (http://atlasgeneticsoncology.org/ Genes/MLL.html). Most of the common MLL fusion proteins are nuclear proteins with transcriptional activating activity [42-44]. The most common translocations in ALL are t(11;19) and t(4;11), resulting in the fusion proteins MLL-ENL and MLL-AF4, respectively. In contrast, the t(9;11) translocation, resulting in the MLL-AF9 fusion protein, is more frequently found in AML [42]. In addition to the nuclear translocation partners, another class of MLL fusion partners consists of cytoplasmic proteins that contain dimerization domains, such as AF6. Dimerization of these MLL fusion proteins leads to potent transcriptional activation and is essential for their leukemogenic capacity [45, 46]. The *MLL* gene is approximately 89kb long, consisting of 37 exons [47] and encodes a 3968 amino acid nuclear protein with a complex domain structure. A summary of the wild type MLL protein domain structure is presented in Figure 1.2, and has been described thoroughly in several reviews [15, 42]. Briefly, MLL contains three short AT-hook motifs on the N-terminal region, and these mediate binding to the minor groove of AT-rich genomic DNA sequences [48]. There are two nuclear localization sites (SNL1 and SNL2) immediately adjacent, on the C-terminal side of the AT-hooks that are followed by a DNA methyltransferase (DMT) homology domain that includes a

CxxC zinc-finger motif [49, 50]. There are several PHD (Plant Homology) fingers in MLL [51]. The C-terminus SET domain possesses histone H3 lysine 4 (H3K4) methyltransferase activity [52, 53] and is structurally homologous to *Drosophila melanogaster* trithorax. The MLL protein is cleaved by an aspartic protease named taspase into an N-terminal fragment (MLL^N) and a C-terminal subunit (MLL^C) [54, 55]. The MLL^N fragment contains several functional regions that are considered essential for correct localization of the MLL complex.

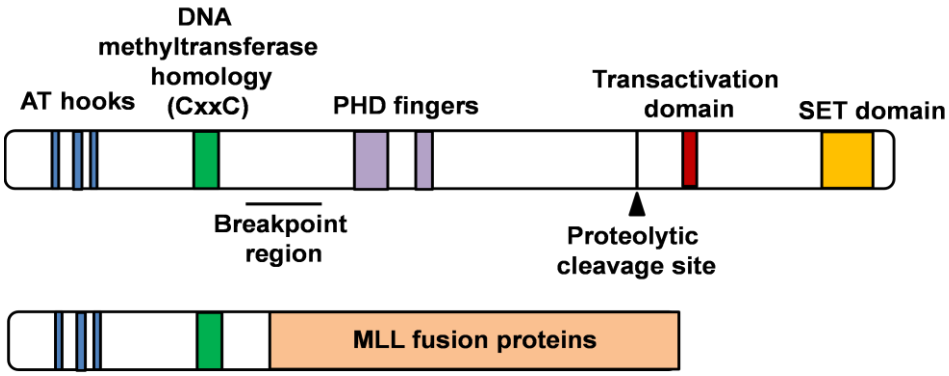


Figure 1.2 Schematic of wild type MLL and MLL-rearranged fusion proteins. Major functional domains and the proteolytic cleavage site associated with wild-type MLL are indicated. MLL fusion proteins consist of the N-terminus of wild-type MLL (up to the breakpoint region) fused in a frame with a translocation partner such as AF9, ENL and AF4, or a cytoplasmic protein, such as AF6.

The MLL^N portion binds to a large number of different proteins. Menin (MEN1) and lens-epthelium-derived growth factor (LEDGF) are two of the proteins interact with MLL^N and are found to be essential for the recruitment of wild type MLL and MLL fusion proteins to gene targets. Direct interactions between Menin and the extreme N-terminus of MLL are essential for MLL-rearranged leukemogenesis [56, 57]. The MLL^C subunit associates with several proteins, including MOF, WDR5, ASH2L and RBBP5, to modify chromatin structure, thereby facilitating transcription activation [58, 59]. All identified MLL

fusions contain the first 8-13 exons of MLL and a variable number of exons from fusion partner genes [60]. The breakpoint region in MLL is immediately next to the CxxC domain, with deletion of the distal C-terminal SET domains and in-frame fusion to one of many different translocation partners (Figure 1.2). The AF4 and ENL protein families are the most frequent MLL fusion partners, accounting for two thirds of MLL associated leukemias [40]. *MLL* gene rearrangements create fusion genes that contain the 5' portion of MLL and the 3' portion of its fusion partner. Those fused gene products cause the sustained expression of MLL target genes, for example, *Hox* genes and consequent enhanced proliferation of hematopoietic progenitors. However, the mechanisms by which the major fusion partners contribute to MLL-rearranged leukemogenesisare only beginning to be defined.

Although MLL fusion proteins have incredible diversity, they have several features in common. First, all the MLL fusion proteins share a common structure with the respective partners invariably fused in a frame to MLL^N right after the CxxC domain but excluding the PHD fingers. Second, all the MLL fusion proteins can be divided into two classes, the nuclear proteins, which include AF4, AF9, ENL, AF10 and ELL, and the cytoplasmic proteins, such as AF6, GAS17 and AP1.

The molecular pathways of the nuclear MLL fusion proteins include histone lysine methylation, acetylation and arginine methylation and involve transcriptional elongation and chromatin modification. The most frequent nuclear

fusion partners of the ENL and AF4 family are members of the EAP complex that combines histone H3K79 methyltransferase activity catalyzed by DOT1L with transcriptional elongation stimulation by pTEFb (positive transcription elongation factor b, which is composed of catalytic subunit Cdk9 and a regulatory subunit, cyclin T). The EAP complex phosphorylates the C-terminal repeat domain of RNA polymerase II. It is speculated that recruitment of one or more of these complexes and introduction of histone lysine methylation by chimeric MLL fusion proteins facilitates sustained expression of MLL-target genes, resulting in leukemic transformation.

In addition to lysine methylation, recent studies have also indicated that fusions of MLL regulate its target gene expression through histone arginine methylation [61, 62]. MLL-EEN was found to recruit histone H4 arginine 3-methyltransferase (*PRMT1*) through the adaptor protein SAM68. Additionally, this modification has been shown to be correlated with an increased histone acetylation, suggesting cross talk in the MLL molecular mechanism between acetylation and arginine methylation.

Fusions of MLL with the histone acetyltransferases CREB binding protein (CBP) and the related p300 have also been observed [63-65]. Structure function analyses clearly singled out the bromo- and histoneacetyltransferase domains of CBP as necessary and sufficient for the oncogenic function of the respective fusion proteins.

Unlike nuclear proteins, the cytoplasmatic MLL fusion proteins, including AF6, GAS17 and AFp1, employ different mechanisms for the leukemogenic transformation. Studies indicate that cytoplasmatic MLL fusion proteins dimerize *via* coiled-coil or other dimerization domains provided by the fusion partners. The dimerization plays an important role in these MLL transformed leukemias [45, 46]. However, the mechanism by which dimerization of the fusion protein activates the target gene has not been delineated.

1.3.2 DOT1L and MLL fusion proteins

Although the leukemogenic mechanisms underlying MLL fusion forms are poorly understood, recent studies have begun to unveil a common transformation pathway for the most common MLL fusion forms. DOT1L was revealed to be a very important protein involved in MLL transformed leukemia.

Table 1.1 Immunoprecipitation complexes of MLL fusion protein complexes

Complex	Components	Ref.
EAP	ENL, AF4, p-TEFb, DOT1L, AF9, AF5q31, CBX8, BCoR, HSP70, RING1	[26]
AEP	ENL, AF9, AF4, p-TEFb, AF5q31	[44]
SEC	ENL, AF4, p-TEFb, AF9, AF5q31, ELL1, ELL2, ELL3	[70]
DotCom	ENL, DOT1L, AF9, AF10, AF17, TRRAP, SKP1, β-Catenin	[68]

Aberrant induction of H3K79me was observed in leukemia-promoting oncogenes, such as *Hoxa9*, in leukemia cells transformed by MLL-AF9, MLL-ENL, MLL-AF4 and, MLL-AF10, which represent the most common MLL fusion forms. It has been reported that MLL-ENL, MLL AF9, MLL-AF10 and CALM-

AF10 fusions recruit DOTL directly. Mutations of MLL-ENL, MLL-AF9 or MLL-AF10 that disrupt the interactions with DOT1L also abolish leukemia transformation [26, 28, 44]. Recent biochemical studies have further revealed that DOT1L actually associates itself with a long list of factors that are known MLL fusion partners. These include ENL [26, 44, 66], AF9 [66, 67], AF10 [28, 68], AF17[68], AF4 (also known as MLLT2 or AFF1) [26, 69], AF5q31 (also known as MCEF or AFF4) [26] and LAF4 [26]. These MLL fusion proteins associate directly or indirectly with DOT1L and form multiple protein complexes. As summarized in Table 1.1, the reported MLL fusion protein complexes include the ENL associated protein complex (EAP) containing ENL, AF4, DOT1L and pTEFb [26]; the super elongation complex (SEC), containing AF4, AF9, ENL, AFF4, ELL1 and pTEFb [70], AEP (containing AF4, AF5, ENL and pTEFb) [44], and DotCom (containing DOT1L, AF9, ENL, AF10, AF17 along with several Wnt signaling pathway components) [68]. AF9, ENL, AF10 and AF17 were identified as stable components of DOT1L-containing complexes [68].

In an intriguing model, DOT1L is responsible for aberrant transcription in many MLL fusion-induced leukemias. AF9 and ENL are the shared components of the above complexes [26, 44, 68, 70] and they can interact directly with DOT1L. DOT1L-AF10-ENL/AF9 complexes are also further associated *via* a protein-protein interaction network, with a transcription elongation-promoting complex, AEP/EAP (which contains AF4, AF5q31, ENL and p-TEFb). Together with DOT1L, recruitment of p-TEFb transcription elongation complexes promotes

the transcription of downstream targets such as *Hox* genes. These putative mechanism models of various MLL fusion protein complexes are illustrated schematically in Figure 1.3.

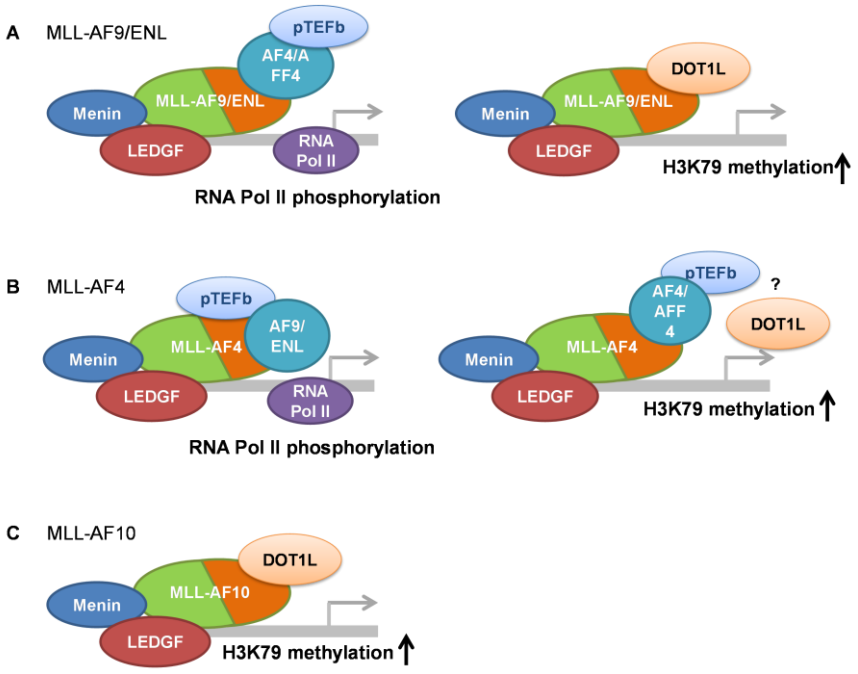


Figure 1.3 Putative model of DOT1L recruitment by MLL fusion proteins. A. MLL-AF9/ENL fusion protein can exist in a complex either with DOT1L or with AF4/AFF4. Association of AEP leads to phosphorylation and transcription elongation of RNA Pol II, and direct recruitment of DOT1L by MLL-AF9/ENL, introducing H3K79 methylation. B. MLL-AF4 fusion proteins can recruit pTEFb for RNA Pol II phosphorylation and transcription elongation. However, since AF4 does not interact with DOT1L directly, recruitment of DOT1L by MLL-AF4 might through an indirect protein-protein interaction, *e.g.* pTEFb. C. In MLL-AF10 transformed leukemia, MLL-AF10 interacts directly with DOT1L and recruits DOT1L for downstream gene transcription.

One of the common features of MLL-transformed leukemias is the involvement of transcription initiation and elongation factor pTEFb and histone H3K79 metyltransferase DOT1L. In MLL-ENL, the MLL-AF9/ENL fusion protein can exist in a complex either with DOT1L or with AF4/AFF4. Association of AEP

leads to RNA Pol II phosphorylation and transcription elongation, and direct recruitment of DOT1L by MLL-AF9/ENL leading to H3K79 methylation. MLL-AF4 fusion proteins can recruit pTEFb for RNA Pol ll phosphorylation and transcription elongation but since AF4 does not interact with DOT1L directly, recruitment of DOT1L by MLL-AF4 might proceed through indirect protein-protein interactions. In this connection, it has been reported that DOT1L interacts with the phosphorylated RNA Pol II CTD [29], which might be a bridge for recruitment of DOT1L by MLL-AF4. In MLL-AF10 transformed leukemia, MLL-AF10 interacts directly with DOT1L and recruitment of DOT1L for downstream gene transcription [28].

1.3.3 *Hox* genes are crucial downstream targets of MLL

The clustered homeobox (*Hox*) genes have been identified as targets of mammalian MLL [71-73]. In humans, a total of 39 *Hox* genes have been identified. They are linked in four separate clusters*, Hoxa*, *Hoxb*, *Hoxc* and *Hoxd*, located on chromosomes 7, 17, 12 and 2, respectively [74]. Based on sequence similarities and their positions within the clusters, they are further classified into 13 groups laid out from position 1 to 13 in a 3'-5' direction [75, 76]. HOX proteins constitute a particular subgroup of homeodomain transcription factors that serve critical functions in embryonic development [72] and normal hematopoietic differentiation [77, 78]. The A cluster *Hox* genes, including *Hoxa7* and *Hoxa9* and the *Hox* cofactor *Meis1*, are normally involved in early hematopoietic stem cells [79]. *Hoxa9* is an important regulator of hematopoiesis and acts, in part, by promoting stem-cell renewal [80]. The deregulation of *Hox* genes appears to be

the most crucial factor for MLL fusion-induced leukemogenesis [81, 82]. Conditionally transforming this version of MLL–ENL shows that the expression of *Hoxa9* and *Meis1* can replace the gain-of-function activity of the fusion proteins, indicating that these are crucial MLL fusion protein targets [81]. Additionally, retroviral transformation assays show convincingly that MLL-ENL transformation requires functional *Hoxa7* and *Hoxa9* [83].

1.4 DOT1L is a validated target in MLL

1.4.1 Role of DOT1L in MLL

Transcriptional activation by MLL fusion proteins is accompanied by a dramatic increase of histone H3 lysine 79 (H3K79) methylation levels at the prominent MLL-fusion downstream target loci Hoxa9 and Meis1 in MLL-ENL transformed hematopoietic progenitors [84]. Subsequent genome-wide analysis revealed a distinct pattern of H3K79 methylation in an MLL–AF4 mouse model, and in human MLL-rearranged primary leukemia samples when compared to normal proB cells and leukemias with other cytogenetic abnormalities [85]. This link was further strengthened by analysis of H3K79 methylation patterns on MLL–AF4 direct target loci in a human MLL-rearranged leukemia cell line, and MLL–AF9 direct targets in a mouse model [86, 87]. All this evidence indicated that MLL-fusion target genes are associated with histones methylated at H3K79, and H3K79 methylation patterns on MLL–AF9 target loci appear abnormal [87]. Since DOT1L is the only known histone methyltransferase that is capable of introducing

H3K79 methylation, the role of DOT1L in MLL transformed leukemia was studied extensively. In recent years, several groups have used shRNA-knockdown, or conditional knockout models to confirm that DOT1L is required for MLL-fusion-mediated leukemic transformation, and *in vivo* leukemia development and maintenance [34, 87-90].

Analysis of H3K79 methylation profiles and gene-expression changes after loss of DOT1L in MLL–AF9 driven leukemia cells revealed that only a small subset of genes demonstrate a change in expression shortly after DOT1L inactivation. However, this set of genes is highly enriched for MLL–AF9 targets and other genes with known functional significance in MLL-rearranged leukemia biology [87]. It appears that functional DOT1L is required for the most common MLL-rearranged leukemias, and displays specificity with respect to several other types of leukemia.

MLL fusions are potent oncogenes and they alter gene expression in hematopoietic cells through interactions with DOT1L. The MLL-ENL or MLL-AF9 interaction with DOT1L is critical in the transformation capability of the fusion protein, since deletion or mutations in the AF9 or ENL C-terminal domain, which is necessary for DOT1L interaction, eliminates the transformation capability of MLL fusion protein [26, 27]. Additionally, transient knockdown of ENL diminished genome-wide as well as HOXA9 gene-specific H3K79 dimethylation, providing further circumstantial evidence of a critical role for the ENL - DOT1L interaction in leukemia related to MLL-ENL [26].

DOT1L is involved in transformation by other MLL fusion proteins In addition to MLL-AF9 and MLL-ENL. These include MLL-AF4 [85], MLL-AF10 [28] and MLL-AF6 [91]. In MLL-AF10 transformed leukemia, both MLL-AF10 and H3K79 methylation were specifically enriched at the Hoxa9 locus in MLL-AF10 transformed cells. The transformation capability of MLL-AF10 was dependent on DOT1L enzymatic activity and recruitment of DOT1L by interaction with MLL-AF10 is important for MLL-AF10 leukemia, since disruption of the interaction between MLL-AF10 and DOT1L abolished the MLL-AF10 transformation [28]. Although there is no evidence indicating the direct interaction between DOT1L and MLL-AF4 or MLL-AF6, there is a link between DOT1L and these MLL fusion proteins since knockdown of DOT1L significantly interferes with MLL-AF4 induced transformation and also with the activation of *Hox* genes [85]. Targeted disruption of Dot1l using a conditional knockout mouse model also inhibited leukemogenesis mediated by the MLL-AF6 fusion oncogene [91].

It is clear that DOT1L is broadly involved in leukemiogenesis, particularly when mediated by MLL fusion proteins. However, the exact mechanism underlying the contribution of DOT1L to the leukemogenic gene activation process requires further elucidation. DOT1L has been associated with protein complexes with transcription elongation activities [26, 27, 69] and evidence that DOT1L associates with MLL fusion partners [26, 67, 69, 73, 92] might help to elucidate how DOT1L is involved in MLL hematological malignancy. The discovery of DOT1L related complexes such as EAP, DotCom, or AEP (Table 1.1), prompted the hypothesis that members of the complex, when fused to MLL,

misguided DOT1L and pTEFb to MLL target loci. DOT1L would then presumably methylate H3K79 to induce an open chromatin formation, and pTEFb would stimulate RNA Pol II to transcribe the respective locus. DOT1L and pTEFb appear to be the 'effector units' of these complexes. A suggested general mechanism by which DOT1L contributes to the leukemogenesis process is shown in Figure 1.3 and can be described as interaction between DOT1L and MLL fusion partners leading to mistargeting of DOT1L to the gene targets of MLL fusion proteins, such as the *Hoxa* cluster and Meis1. Aberrant hypermethylation of H3K79 leads to constitutive transcriptional activation of these genes, which in turn results in leukemic transformation.

1.4.2 Rationale and strategies for targeting DOT1L functions

The crucial role for DOT1L in *MLL*-rearranged leukemias is now well established and is strongly supported by functional evidence demonstrating that a loss of DOT1L profoundly affects the leukemogenic gene expression program and leukemogenic activity of *MLL* fusion-transformed cells, while several non-*MLL* leukemia cells are unaffected [85]. The development of small molecule inhibitors for DOT1L is clearly warranted because such molecules will help to address some of the outstanding mechanistic questions, and most importantly, they may change the outlook for patients with this devastating type of leukemia.

A critical step in transformation and disease progression in MLL is the aberrant recruitment of DOT1L by MLL fusion proteins. Two plausible strategies for therapies against DOT1L-related leukemia are: 1) specifically targeting the

enzymatic activity of DOT1L in transformed cells; and 2) blocking the protein-protein interactions between DOT1L and oncogenic fusion proteins. The recent development [88] of a DOT1L enzymatic inhibitor, a small molecule inhibitor, EPZ004777 which competes with S-adenosyl methionine, provided proof of principle for the development of DOT1L enzymatic activity inhibitors as targeted therapeutics for MLL-rearranged leukemia. Treatment of MLL cells with EPZ004777 selectively inhibits H3K79 methylation and blocks expression of leukemogenic genes. Exposure of leukemic cells to EPZ004777 results in selective killing of cells bearing the MLL gene translocation, with little effect on non-MLL-translocated cells. Additionally, *in vivo* administration of EPZ004777 leads to extension of survival in a mouse MLL xenograft model. These results provide compelling support for DOT1L inhibition as a basis for targeted therapeutics against MLL. However, it is important to note that potential inhibitors of DOT1L enzymatic activity must be used with caution since DOT1L is the only known H3K79 methyltransferase and is required for survival of normal somatic cells. Constitutive and conditional knockout studies of DOT1L, have shown it to be essential for embryonic development, prenatal and postnatal hematopoiesis and cardiac function [34, 35, 38, 87, 90]. This universal and essential function of DOT1L in multiple cell types [93] suggests that direct inhibition of DOT1L histone methyltransferase activity might be toxic. Consequently, development of therapeutic strategies allowing selective inhibition of DOT1L function is important and necessary. The disruption of DOT1L interactions with oncogenic fusion proteins is an attractive targeting approach, because 1) recruitment of DOT1L by

MLL fusion protein is important for MLL transformation; 2) targeted disruption of DOT1L interactions with MLL fusion proteins might be more selective to leukemia cells compared to universal inhibition of DOT1L enzymatic activity.

1.5 Summary

In summary, genetic and biological studies support DOT1L as a *bona fide* drug target for treatment of MLL. Besides the enzymatic function of DOT1L, the protein-protein interactions (PPIs) between DOT1L and MLL fusion proteins are also interesting potential targeting sites.

Protein-protein interactions are emerging as promising drug targets and reports of PPI inhibitors have increased in the last few years[94]. The first step toward development of inhibitors of PPIs is to define the interaction sites and residues crucial to the interaction. To date, interactions of MLL fusion protein with DOT1L were only observed in Co-IP and yeast two hybrid approaches and there is lack of further understanding of this PPI. In this dissertation work, we have addressed these key questions regarding MLL fusion protein AF9/ENL and DOT1L PPIs. In Chapter 2, we have undertaken the initial steps toward characterized the AF9/ENL-DOT1L PPIs, including mapping the interaction site, identifying the crucial residues for interaction, and proving that the identified interaction site was important for MLL transformation. In Chapter 3, we applied a high-throughput screening (HTS) approach to identify small molecule inhibitors of the AF9-DOT1L interaction. Toward this goal, we have developed and optimized

fluorescence polarization based binding assays for the AF9-DOT1L interaction. The initial HTS for potential inhibitors of the DOT1L-MLL fusion protein interaction were performed and the identified hits were further characterized by FP and SPR-based assays. Finally in Chapter 4, we discuss the implications of these studies and potential future strategies for targeting MLL fusion proteins and DOT1L interactions.

1.6 References

1. Berdasco, M. and M. Esteller, *Aberrant epigenetic landscape in cancer: how cellular identity goes awry.* Dev Cell, 2010. **19**(5): p. 698-711.
2. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-5.
3. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.
4. Chi, P., C.D. Allis, and G.G. Wang, *Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers.* Nat Rev Cancer, 2010. **10**(7): p. 457-69.
5. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.
6. Gelato, K.A. and W. Fischle, *Role of histone modifications in defining chromatin structure and function.* Biol Chem, 2008. **389**(4): p. 353-63.
7. Yun, M., et al., *Readers of histone modifications.* Cell Res, 2011. **21**(4): p. 564-78.
8. Cohen, I., et al., *Histone modifiers in cancer: friends or foes?* Genes Cancer, 2011. **2**(6): p. 631-47.
9. Lee, M.J., et al., *Histone deacetylase inhibitors in cancer therapy.* Curr Opin Oncol, 2008. **20**(6): p. 639-49.
10. Martens, J.H., et al., *PML-RARalpha/RXR Alters the Epigenetic Landscape in Acute Promyelocytic Leukemia.* Cancer Cell, 2010. **17**(2): p. 173-85.
11. Kouzarides, T., *Histone methylation in transcriptional control.* Curr Opin Genet Dev, 2002. **12**(2): p. 198-209.
12. Peterson, C.L. and M.A. Laniel, *Histones and histone modifications.* Current biology : CB, 2004. **14**(14): p. R546-51.
13. Martin, C. and Y. Zhang, *The diverse functions of histone lysine methylation.* Nat Rev Mol Cell Biol, 2005. **6**(11): p. 838-49.
14. Fog, C.K., K.T. Jensen, and A.H. Lund, *Chromatin-modifying proteins in cancer.* APMIS, 2007. **115**(10): p. 1060-89.
15. Hess, J.L., *MLL: a histone methyltransferase disrupted in leukemia.* Trends Mol Med, 2004. **10**(10): p. 500-7.
16. Min, J., et al., *Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase.* Cell, 2003. **112**(5): p. 711-23.
17. Greenberg, R.A., *Histone tails: Directing the chromatin response to DNA damage.* FEBS Lett, 2011. **585**(18): p. 2883-90.
18. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.
19. Eissenberg, J.C. and A. Shilatifard, *Histone H3 lysine 4 (H3K4) methylation in development and differentiation.* Dev Biol, 2010. **339**(2): p. 240-9.

20. Albert, M. and K. Helin, *Histone methyltransferases in cancer.* Semin Cell Dev Biol, 2010. **21**(2): p. 209-20.
21. Singer, M.S., et al., *Identification of high-copy disruptors of telomeric silencing in Saccharomyces cerevisiae.* Genetics, 1998. **150**(2): p. 613-32.
22. Feng, Q., et al., *Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain.* Current biology : CB, 2002. **12**(12): p. 1052-8.
23. Jenuwein, T., et al., *SET domain proteins modulate chromatin domains in eu- and heterochromatin.* Cell Mol Life Sci, 1998. **54**(1): p. 80-93.
24. Frederiks, F., et al., *Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states.* Nat Struct Mol Biol, 2008. **15**(6): p. 550-7.
25. Vidgren, J., L.A. Svensson, and A. Liljas, *Crystal structure of catechol O-methyltransferase.* Nature, 1994. **368**(6469): p. 354-8.
26. Mueller, D., et al., *A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification.* Blood, 2007. **110**(13): p. 4445-54.
27. Biswas, D., et al., *Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes.* Proc Natl Acad Sci U S A, 2011. **108**(38): p. 15751-6.
28. Okada, Y., et al., *hDOT1L links histone methylation to leukemogenesis.* Cell, 2005. **121**(2): p. 167-78.
29. Kim, S.K., et al., *Human histone H3K79 methyltransferase DOT1L protein [corrected] binds actively transcribing RNA polymerase II to regulate gene expression.* The Journal of biological chemistry, 2012. **287**(47): p. 39698-709.
30. Kouskouti, A. and I. Talianidis, *Histone modifications defining active genes persist after transcriptional and mitotic inactivation.* The EMBO journal, 2005. **24**(2): p. 347-57.
31. Vakoc, C.R., et al., *Profile of histone lysine methylation across transcribed mammalian chromatin.* Molecular and cellular biology, 2006. **26**(24): p. 9185-95.
32. Steger, D.J., et al., *DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells.* Molecular and cellular biology, 2008. **28**(8): p. 2825-39.
33. Mueller, D., et al., *Misguided transcriptional elongation causes mixed lineage leukemia.* PLoS Biol, 2009. **7**(11): p. e1000249.
34. Jo, S.Y., et al., *Requirement for Dot1l in murine postnatal hematopoiesis and leukemogenesis by MLL translocation.* Blood, 2011. **117**(18): p. 4759-68.
35. Jones, B., et al., *The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure.* PLoS Genet, 2008. **4**(9): p. e1000190.
36. Illi, B., et al., *Epigenetic histone modification and cardiovascular lineage programming in mouse embryonic stem cells exposed to laminar shear stress.* Circ Res, 2005. **96**(5): p. 501-8.

37. Nguyen, A.T., et al., *DOT1L regulates dystrophin expression and is critical for cardiac function.* Genes Dev, 2011. **25**(3): p. 263-74.
38. Feng, Y., et al., *Early mammalian erythropoiesis requires the Dot1L methyltransferase.* Blood, 2010. **116**(22): p. 4483-91.
39. Biondi, A., et al., *Biological and therapeutic aspects of infant leukemia.* Blood, 2000. **96**(1): p. 24-33.
40. Huret, J.L., P. Dessen, and A. Bernheim, *An atlas of chromosomes in hematological malignancies. Example: 11q23 and MLL partners.* Leukemia, 2001. **15**(6): p. 987-9.
41. Daser, A. and T.H. Rabbitts, *Extending the repertoire of the mixed-lineage leukemia gene MLL in leukemogenesis.* Genes Dev, 2004. **18**(9): p. 965-74.
42. Krivtsov, A.V. and S.A. Armstrong, *MLL translocations, histone modifications and leukaemia stem-cell development.* Nat Rev Cancer, 2007. **7**(11): p. 823-33.
43. Zhang, W., et al., *Aldosterone-induced Sgk1 relieves Dot1a-Af9-mediated transcriptional repression of epithelial Na+ channel alpha.* The Journal of clinical investigation, 2007. **117**(3): p. 773-83.
44. Yokoyama, A., et al., *A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription.* Cancer Cell, 2010. **17**(2): p. 198-212.
45. Martin, M.E., et al., *Dimerization of MLL fusion proteins immortalizes hematopoietic cells.* Cancer Cell, 2003. **4**(3): p. 197-207.
46. So, C.W., et al., *Dimerization contributes to oncogenic activation of MLL chimeras in acute leukemias.* Cancer Cell, 2003. **4**(2): p. 99-110.
47. Tkachuk, D.C., S. Kohler, and M.L. Cleary, *Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias.* Cell, 1992. **71**(4): p. 691-700.
48. Zeleznik-Le, N.J., A.M. Harden, and J.D. Rowley, *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, 1994. **91**(22): p. 10610-4.
49. Birke, M., et al., *The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation.* Nucleic Acids Res, 2002. **30**(4): p. 958-65.
50. Cierpicki, T., et al., *Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia.* Nat Struct Mol Biol, 2010. **17**(1): p. 62-8.
51. Wang, Z., et al., *Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression.* Cell, 2010. **141**(7): p. 1183-94.
52. Milne, T.A., et al., *MLL targets SET domain methyltransferase activity to Hox gene promoters.* Molecular cell, 2002. **10**(5): p. 1107-17.

53. Nakamura, T., et al., *ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation.* Molecular cell, 2002. **10**(5): p. 1119-28.
54. Hsieh, J.J., E.H. Cheng, and S.J. Korsmeyer, *Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression.* Cell, 2003. **115**(3): p. 293-303.
55. Hsieh, J.J., et al., *Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization.* Molecular and cellular biology, 2003. **23**(1): p. 186-94.
56. Yokoyama, A., et al., *The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis.* Cell, 2005. **123**(2): p. 207-18.
57. Grembecka, J., et al., *Menin-MLL inhibitors reverse oncogenic activity of MLL fusion proteins in leukemia.* Nat Chem Biol, 2012. **8**(3): p. 277-84.
58. Dou, Y., et al., *Regulation of MLL1 H3K4 methyltransferase activity by its core components.* Nat Struct Mol Biol, 2006. **13**(8): p. 713-9.
59. Ruthenburg, A.J., et al., *Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex.* Nat Struct Mol Biol, 2006. **13**(8): p. 704-12.
60. Ayton, P.M. and M.L. Cleary, *Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins.* Oncogene, 2001. **20**(40): p. 5695-707.
61. Cheung, N., et al., *Protein arginine-methyltransferase-dependent oncogenesis.* Nat Cell Biol, 2007. **9**(10): p. 1208-15.
62. Pal, S. and S. Sif, *Interplay between chromatin remodelers and protein arginine methyltransferases.* J Cell Physiol, 2007. **213**(2): p. 306-15.
63. Sobulo, O.M., et al., *MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3).* Proc Natl Acad Sci U S A, 1997. **94**(16): p. 8732-7.
64. Rowley, J.D., et al., *All patients with the T(11;16)(q23;p13.3) that involves MLL and CBP have treatment-related hematologic disorders.* Blood, 1997. **90**(2): p. 535-41.
65. Lavau, C., et al., *Chromatin-related properties of CBP fused to MLL generate a myelodysplastic-like syndrome that evolves into myeloid leukemia.* The EMBO journal, 2000. **19**(17): p. 4655-64.
66. Park, G., et al., *Characterization of the DOT1L network: implications of diverse roles for DOT1L.* Protein J, 2010. **29**(3): p. 213-23.
67. Zhang, W., et al., *Dot1a-AF9 complex mediates histone H3 Lys-79 hypermethylation and repression of ENaCalpha in an aldosterone-sensitive manner.* The Journal of biological chemistry, 2006. **281**(26): p. 18059-68.
68. Mohan, M., et al., *Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom).* Genes Dev, 2010. **24**(6): p. 574-89.
69. Bitoun, E., P.L. Oliver, and K.E. Davies, *The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional*

elongation and mediates coordinated chromatin remodeling. Hum Mol Genet, 2007. **16**(1): p. 92-106.
70. Lin, C., et al., *AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia.* Molecular cell, 2010. **37**(3): p. 429-37.
71. Milne, T.A., et al., *MLL associates specifically with a subset of transcriptionally active target genes.* Proc Natl Acad Sci U S A, 2005. **102**(41): p. 14765-70.
72. Yu, B.D., et al., *Altered Hox expression and segmental identity in Mll-mutant mice.* Nature, 1995. **378**(6556): p. 505-8.
73. Ernst, P., et al., *An Mll-dependent Hox program drives hematopoietic progenitor expansion.* Current biology : CB, 2004. **14**(22): p. 2063-9.
74. Ansari, K.I. and S.S. Mandal, *Mixed lineage leukemia: roles in gene expression, hormone signaling and mRNA processing.* FEBS J, 2010. **277**(8): p. 1790-804.
75. He, H., X. Hua, and J. Yan, *Epigenetic regulations in hematopoietic Hox code.* Oncogene, 2011. **30**(4): p. 379-88.
76. Krumlauf, R., *Hox genes in vertebrate development.* Cell, 1994. **78**(2): p. 191-201.
77. Jude, C.D., et al., *Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors.* Cell Stem Cell, 2007. **1**(3): p. 324-37.
78. McMahon, K.A., et al., *Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal.* Cell Stem Cell, 2007. **1**(3): p. 338-45.
79. Pineault, N., et al., *Differential expression of Hox, Meis1, and Pbx1 genes in primitive cells throughout murine hematopoietic ontogeny.* Exp Hematol, 2002. **30**(1): p. 49-57.
80. Thorsteinsdottir, U., et al., *Overexpression of the myeloid leukemia-associated Hoxa9 gene in bone marrow cells induces stem cell expansion.* Blood, 2002. **99**(1): p. 121-9.
81. Zeisig, B.B., et al., *Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization.* Molecular and cellular biology, 2004. **24**(2): p. 617-28.
82. Horton, S.J., et al., *Continuous MLL-ENL expression is necessary to establish a "Hox Code" and maintain immortalization of hematopoietic progenitor cells.* Cancer research, 2005. **65**(20): p. 9245-52.
83. Ayton, P.M. and M.L. Cleary, *Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9.* Genes Dev, 2003. **17**(18): p. 2298-307.
84. Milne, T.A., et al., *Leukemogenic MLL fusion proteins bind across a broad region of the Hox a9 locus, promoting transcription and multiple histone modifications.* Cancer Res, 2005. **65**(24): p. 11367-74.
85. Krivtsov, A.V., et al., *H3K79 methylation profiles define murine and human MLL-AF4 leukemias.* Cancer Cell, 2008. **14**(5): p. 355-68.

86. Guenther, M.G., et al., *Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia.* Genes Dev, 2008. **22**(24): p. 3403-8.
87. Bernt, K.M., et al., *MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L.* Cancer Cell, 2011. **20**(1): p. 66-78.
88. Daigle, S.R., et al., *Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor.* Cancer Cell, 2011. **20**(1): p. 53-65.
89. Chang, M.J., et al., *Histone H3 lysine 79 methyltransferase Dot1 is required for immortalization by MLL oncogenes.* Cancer research, 2010. **70**(24): p. 10234-42.
90. Nguyen, A.T., et al., *DOT1L, the H3K79 methyltransferase, is required for MLL-AF9-mediated leukemogenesis.* Blood, 2011. **117**(25): p. 6912-22.
91. Deshpande, A.J., et al., *Leukemic transformation by the MLL-AF6 fusion oncogene requires the H3K79 methyltransferase Dot1l.* Blood, 2013. **121**(13): p. 2533-41.
92. Zeisig, D.T., et al., *The eleven-nineteen-leukemia protein ENL connects nuclear MLL fusion partners with chromatin.* Oncogene, 2005. **24**(35): p. 5525-32.
93. Nguyen, A.T. and Y. Zhang, *The diverse functions of Dot1 and H3K79 methylation.* Genes Dev, 2011. **25**(13): p. 1345-58.
94. Arkin, M.R. and J.A. Wells, *Small-molecule inhibitors of protein-protein interactions: progressing towards the dream.* Nat Rev Drug Discov, 2004. **3**(4): p. 301-17.

Chapter 2

Biological Characterization of MLL fusion proteins and DOT1L Protein-Protein Interactions

2.1 Abstract

The MLL fusion proteins, AF9 and ENL, activate target genes in part *via* recruitment of the histone methyltransferase DOT1L (Disruptor Of Telomeric silencing 1-Like). We have characterized the interaction between DOT1L and the MLL-fusion proteins, AF9/ENL. The AF9/ENL binding site in human DOT1L was mapped and the interaction site was identified as a 10 amino acid region ($DOT1L_{865-874}$), a region that is highly conserved in DOT1L from a variety of species. Alanine scanning mutagenesis analysis shows that four conserved hydrophobic residues from the identified binding motif are essential for the interactions with AF9/ENL. Phosphorylation of DOT1L serine 868 in the AF9/ENL interaction region diminished the protein-protein interaction significantly and binding studies demonstrated that the entire intact C-terminal domain of AF9/ENL is required for optimal interaction with DOT1L. Functional studies show that the AF9/ENL interacting site that was mapped is essential for immortalization by MLL-AF9, indicating that DOT1L interaction with MLL-AF9 and its recruitment are required for transformation by MLL-AF9. These results strongly suggest that

disruption of the interaction between DOT1L and AF9/ENL is a promising therapeutic strategy for MLL-fusion protein-associated leukemia with potentially fewer adverse effects than enzymatic inhibition of DOT1L.

2.2 Background

2.2.1 AF9 and ENL proteins were involved in multiple protein-protein interactions

AF9 and ENL, the most common fusion partners in MLL rearrangement leukemia [1] belong to the YEATS family of proteins. The AF9/ENL protein consists of an N-terminal YEATS domain, a poly-serine and -proline region, and a coiled-coil hydrophobic domain at the C-terminus [2] (Figure 2.1). The YEATS domain is named after the first proteins, Yaf9, ENL, AF9, Taf14, Sas5, that were found to contain this module [3]. The primary amino acid sequence of the YEATS domain is evolutionarily conserved from yeast to humans and is found in proteins belonging to a variety of chromatin-modifying complexes and transcription factors [4]. The C-terminal domains of AF9 and ENL share high homology [5] and are reported to be involved in interactions with many binding partners with roles in transcription and chromatin modification. In addition to the sequence similarity, the C-terminal domain of AF9 or ENL shares similar protein-protein interactions. Both AF9 and ENL bind to the mouse

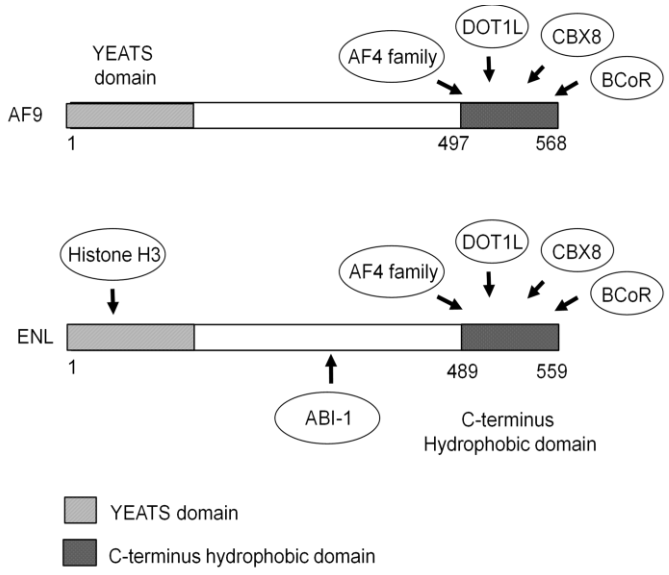


Figure 2.1 The domain structure of AF9/ENL protein. AF9/ENL has an N-terminal YEATS domain, and a C-terminal hydrophobic domain. The proteins interacting with AF9/ENL are indicated**.**

or human Polycomb proteins [6] [7, 8] [9-11], AF4 proteins (Srinivasan, Nesbit et al. 2004,) BCL-6 co-repressor (BCoR) [12] and DOT1L [13, 14]. ENL also has a distinct set of protein-protein interactions and it interacts with another MLL fusion partner, ABI-1. However, the ABI-1 interaction region of ENL is not in the C-terminal domain which is important for MLL transformation. [15]. ENL specifically associates with histones H3 and H1 whereas it shows no affinity for H4, H2A. The histone binding capacity of ENL is crucially dependent on the YEATS domain as N-terminal truncation mutants are unable to interact with histones. However, the deletion of C-terminal domain has no influence on H1/H3 affinity [7]. ENL was also found to be a subunit of 2 novel SWI/SNF complexes, EBAFa and EBAFb, ENL-containing BAFa and BAFb complexes [16].

The C-terminal hydrophobic domain (also called the ANC homology domain) of about 90 amino acids of the AF9 and ENL proteins is retained in MLL fusion proteins and has been shown to be crucial for recruitment of DOT1L in MLL. Deletion or mutations within this region, which can disrupt the PPIs between AF9-DOT1L and ENL-DOT1L are also able to abolish the MLL transformation [11, 14]. Additionally, AF9 and ENL are able to recruit and activate p-TEFb by recruitment of AF4 family members through the C-terminus domain. Disruption of AF9-AF4 interaction results in necrotic cell death in several cell lines harboring MLL translocations [17].

2.2.2 Protein-Protein interactions between DOT1L and MLL fusion proteins

DOT1L is a protein with 1537 amino acids. The N-terminal (1-416) of the protein is the catalytic domain [18] (Figure 2.2). The C-terminal region of DOT1L, outside the catalytic domain is reported to be involved in multiple protein-protein interactions. The reported DOT1L interaction partners include AF9, ENL, AF10, AF17 and CTD of RNA Pol II [11, 13, 19-22]. The AF9/ENL-binding domain within the DOT1L C-terminus has been identified as being between amino acids 826 and 1095, a region without any recognizable motif [11, 13].

The physical interaction of DOT1L with AF9/ENL has been demonstrated by co-immunoprecipitation and *in vivo* experiments [11, 23] but a detailed characterization of this interaction is currently unavailable and the exact DOT1L fragment involved in the interaction with AF9/ENL has not been identified. We have characterized the DOT1L-AF9/ENL interactions using a collection of biochemical and biophysical methods. Our work, described below, establishes the molecular basis of the DOT1L-AF9/ENL

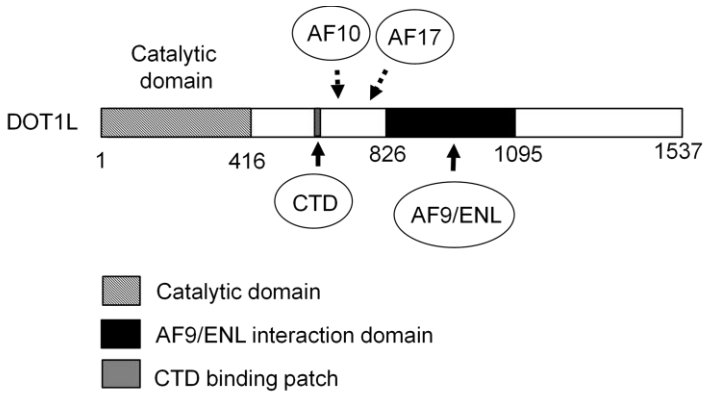


Figure 2.2 DOT1L interacting proteins. The DOT1L interacting proteins and their interaction site in DOT1L are indicated. AF9/ENL interacts with DOT1L in the C-terminal region (826-1095) lacking any apparent homology motif. The AF10 and AF17 interaction sites are indicated by dashed line arrows since their interaction sites in DOT1L are not precisely identified. The phosphorylated CTD of RNA Pol II interacts with DOT1L (618-627), which is a conserved motif call CTD binding patch.

interactions and provides the necessary foundation for development of small molecule inhibitors targeting this interaction in MLL translocated leukemias.

2.3 Experimental procedures

2.3.1 Plasmids and Cloning

The full-length hDOT1L was generously provided by Dr. Yi Zhang from the University of North Carolina at Chapel Hill. All different DOT1L constructs, tested in this study (Table 2.1), were made using full length hDOT1L as a template. The plasmids for AF9, ENL and AF4, were made using MLL-AF9, MLL-ENL and MLL-AF4 fusion proteins as templates. All these constructs for protein expression were cloned by ligation independent cloning (LIC) methods as described before [24]. Different DOT1L constructs (Table 2.1) and AF4 proteins (749-775) were cloned into pMocr-LIC vector obtained from the UM CSB HTP facility. ENL (489-559) was cloned into pMSCG9-LIC vector and AF9 (497-568) was cloned into pGB1-LIC vector. MSCV based HA tagged wild type mouse DOT1L, methyltransferase mutant DOT1L (RCR), and neomycin vector constructs have been reported before [25]. MSCV-HA-mDOT1L 10aa∆ construct where 10 residues identified as AF9/ENL interaction site (863-872 aa) were deleted by using the QuickChange site-directed mutagenesis kit (Agilent Technologies). These residues in mouse are conserved and correspond to the identified human DOT1L interaction site, 865-874 aa (Figure 2.5B). ENL L550E mutant was also created by site-directed mutagenesis (QuickChange, Agilent Technologies). To construct the pCMV-Myc CxxC-AF9 plasmid, a fragment of an MLL-AF9 fusion protein was cut from an MLL-AF9 fusion protein expression vector using MfeI and XhoI and inserted into the pCMV-Myc vector (Clontech) following digestion with EcoR1 and XhoI. A dual nuclear localization signal was inserted into the SfiI

site downstream of the Myc tag. The sequence of MLL includes amino acids 1116 through 1422 and is followed by AF9 sequence including amino acids 479 through 568. The positive clones for all desired constructs were confirmed by DNA sequencing (University of Michigan DNA sequencing core).

2.3.2 Protein expression and purification

All recombinant proteins were expressed in *Escherichia coli* strain BL21 (DE3) (Invitrogen). The medium for bacterial growth was Luria Broth (for all binding studies) and M9 minimal medium supplemented with $^{15}NH_4Cl$ (Cambridge Isotope Laboratories) (for HSQC NMR studies). All the proteins were induced with 200 µM IPTG at 20 °C except GB1-AF9 (489-497) was induced at 16 °C. The cells were harvested after 20h and resuspended in cold lysis buffer (50 mM Tris HCl, pH 7.5, 150 mM NaCl, 0.01 % β-mercatoethanol) and purified by affinity chromatography employing Ni-agarose (Qiagen). Mocr- DOT1L and -AF4 proteins were further purified by anion exchange chromatography in 25 mM Tris HCl pH 7.5, with NaCl gradient ranging from 25 mM to 1 M, and 3 mM DTT, while MBP-ENL (489-559), MBP-ENL (489-559) L550E, and GB1-AF9 (497-568) proteins were further purified with size exclusion chromatography in 50 mM Tris HCl, pH 7.5, 150 mM NaCl, 3 mM DTT. The MBP and Mocr-tags from ENL (489-559) and AF9 (497-568), respectively, were removed by proteolysis using tobacco etch virus (TEV) protease, followed by nickel-NTA column and size exclusion chromatography using the same buffer as above to obtain purified cleaved proteins. All purified recombinant proteins were stored at -80 °C for further experiments.

2.3.3 Peptide synthesis

Peptides were synthesized using standard F-moc solid phase peptide synthesis techniques on an ABI 433A automated peptide synthesizer. NovaPEG Rink amide resin (EMD) was used to prepare all C-terminal amide-capped peptides. Standard side chain protecting groups were used for all amino acids. All the peptides were acetylated on the N-terminus. Peptides containing an N-terminal biotin group or fluorescein were coupled to lysine and two β-alanine residues as a spacer. All crude peptides were purified by semi-preparative reverse-phase high-performance liquid chromatography (RP-HPLC) and their sequence and purity were verified by electrospray ionization mass spectrometry (ESI-MS) and analytical RP-HPLC.

2.3.4 Surface Plasmon Resonance (SPR) binding studies

All SPR based experiments were performed on a BIAcore 2000 (BIAcore, GE Healthcare) instrument. Different tested DOT1L recombinant proteins were immobilized on a CM5 sensor chip by standard EDC/NHS coupling chemistry followed by ethanolamine deactivation of the surfaces, with immobilization level of 2000 to 3000 RU. For immobilization of full length Flag-DOT1L protein, anti-Flag antibody (Sigma, Anti-FLAG M2) was immobilized on CM5 chip by the same amine coupling chemistry. HEK293 whole cells lysate with overexpressed Flag-DOT1L at 2.4 µg/µl total protein concentration was injected over the Flag antibody surface, with immobilized level of 1000RUfollowed by injecting of AF9 and ENL in different concentrations. For determination of their binding affinity AF9 and ENL were tested in concentration from 0.01 to 3 µM in HBS-P buffer (10

mM HEPES pH 7.4, 150 mM NaCl, 0.005% v/v P20, GE Healthcare). Injection of 50 mM NaOH was used to regenerate the proteins bound to the surfaces. The Fc1 surface was used as a control surface and was treated in the same manner as the active surfaces but in the absence of immobilized protein. Binding parameters, k_{on}, k_{off} and K_d were calculated by simultaneous non-linear regression using BIAEvaluation software.

Solution competitive SPR based assay was performed to determine the IC_{50} values of DOT1L peptides. The tested DOT1L wild type and alanine mutated peptides were pre-incubated with AF9 or ENL proteins (500 nM) for at least 30 minutes and then the reaction mixture was injected over the surface of Mocr-DOT1L (826-1095) immobilized CM5 chip. Response units (RU) were measured at 15 seconds in the dissociation phase and the specific binding was calculated by subtracting the control surface (Fc1) signal from the surfaces with immobilized Mocr-DOT1L. IC_{50} values were determined by non-linear least squares analysis using Graph Pad Prism 5.0 software.

2.3.5 Fluorescence polarization (FP) assay

FP experiments were performed in 96-well, black, round-bottom plates (Corning) using the plate reader (Biotek H1 hybrid). Fluorescein tagged DOT1L 10mer peptide (Flu-DOT1L) was used as a fluorescent probe in the FP based binding assays. The *Kd* values of GB1-AF9, MBP-ENL wild type and L550E mutated proteins, were determined using a fixed concentration of the tracer (10 nM Flu-DOT1L) and different concentrations of the tested proteins, in a final

volume of 125 µl and assay buffer (100 mM Na_2HPO_4, pH 7.4, 150 mM NaCl, 0.01 % Triton X-100 and 4 % DMSO). The plate was mixed on a shaker and incubated at room temperature to reach equilibrium. The polarization values in milli-polarization units (mP) were measured at an excitation wavelength at 485 nm and an emission wavelength at 530 nm. Polarization data were analyzed using GraphPad Prism 5.0 by non-linear fitting with a one-site binding model.

AF4 (749-775) protein and several peptides (AF4 14mer, un-labeled and biotin-labeled DOT1L 10mer peptide) were tested for their ability to displace the Flu-DOT1L from GB1-AF9 protein. The dose-dependent binding experiments were carried out in 96-well plates with serial dilutions of tested protein or peptides, GB1-AF9 protein (500 nM) and Flu-DOT1L (10 nM) in the same assay buffer and final volume of 125 µL. The polarization values were measured after 3h incubation. Negative controls containing GB1-AF9 protein and probe (equivalent to 0 % inhibition), and positive controls containing only free Flu-DOT1L probe (equivalent to 100 % inhibition), were included on each assay plate. IC_{50} values were determined by nonlinear regression fitting of the competition curves (GraphPad Prism 5.0).

2.3.6 Isothermal titration calorimetry

ITC was carried out using a Nano-ITC Micro Calorimeter (TA Instruments) at 20 °C. MBP-ENL (489-559) was dialyzed extensively against 50 mM Na_2HPO_4, (pH 7.5) and 100 mM NaCl, and 80 µM was used for titration studies of DOT1L 16-mer and 10-mer peptides, and 110 µM for DOT1L 7-mer titration. Tested

peptides dissolved in the same buffer at 400 µM (DOT1L 16-mer and 10-mer) or 500 µM (DOT1L 7-mer) were tested by injecting 2 µl aliquots into the protein sample, at time intervals of 30 seconds, to ensure that the titration peak returned to the baseline. The ITC data were analyzed by NanoAnalyze software package using a one site-binding model. The ITC of MBP-AF9 with DOT1L 16-mer peptides was performed on VP-ITC system (MicroCal) at 25 °C. Tested peptides were diluted in the same buffer at 300 µM in the presence of 5 % DMSO. The peptide was injected in 10 µL aliquots into the protein sample, at time intervals of 2 minute. The ITC data was analyzed by Origin 7.0 (OriginLab) using a one site-binding model.

2.3.7 NMR spectroscopy

^{15}N-labeled ENL protein sample was prepared in a 50 mM Na_2HPO_4, pH 7.5, 50 mM NaCl in 7 % D_2O. The binding of DOT1L protein and peptides has been characterized by recording ^{1}H, ^{15}N -HSQC spectra of uniformly ^{15}N-labeled ENL in the absence and presence of tested protein and peptides in 1:1 molar ratio. The presence of tested peptides in solution was confirmed by ^{1}H-1D NMR. All spectra were acquired at 30 ^{o}C on a Bruker 600 MHz NMR spectrometer equipped with a cryogenic probe, processed using Bruker TopSpin and rNMR [26], and were analyzed with Sparky [27].

2.3.8 Circular dichroism

DOT1L peptides were dissolved in phosphate buffer (50 mM Na_2HPO_4, pH 7.4, 100 mM NaCl). ENL and AF9 protein were dialyzed against the above

phosphate buffer overnight. CD measurements were performed at room temperature using a Jasco J-715 and a quartz flow cell with a 1 mm path length. The spectra were averaged over 10 scans and the baseline (buffer scan) was subtracted from each spectrum. The percentage of α helicity was calculated by K2D2 online tool.

2.3.9 Generating the homology model of the complex ENL-DOT1L

The NMR structure of 2LM0, in which the C-terminal hydrophobic domain of AF9 is in complex with the elongation factor AF4, was employed for the homology modeling of the complex between ENL and DOT1L. The reported NMR structure contains ten conformers and the fourth was randomly selected as the template because the alignment of these conformers showed that they are virtually identical. The sequence of ENL (489 – 559) was aligned with this minimized AF9 – AF4 complex using the program of MOE-align. The homology model was built using the Homology Model module of MOE 2010.10. During this process, the peptide of LMVKITL was kept as environment for induced fit. In the top-scored model produce by MOE, the peptide of LMVKITL was mutated to LPISIPL and then the obtained homology model was minimized.

2.3.10 Cell Culture, Transfections, and Immunoprecipitation

HEK293 cells were plated in 100 mm culture dishes and cultured in Dulbecco's modified Eagle's medium (DMEM) medium supplemented with 10 % FBS and antibiotics. Co-transfection with FLAG-DOT1L and Myc-CxxC-AF9 was performed using Lipofectamine 2000 (Invitrogen). Forty eight hours post-

transfection, cells were collected and lysed using BC-300 lysis buffer: 20 mM Tris-HCl pH 8.0, 300 mM KCl, 1 mM EDTA, 10 % glycerol, 0.1 % NP-40 and protease inhibitor cocktail (Roche). The lysates was pre-cleared for 2h in Mouse IgG Agarose (Sigma-Aldrich). The pre-cleared lysates were incubated with different concentrations of DOT1L 10mer peptide at 4 °C overnight. The next day, the cell lysates were immunoprecipitated with Anti-FLAG M2 magnetic beads (Sigma-Aldrich) at 4 °C for 2h. After incubation, the beads were washed extensively, boiled in SDS loading buffer and analyzed by western blotting using mouse monoclonal anti-FLAG M2 (Sigma-Aldrich) and goat monoclonal anti-Myc tag (Abcam) antibody.

Primers for and Quantitative PCR (qPCR):

5S rRNA: TCTACGGCCATACCACCCTGA and

GCCTACAGCACCCGGTATTCC;

HA-Dot1l: GCCACCATGTACCCCTACGACGTG and

GATTTCCTCGCAGACCCACCGGAT

2.3.11 Pull-down assay

HEK293 cells transfected with Myc-CxxC-AF9 were lysed in BC-300 lysis buffer in the same way as the Co-IP experiment. The supernatant was pre-cleared for 2h in streptavidin agarose beads (Thermal Scientific). The pre-cleared lysates were incubated with different concentrations of biotin-labeled DOT1L 10mer peptide at 4 °C overnight. The next day, the cell lysate was incubated with streptavidin agarose at 4 °C for 2h. After extensive washing the beads with the

BC-300 lysis buffer without NP-40, the pull down samples were applied to SDS-PAGE electrophoresis and pull-down Myc-CxxC-AF9 protein was probed with goat monoclonal anti-Myc tag antibody (Abcam).

2.3.12 Retroviral Transductions and CFU Assay

Retroviral production and transduction of bone marrow progenitor cells were carried out as described [25]. Briefly, retroviruses were generated by transfecting MSCV-HA wild type mouse DOT1L, methyltransferase mutant mDOT1L (RCR), mDOT1L deletion (10aaΔ; deletion of the residues 863-872 aa) constructs, and neomycin vector control into Plat-E cell line with Fugene 6 (Roche). Fresh viral supernatants were used for transducing MLL-AF9 transformed cells described [25]. 150,000 cells were used for transduction per viral supernatant from 10 cm dish. Cells were then plated on Methocult media (Stem Cell Technologies, M3234) with 1% penicillin/ streptomycin (Invitrogen), 10 ng/ml IL3 (R & D Systems), 5 nM 4-OHT (Sigma) or 100% ethanol, and 1 mg/ml G418 (Invitrogen). Colonies were scored 5-7 days after plating for two rounds. In the final round, colonies were stained with 0.1 % p-iodonitrotetrazolium violet (Sigma) for visualization.

2.3.13 RNA extraction, cDNA generation, and Protein extraction

RNA was extracted from cells using TRIzol Reagent (Invitrogen) and converted to cDNA using SuperScript II (Invitrogen) according to manufacturer's instructions. Whole cell lysate samples were prepared by directly resuspending cells in Tris-Glycine SDS sample buffer (Novex) and sonicating for 15 minutes

(Bioruptor, Diagenode). Primers for quantitative PCR (qPCR) are provided in the supplementary information

2.3.14 In-gel digestion with trypsin

Anti-Flag affinity purified DOT1L protein was separated on a SDS-PAGE and visualized using Simply Blue Coomassie stain. Protein band corresponding to DOT1L was excised and destained with 30 % methanol for 4 h. Cysteines were reduced with 150 µL of 10 mM DTT followed by alkylation with 65 mM 2-chloroacetamide (in 0.1 M ammonium bicarbonate, pH 8.0) at RT for 30 min, each. Gel slices were crushed, dried using a vacufuge and re-swollen in 30 µL of ammonium bicarbonate buffer containing 500 ng of sequencing grade, modified trypsin (Promega). After 45 min incubation on ice, another 40 µL of ammonium bicarbonate buffer was added and digestion was carried out at 37 °C, overnight. Additional 250 ng of trypsin was added 2 h prior to the extraction of peptides. Peptides were extracted once each with 150 µL of 60 % acetonitrile/0.1 % TFA and acetonitrile/0.1 % TFA. All extracts were pooled and dried using a vacufuge. Sample was re-dissolved in 30 µL of HPLC loading solvent (2 % acetonitrile/0.1%TFA) and analyzed immediately.

2.3.15 Protein identification by LC-tandem MS

Ten µl of the digest was separated on a reverse phase column (Aquasil C18, 15 m tip x 75 m i.d. x 15 cm, Picofrit column, New Objectives, Woburn MS) using acetonitrile/1 % acetic acid gradient system (5-60 % acetonitrile over 40 min followed by 95 % acetonitrile wash for 5 min) at a flow rate of ~300 nl/min.

Eluted peptides were directly introduced into an ion-trap mass spectrometer (LTQ Orbitrap XL, ThermoFisher) equipped with a nano-spray source. The mass spectrometer was operated in data-dependent acquisition mode to acquire a high-resolution full MS scan in Orbitrap analyzer (FWHM 30,000 at 400 m/z) followed by collision induced dissociation MS/MS spectrum on the top 9 ions in the linear ion-trap (relative collision energy ~35 %). Dynamic exclusion was set to collect 2 MS/MS spectra on each ion and exclude it for further 2 min. Raw files were converted to mzXML format and searched against SwissProt human protein database (release 2013) appended with reverse database and common contaminants using XTandem with *k-score* plug-in, an open-source search engine developed by the Global Proteome Machine. Search parameters included a precursor peptide mass tolerance of 50 ppm and fragment mass tolerance of 0.5 Da. Oxidation of methionine (+15.9949 Da) and carbamidomethylation of cysteines (+57.0214 Da), phosphorylation on S, T and Y (+79.966Da) were considered as variable modifications. Search was restricted to tryptic peptides with two missed cleavages. Tandem outputs were subjected to TransProteomic Pipeline (TPP) – A software suite containing PeptideProphet [28]and ProteinProphet [29], analysis using default parameters of each program. All proteins with a ProteinProphet probability of >0.9 (fdr <2 %) were considered for further analysis. Phosphopeptide matches were manually verified where possible.

2.4 Results

2.4.1 Mapping the AF9/ENL binding site in DOT1L

The interaction of AF9 or ENL with DOT1L has been studied previously by co-immunoprecipitation and yeast 2-hybrid studies [10, 14, 23, 30, 31]. To understand the molecular basis of DOT1L interaction with MLL-fusion proteins, we analyzed these protein-protein interactions (PPIs) using a surface plasma resonance (SPR) assay to quantify the binding affinity between full length DOT1L protein and MLL-fusion proteins. The flag-tagged full length DOT1L protein was transiently transfected in HEK 293 cells. After 48 hours transfection the DOT1L protein was captured from the HEK 293 whole cell lysate on an anti-flag antibody-coated biosensor chip (Figure 2.3).

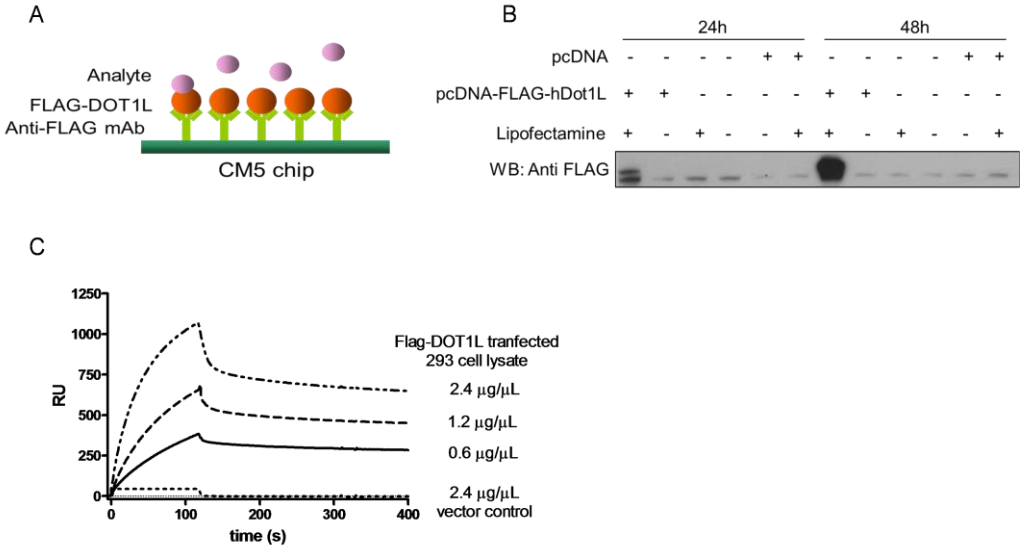


Figure 2.3 Immobilization of Flag-DOT1L for interaction studies. (A) Design of the binding experiments. (B) Western Blot of the expression level of transfected Flag-DOT1L protein. (C) The specificity of immobilizing Flag-DOT1L protein.

The recombinant C-terminal domain from human AF9 (residues 497-568) and the corresponding segment from the ENL protein (residues 489-559) (Figure 2.1) were cloned, expressed and purified for the biochemical binding studies. We determined that AF9 and ENL bind to the immobilized full length DOT1L with dissociation constants (K_d) of 33 nM and 206 nM respectively, agreeing well with a 1:1 interaction model (Figure 2.4A).

These studies confirmed that DOT1L directly interacts with AF9 and ENL and for the first time the binding affinity of DOT1L to AF9 and ENL were

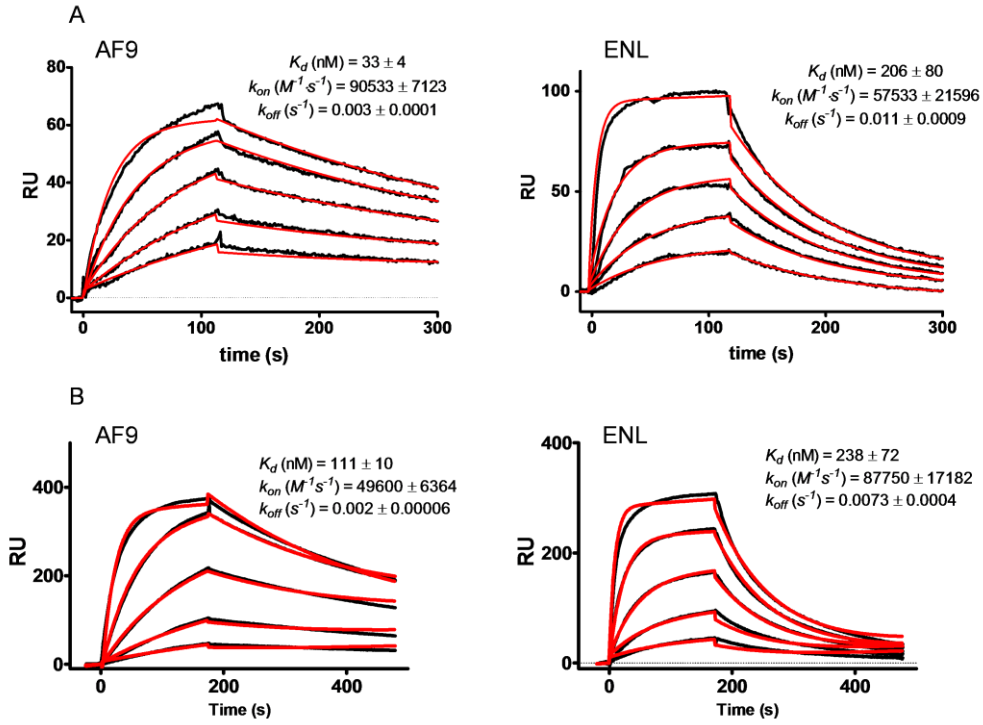


Figure 2.4 Binding of AF9/ENL protein to DOT1L. (A) and (B) Sensorgrams representing the concentration-dependent binding of the C-terminal domain of AF9 and ENL tested in a concentration range from 0.01 to 3 μM, with (A) full length Flag-DOT1L and (B) Mocr-DOT1L (826 - 1095), both immobilized on a CM5 sensor chip. The k_{on}, k_{off} and K_D were calculated by simultaneous non-linear regression using a 1:1 binding model and BIAevaluation 3.1 software. The experimental data are shown in black while the global fit analyses are in red.

determined quantitatively.

The C-terminal unstructured region of DOT1L has been reported to be involved in the interaction with AF9 and ENL [10, 23] (Figure 2.2). Consequently, we cloned and expressed the 826-1095 fragment from DOT1L and studied the interaction of this recombinant protein with AF9/ENL. SPR analysis showed that this segment of the DOT1L protein has K_d values of 111 nM and 238 nM to AF9 and ENL protein respectively (Figure 2.4B), similar to those of the full length DOT1L (Figure 2.4A), confirming that this region in the DOT1L protein is essential for interactions with AF9/ENL proteins.

To further map the region of DOT1L required for the AF9/ENL interaction, a series of truncated constructs of DOT1L (826-1095), devised according to predicted stability, were designed, cloned, expressed and tested for their binding to AF9 and ENL (Table 2.1) .

Table 2.1 Binding affinities of AF9 and ENL proteins to full length DOT1L and different constructs of DOT1L immobilized on a CM5 sensor chip

Immobilized Proteins	AF9 (497-568) K_d ± SD [nM]	ENL (489-559) K_d ± SD [nM]
DOT1L (1-1537)*	33 ± 4	206 ± 80
DOT1L (826-1095)	111 ± 10	238 ± 72
DOT1L (826-958)	92 ± 8	122 ± 16
DOT1L (826-937)	78 ± 15	81 ± 6
DOT1L (854-1095)	135 ± 4	148 ± 14
DOT1L (925-1095)	1,825 ± 276	6,220 ± 113
DOT1L (854-925)	42 ± 2	90 ± 12

The SPR studies show that truncation of the DOT1L protein on the N-terminal site up to residue 854 and C-terminal deletions of up to residue 937 failed to affect the interactions with AF9 and ENL and the measured binding affinities were in a similar range. However, further deletion of the N-terminal region and testing of the truncated DOT1L construct (925 to 1095 residues) showed 30-fold and 50-fold decreases in binding affinity for ENL and AF9 respectively, clearly demonstrating the importance of the sequence 854 to 925 for the binding to MLL fusion proteins. Based on this and the fact that the DOT1L (854-1095) fragment binds with binding affinity similar to that of the full length DOT1L protein, the DOT1L construct (854-925) was cloned and tested. Consistent with predictions, this shortest DOT1L fragment showed similar binding affinity to AF9 and ENL to that of the full length DOT1L with K_d values of 42 nM and 90 nM, respectively (Table 2.1). These results confirm that DOT1L (854-925) contains a domain that interacts specifically with MLL-fusion proteins, AF9 and ENL.

2.4.2 Identification of a conserved peptide motif in DOT1L essential for binding AF9/ENL

It is known that the C-terminal hydrophobic AF9/ENL domains in MLL-fusion proteins retain the ability to form independently higher order complexes with AF4/p-TEFb and with DOT1L, demonstrating that the associations of AF9/ENL with AF4 and DOT1L are mutually exclusive [14, 23]. The AF9-binding domain of AF4 is known and is conserved among the AF4 homologs [17]. Based on these findings we predicted that the AF9/ENL interacting site in DOT1L might

share some similarity with the AF4 protein. A small 16-residue region in DOT1L (865-880) was identified by alignment of the shortest fragment of DOT1L (854-925) involved in AF9/ENL interactions with the AF9-binding domain of AF4 (761-774). This region shares sequence homology with AF4, and includes four conserved and four similar residues (Figure 2.5A).

Significantly, alignment of DOT1L from different species demonstrated that this fragment is highly conserved (Figure 2.5B). Based on this result we synthesized two peptides, DOT1L10mer (865-874) and DOT1L16mer (865-880) and, using several binding assays such as SPR, fluorescence polarization (FP), and isothermal titration calorimetry (ITC) examined whether the conserved peptide motif is sufficient for interaction with AF9 and ENL. Applying a competitive SPR assay, we tested the ability of 16mer and 10mer synthetic peptides to inhibit the PPI between DOT1L and AF9 or ENL.

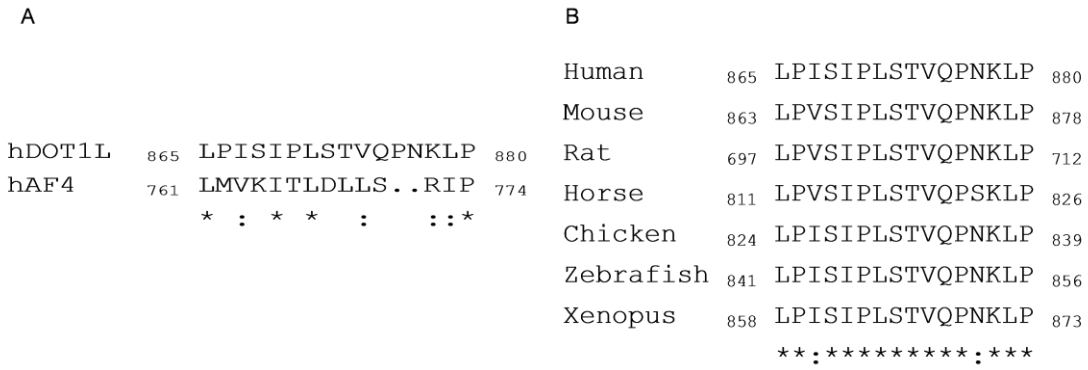


Figure 2.5 (A). Alignment of human DOT1L (865-880) with human AF4 (761-774). The conserved residues are marked with an asterisk, and the similar amino acids with dots. B. Alignment of the identified AF9/ENL binding site in DOT1L protein from different species.

The results obtained demonstrate that both 16mer and 10mer peptides can block the PPI between DOT1L/AF9 and DOT1L/ENL with similar IC_{50} values,

0.49 µM and 0.32 µM respectively, against AF9, and 1.34 µM and 1.56 µM respectively, against ENL (Figure 2.6A). We tested the fluorescent labeled

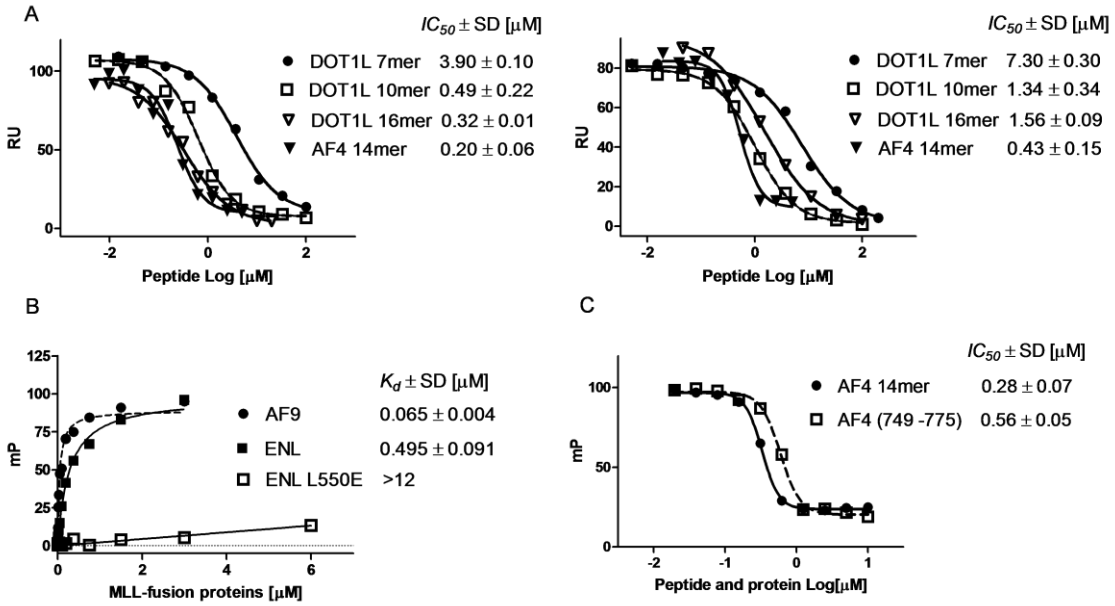


Figure 2.6 Binding studies of DOT1L peptides with AF9/ENL proteins. (A) SPR competitive binding curves of DOT1L 16mer, 10mer, 7mer and AF4 14mer peptides against AF9 (497-568) and ENL (489-559) proteins using different tested concentrations over CM5 chip with immobilized DOT1L (826-1095). (B) Binding affinities of fluorescent-labeled DOT1L 10-mer peptide against AF9, ENL and mutated ENL L550E. (C) Fluorescence polarization (FP) competitive binding curves of AF4 (749-775) recombinant protein and AF4 14mer peptide using fluorescein-labeled DOT1L 10mer peptide.

DOT1L 10mer peptide and found that it binds to the MLL fusion proteins, AF9 and ENL, with K_d values of 0.065 µM and 0.495 µM respectively (Figure. 2.6B).

To further confirm these binding results, an ITC assay using MBP-tagged ENL protein (Figure 2.7) was employed. Consistent with the FP-based results, both peptides, DOT1L 16mer and 10mer, bind to the ENL protein with similar binding affinities and K_d values of 0.9 µM and 1.1 µM, respectively in 1:1 stoichiometry (Figure 2.7A and B). DOT1L 16mer binds to MBP-AF9 with a Kd of 0.38 µM in 1:1 stoichiometry (Figure 2.7D). The binding results obtained for

DOT1L 16mer and 10mer peptides are consistent with our direct binding studies of the PPI (Table 2.1) and provide strong evidence that peptide motif identified in DOT1L is both necessary and sufficient for interaction with AF9 and ENL.

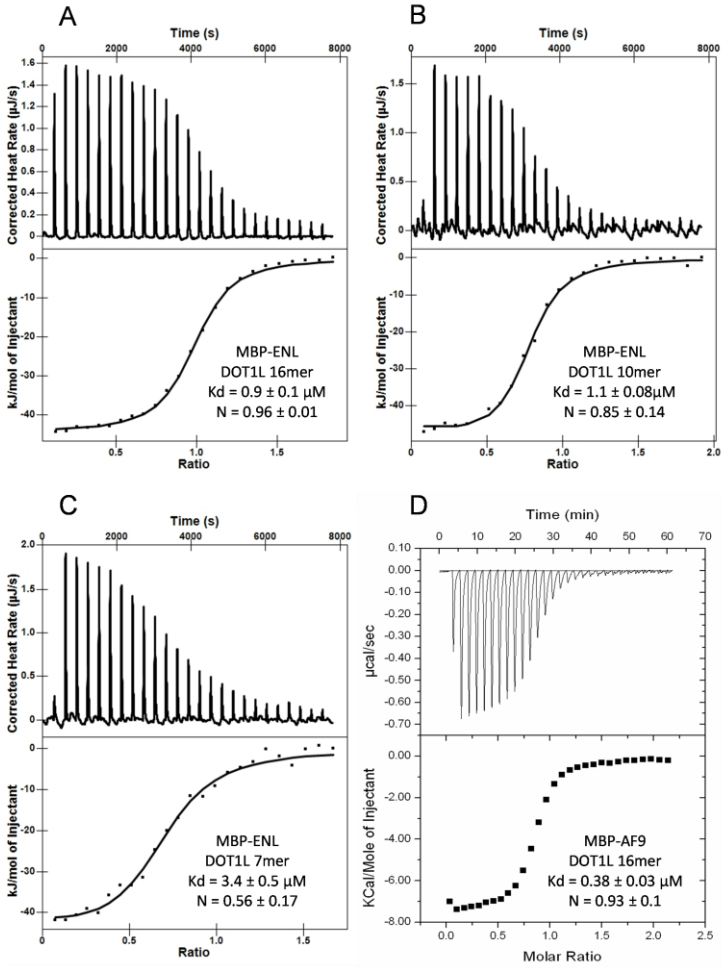


Figure 2.7 Isothermal titration calorimetry (ITC) of MBP-ENL (489-559) (80 µM) with a solution of DOT1L 16-mer (400 µM) (A) and 10-mer (400 µM) (B), and for DOT1L 7-mer peptide (500 µM) (C) and MBP-AF9 (497-568) (30 µM) with DOT1L 16-mer peptide (300 µM) (D).110 µM MBP-ENL was used. For all titrations, the raw data are shown in the upper panel, and the integrated calorimetry data are shown in the lower panel.

Since the identified DOT1L peptide shows sequence homology with the reported AF4 peptide, and the associations of AF9 and ENL with AF4 and

DOT1L through their conserved C-terminal domain have been reported to be mutually exclusive [14, 23], competition experiments were carried out between DOT1L and AF4 for binding to MLL-fusion proteins. For this purpose we expressed and purified AF4 recombinant protein (749-775 residues) and synthesized the AF4 14mer peptide (761-774) [17]. As expected, we found that the recombinant AF4 protein and AF4 14mer peptide efficiently competed away the binding of DOT1L 10mer fluorescein-labeled peptide to the C-terminal domain of AF9 protein with similar IC_{50} values of 0.56 and 0.28 μM (Figure. 2.6C). The SPR solution competitive binding assay confirmed these results and showed that the AF4 14mer peptide is also able to inhibit the PPI between AF9-DOT1L and ENL-DOT1L, with potency similar to that of DOT1L 16mer and 10mer peptides, showing IC_{50} of 0.20 μM against AF9 and 0.43 μM against ENL (Figure. 2.6A). These findings clearly demonstrate that DOT1L and AF4 bind with similar binding affinity and compete for the same AF9/ENL interaction site, the C-terminal hydrophobic domain.

2.4.3 Alanine-scanning mutagenesis studies of the interaction between DOT1L 10mer peptide and MLL-fusion proteins

Alanine scanning mutagenesis involves systematic replacement of amino acid residues with alanine. Alanine was the residue chosen to replace the amino acids under examination because it maintains the stereochemistry of the backbone, while eliminating all of the side chain atoms beyond the β-carbon. It can yield important functional insight about PPIs for which there is no structural information. We performed systematic alanine mutagenesis to identify positions

in DOT1L that make essential contacts with the MLL-fusion proteins, AF9 and ENL. We mutated to alanine each of the DOT1L residues from Leu865 to Val874, thus creating 10 different DOT1L mutants and, in an SPR-based competitive assay determined their potency as inhibitors of the PPI.

Using wild type DOT1L10mer peptide as a standard, we compared the binding affinities of each of these ten alanine-mutated DOT1L - based peptides, obtaining the results shown in Figures 2.8.

The three conserved residues in DOT1L, L865, I869 and L871, as well as the similar residue, I867, all failed to tolerate alanine substitution, showing significant decrease in binding affinity to both AF9 and ENL proteins. These results demonstrate the importance of the hydrophobic interactions which is

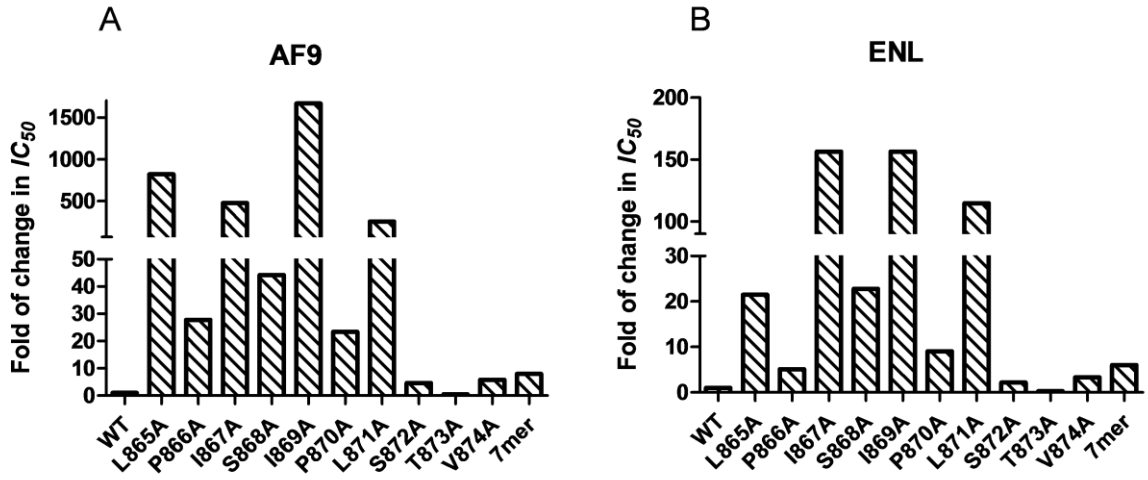


Figure 2.8 Binding affinities of alanine mutated DOT1L 10-mer peptides to AF9 (A) and ENL (B) compared with the wild type DOT1L 10-mer peptide.

consistent with notion of a conserved C-terminal hydrophobic domain in both AF9 and ENL. The non-conserved residues, P866, S868 and P870 showed 10- to 40-fold reductions in their ability to inhibit the PPI. Interestingly, mutation of the last

three residues, S872, T873 and V874 in the DOT1L-10mer peptide, was tolerated well showing only 2- to 6-fold decrease in the binding affinity compared to the wild type peptide. We synthesized wild type 7mer DOT1L peptide (865-871) and tested its binding to AF9 and ENL. Consistent with the mutation data this peptide showed an 8-fold reduction in the binding to AF9 and a 5-fold reduction in binding affinity to ENL protein when compared with the 10mer peptide (Figure 2.8A and B). The binding affinity of DOT1L 7mer peptide to ENL protein was confirmed in an ITC assay showing a K_d of 3.4 μM, 3-fold lower in comparison with the 16mer and 10mer DOT1L peptides (Figure 2.7C). Therefore the alanine scanning mutagenesis studies indicate that the heptapeptide binding motif of DOT1L is essential for interactions with MLL-fusion proteins and in fact is the minimum required fragment. In addition, the results obtained show a similar pattern in the binding of all tested peptides against AF9 and ENL proteins, suggesting that they share similar structural requirements for this interaction; a single amino acid mutation of the hydrophobic conserved residues is sufficient to disrupt the PPI. Overall, these findings provide a rationale towards future efforts in identifying chemical probes that can disrupt the interaction between DOT1L and the MLL-fusion proteins.

2.4.4 The entire C-terminal domain of ENL is required for interaction with DOT1L

Secondary structure predictions of the C-terminal hydrophobic domain in AF9 and ENL indicate the presence of two highly conserved helical segments (Figure 2.9A). Deletion of either of these helical regions completely abrogates the transforming activity of MLL-AF9 and MLL-ENL fusion proteins [32]. To determine if the helical regions are important for the binding of these two MLL-fusion proteins to DOT1L, we prepared ENL H1 (489-544) and ENL H2 (523-559),

A

```
AF9 497GECDKAYLDELVELHRRLMTLRERHILQQIVNLIEE
ENL 489GTYDKAYTDELVELHRRLMALRERNVLQQIVNLIEE
      *  **** ***********:****::**********

AF9 532TGHFHITNTTFDFDLCSLDKTTVRKLQSYLETSGTS568
ENL 525TGHFNVTNTTFDFDLFSLDETTVRKLQSCLEAVAT 559
      ****::********* *** ******** **:
```

B

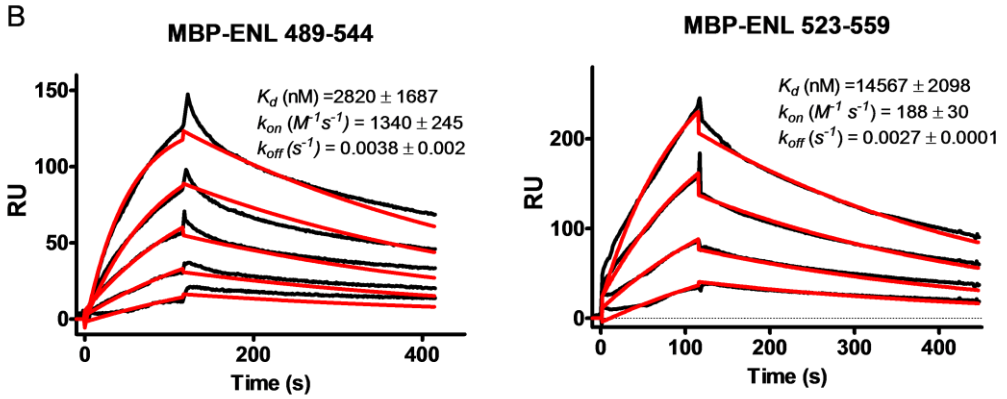


Figure 2.9 (A) Alignment of AF9/ENL C-terminus domains, AF9 (497-568 and ENL (489-559). Conserved residues are highlighted with asterisk. (B) Binding sensorgrams of immobilized Mocr-DOT1L (826 - 1095) with ENL H1 (489-544) and ENL H2 (523-559), tested in the concentration ranges 1 - 16 µM and 5 - 35 µM, respectively. The k_{on}, k_{off} and K_d were calculated by simultaneous non-linear regressions using a 1:1 binding model and BIAevaluation 3.1 software. The experimental data are shown in black while the global fit analyses are shown in red.

two recombinant fragments of ENL protein which correspond to predicted helices 1 and 2, respectively [32]. It was determined that ENL H1 binds to DOT1L (826-1095) with K_D of 2.8 μM and ENL H2 binds with K_d of 14.6 μM (Figure 2.9B), demonstrating 10-fold and 30-fold decreased binding affinity respectively in comparison with the intact ENL C-terminal domain (Table 2.1).

These results demonstrate that the entire C-terminal domain from ENL is required for optimal interaction with DOT1L, which is consistent with the reports that both helical segments are essential for the transformation potential of MLL-ENL fusion protein [5]. To further map and identify the so called “hot-spot” residues in C-terminal domain of ENL, those that are essential for the interaction with DOT1L, we prepared ENL mutant, L550E, which has been reported to block the transforming capacity of the MLL-ENL fusion protein [23]. As was predicted, the fluorescein-labeled DOT1L 10mer peptide failed to bind to this mutant protein (Figure 2.6B). The result obtained identified this conserved residue, L550 and L558 in ENL and AF9, respectively, as a hot-spot and demonstrated that is essential for the interaction with DOT1L.

2.4.5 Binding of DOT1L to the C-terminus domain of ENL induces folding of ENL

NMR spectroscopy was used to gain more information and understanding of the DOT1L-ENL PPI at the structural level, and to determine conclusively whether DOT1L peptides specifically interact with ENL. We obtained high-quality HSQC (Heteronuclear Single Quantum Correlation) spectra of ^{15}N-labeled ENL (489-559) protein. The ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled ENL (489-559)

protein alone (Figure 2.10A) showed poor signal dispersion, indicating that the protein is intrinsically disordered. Addition of the DOT1L protein (826-1095) however (Figure. 2.10A, red spectrum) in a 1:1 ratio, resulted in folding and binding, and the spectrum of the ENL-DOT1L complex shows very extensive chemical shift dispersion consistent with a well folded and stable structure. Comparison of the HSQC NMR spectra of DOT1L-ENL complexes with DOT1L 16mer (red spectrum), 10mer peptide (blue spectrum) and 7mer peptide (green spectrum) (Figure 2.10B), showed very similar chemical shifts and overall, the HSQC spectrum of the ENL-DOT1L peptide complexes overlap with the spectrum of the complex with DOT1L protein. These results further confirm that all three peptides are able to completely recapitulate the chemical shifts of DOT1L protein, and that they bind in a similar way. Consistent with our binding results, the NMR data further demonstrate that all tested DOT1L peptides interact directly with ENL and the heptapeptide binding motif of DOT1L is essential for these interactions and represents the minimum required motif.

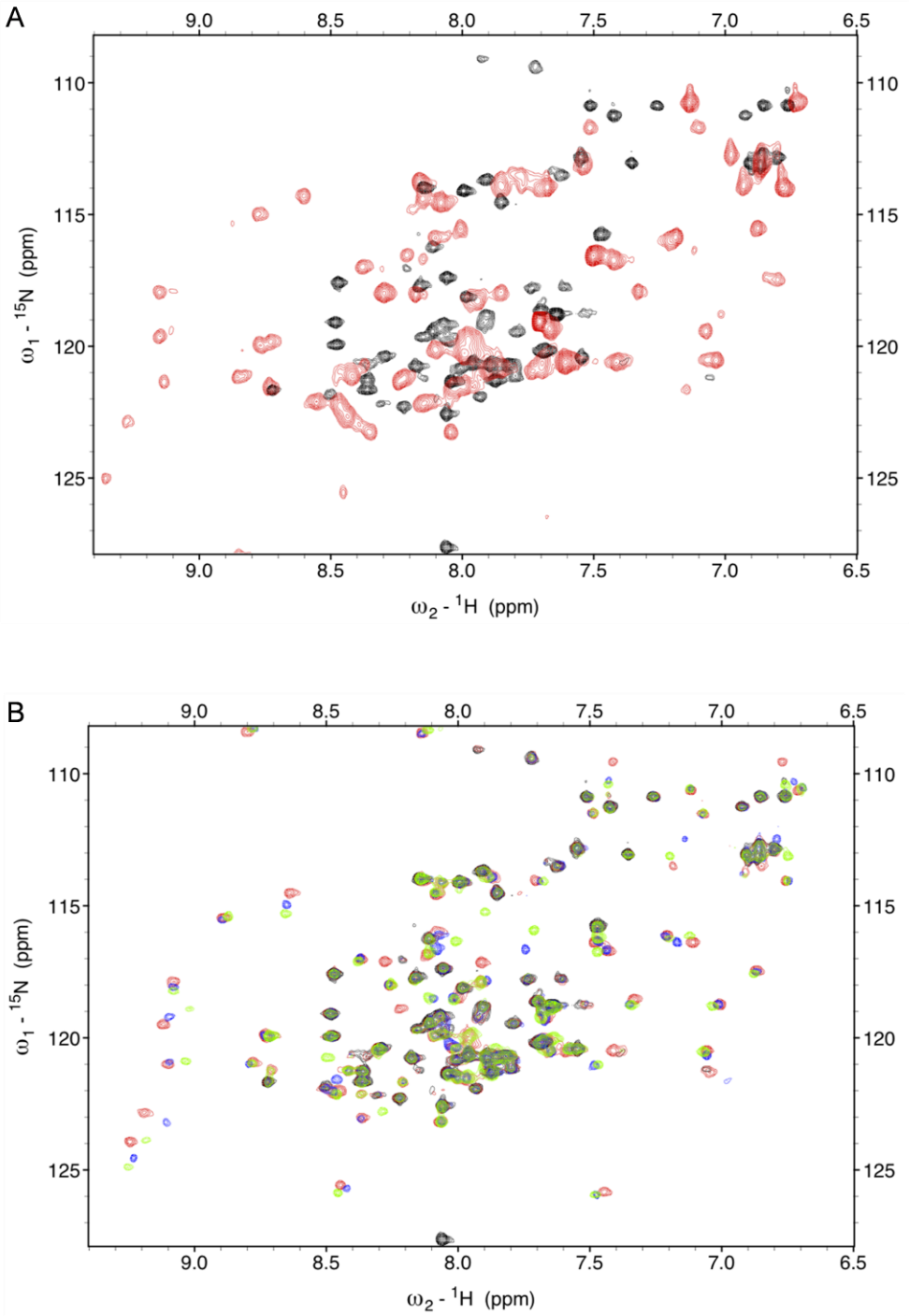


Figure 2.10 Binding of DOT1L to ENL C-terminal domain induces folding of ENL. (A) Overlay of ^{1}H-^{15}N HSQC spectrum from ^{15}N ENL (489-559) (80 μM, black) and ^{15}N ENL (489-559) in complex with DOT1L (826 - 1095) protein in 1:1 molar ratio (red) (B) Overlay of ^{1}H-^{15}N HSQC spectrum from ^{15}N ENL (489-559) (80 μM, black) and ^{15}N ENL (489-559) in a complex with DOT1L 16-mer peptide (red), DOT1L 10mer peptide (blue) or DOT1L 7mer peptide (green) in 1:1 molar ratio.

2.4.6 Homology modeling of DOT1L-ENL complex

Recently the structure of AF9-AF4 complex was determined using NMR spectroscopy (PDB ID: 2LM0) [33]. Consistent with our findings this study also demonstrate that the AF9 C-terminal hydrophobic domain is intrinsically disordered and folds upon binding to AF4. To take advantage of this structural information and to better understand the structural basis of ENL-DOT1L PPI, we have generated a homology model of ENL in complex with DOT1L 7mer peptide. The C-terminal domain of AF9 shows the highest homology to the C-terminus of ENL (amino acids 473 - 568) and therefore the reported NMR structure was used as a template. MOE-align results showed that MLL fusion proteins, ENL and AF9, have high identity and similarity and based on the sequences aligned, the identity of ENL to AF9 is 79.2%. The accuracy of the structures generated by homology modeling is highly dependent on the sequence identity between the target and the template protein. Since ENL and AF9 shared high level od identity the generated homology model is accurate and reliable and can be used for structure based studies .

The generated homology model showed that the interface between ENL and DOT1L 7mer peptide is extensive and hydrophobic, with four aliphatic residues from DOT1L making hydrophobic interaction, especially I867 and I869 which are buried deeply into the hydrophobic core of ENL (Figure 2.11). This model is strongly supported by our alanine mutagenesis studies which demonstrated that mutating these hydrophobic residues to alanine abolished the binding of DOT1L peptide to ENL protein (Figure 2.8). Ser 868 of the DOT1L

peptide is not involved in the hydrophobic interactions and it is in close proximity to the Asp 536 of ENL. Therefore, mutating of Ser 868 to a Ala may induce decrease of the binding affinity which was confirmed with alanine scanning mutagenesis studies. The minimized ENL-DOT1L homology structure indicated that there are at least three hydrogen bonds: two between Phe 537 of ENL and Ile 867 of DOT1L peptide and one between Phe 535 of ENL and Ile 869 of DOT1L.

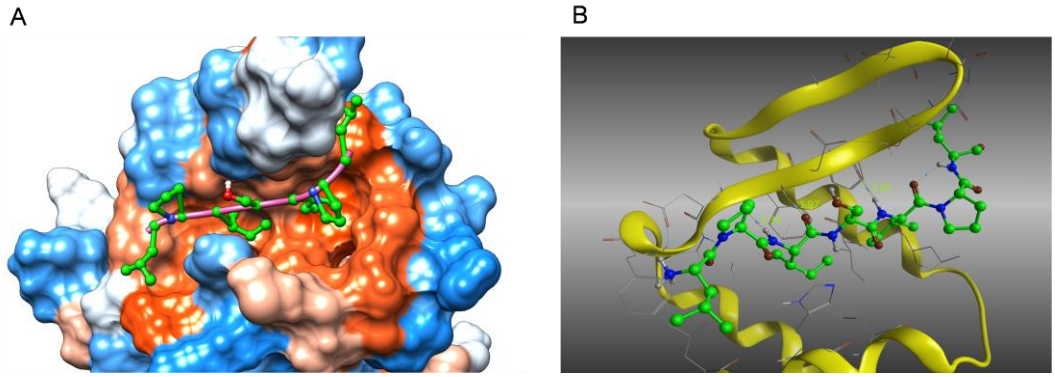


Figure 2.11 Homology model of ENL in complex with DOT1L 7mer peptide. **(**A) The DOT1L 7mer peptide, LPISIPL, is in a long and deep hydrophobic channel. (B) Three hydrogen bonds were revealed for the ENL – DOT1L complex.

In addition, our binding studies demonstrated that L550E point mutation in ENL protein is disrupting the binding of DOT1L peptide (Figure. 2.6B). According to the homology model of ENL complex with DOT1L peptide, L550 is buried in the protein interior through hydrophobic interactions and is involved in packing together secondary structure elements. Mutation of L550 to glutamic acid will have impact on the conformation of the binding pocket and probably on the overall folding of the protein, resulting in no binding of the DOT1L peptide to the protein.

Therefore, this model will allow further understanding of this PPI on the molecular level as well as gaining knowuledge about other PPI that AF9 and ENL are involved. Furthermore, knowing the interactions between DOT1L peptide and

ENL on the structural level will provide critical insights towards developing peptidomimetics as one of the known strategies for targeting PPI.

2.4.7 DOT1L 10mer peptide disrupts in vivo interactions of full length DOT1L with MLL-AF9 fusion protein

To assess whether the DOT1L 10mer peptide can bind and block the interaction between full length DOT1L and MLL-fusion proteins, we transiently co-transfected HEK 293 cells with Flag-tagged human DOT1L protein and Myc-tagged CxxC-AF9 protein. The whole cell lysate was pre-incubated with different concentrations of DOT1L 10mer peptide and co-immunoprecipitation (co-IP) experiments were performed (Figure 2.12A).

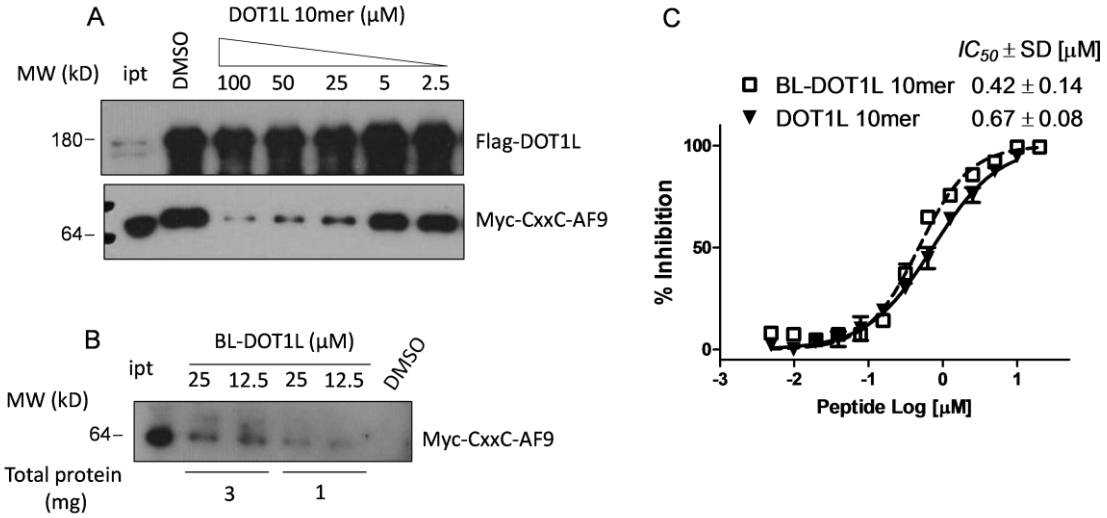


Figure 2.12 DOT1L 10-mer peptide binds cellular MLL-AF9 fusion protein and disrupts its interaction with DOT1L. (A) DOT1L 10-mer peptide disrupts the interaction between DOT1L and MLL-AF9 in cells. Flag-DOT1L and Myc-CxxC-AF9 were co-transfected in HEK 293 cells and co-IP was performed. (B) Pull-down assay using biotin-labeled DOT1L 10-mer peptide. (C) FP competitive curve of BL-DOT1L and DOT1L 10mer peptide in the pull-down assay.

The results obtained demonstrate that the DOT1L 10mer peptide blocks the interaction of cellular DOT1L with CxxC-AF9 in a dose-dependent manner (Figure 2.12A). To further probe and confirm the cellular target for DOT1L 10mer peptide, we synthesized a biotinylated DOT1L 10mer peptide and to determine the binding affinity of this labeled peptide we used FP competitive based assay, which was tested together with the corresponding unlabeled DOT1L 10mer peptide.

The IC_{50} values obtained were similar, confirming that biotin labeling did not affect its binding to AF9 protein (Figure 2.12C). Using this peptide as a tool, we performed streptavidin-biotin pull-down experiments in the HEK 293 cell lysates that showed that the biotinylated DOT1L 10mer peptide recognizes and binds to the cellular CxxC-AF9 protein in a dose- dependent manner (Figure 2.12B). Together, these experiments demonstrate that the DOT1L 10mer peptide binds to the cellular AF9 protein and blocks its interaction with DOT1L, consistent with our *in vitro* binding data using recombinant MLL-fusion proteins.

2.4.8 Identified 10 residues are essential for DOT1L recruitment and MLL-AF9 leukemic transformation

To assess further the functional importance of the recruitment of DOT1L by MLL-fusion proteins and determine whether the 10 residue segment identified in DOT1L is required for MLL-AF9 transformation, we performed colony forming unit (CFU) assays (Figure 2.13).

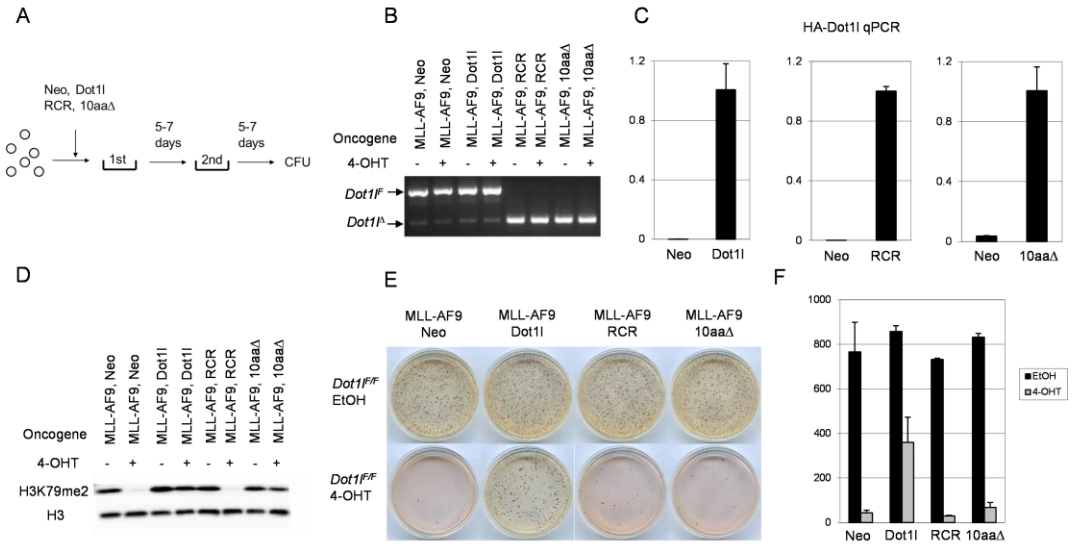


Figure 2.13 AF9 binding site in DOT1L is essential for MLL-AF9 leukemic transformation. (A) Schematic presentation of the colony formation unit (CFU) assay. (B) Genotyping of transduced bone marrow cells. The PCR reaction showed high excision efficiency of endogenous DOT with 4-OHT treatment of all cells. (C) Quantitative PCR of exogenous DOT1L expression. All constructs showed expression compared with Neo vector alone. (D) Western blot of H3K79me2 global level. Histone 3 blot was used as loading control. (E) and (F) Colony formation on methocult plates. Pictures of iodonitrotetrazolium chloride staining (E) and bar graph of colony counts (F) after the second round.

In these experiments MSCV-based vectors were used to transduce mouse bone marrow cells with leukemogenic MLL-AF9 fusion protein from floxed *Dot1l* mice generated as described previously [25]. HA-tagged wild-type mouse DOT1L, methytransferase inactive full-length mDOT1L (RCR) with GSG to RCR mutation in the *S*-adenosylmethionine binding domain and lack enzymatic activity [19, 34], full-length mDOT1L 10aaΔ lacking ten amino acid AF9 interacting residues (863 to 872 which are conserved and correspond to the human DOT1L 865 to 874, Figure 2.5B), or neomycin vector control were introduced into endogenous DOT1L lacking MLL-AF9 transformed cells (Figure 2.13A). Endogenous *Dot1l* excision was confirmed by PCR and expression of exogenous DOT1L by

quantitative PCR of HA tag sequences (Figure 2.13B and C). As was expected, wild-type DOT1L was able to restore H3K79 methylation while mDOT1L (RCR) failed to restore H3K79 methylation (Figure 2.13D). Interestingly, mDOT1L 10aaΔ with an intact histone methyltransferase domain was also able to restore the H3K79 methylation although slightly decreased in comparison with wild-type DOT1L. It is also important to be pointed out that the global level of H3K79 methylation was determined after the first round of replating. It was not possible to check for the global H3K79 methylation after the second round because there were not enough cells for the assay. Consistent with the reported findings, the colony forming potential of MLL-AF9-immortalized cells was completely abolished by introduction of an mDOT1L (RCR) construct after *Dot1l* deletion, while introduction of exogenous wild-type DOT1L was able to rescue the transformation capability and restore CFU formation (Figure 2.13E and F). Despite being able to restore H3K79 methylation level in a similar way as the wild-type DOT1L construct, the mDOT1L 10aaΔ construct failed to restore CFU formation (Figure 2.13E and F). These results strongly suggest that DOT1L interaction with MLL-AF9 and its recruitment are required for transformation by MLL-AF9, and full-length DOT1L lacking the AF9 interacting residues could not rescue CFU formation.

2.4.9 Post-transcriptional modification as a potential mechanism for regulation of the AF9/ENL-DOT1L protein-protein interaction

2.4.9.1 Phosphorylation of serine 868 in DOT1L abolishes the AF9/ENL-DOT1L protein-protein interaction

The effects of phosphorylation regulating protein-protein interactions have been reported earlier [35, 36]. Since serine (S868 and S872) and threonine (T873) are present in the DOT1L peptide, we speculate that phosphorylation might be involved in regulation of the AF9-DOT1L interaction. To investigate the possible role of phosphorylation in regulation of the AF9-DOT1L interaction, we synthesized the corresponding DOT1L peptides, with S868, S872 or T873 phosphorylated.

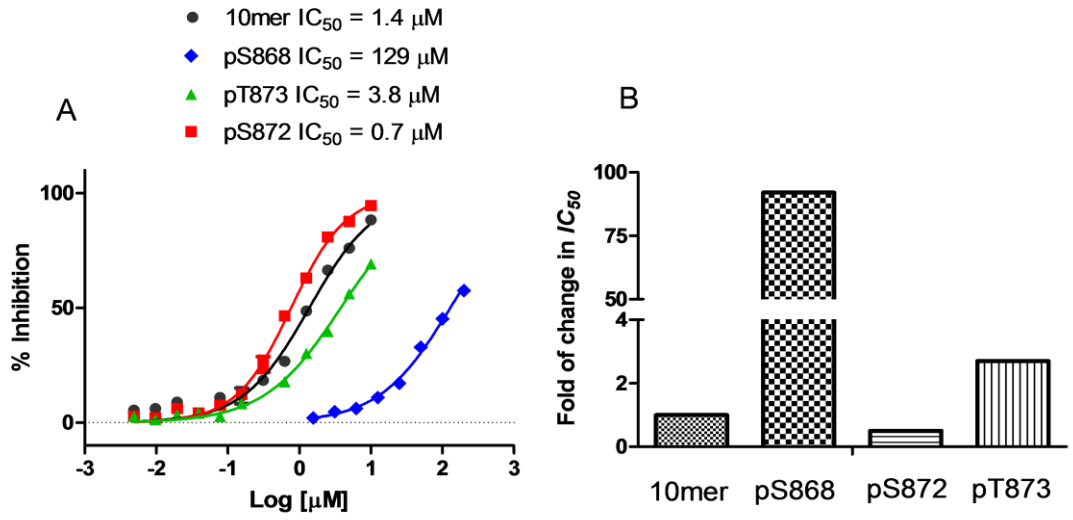


Figure 2.14 Binding of different DOT1L phosphorylated peptides to AF9 protein. (A) Competitive FP binding DOT1L phosphorylated peptides to AF9 protein. (B). Calculated change of fold in IC_{50} of DOT1L phosphorylated peptides compare to DOT1L 10mer.

The binding of these peptides to AF9 proteins was tested by FP-based competitive assay (Figure 2.14A). Phosphorylation of S872 or T873, had no or minimal effect on the AF9-DOT1L interaction, whereas phosphorylation of S868 largely abolished the interaction. DOT1L pS868 10mer peptide showed about a 100-fold increase in the IC_{50} against AF9-DOT1L interaction (Figure 2.14B) compared to the DOT1L 10mer, indicating that phosphorylation of S868 in DOT1L might be involved in regulation of the AF9-DOT1L interaction.

2.4.9.2 Mass spectrometry approach to the identification of phosphorylation sites in DOT1L

There are no reports about the phosphorylation at S868 in DOT1L. To study further the effects of phosphorylation on DOT1L, we prepared the full length Flag tag human DOT1L protein overexpressed in 293 cells and immune-precipitated by anti-Flag antibody (Figure 2.15A).

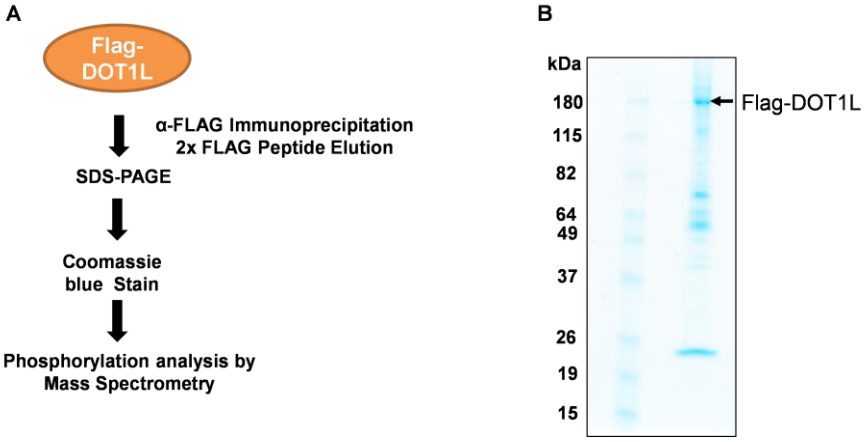


Figure 2.15 Mass spectrometry-based analysis of DOT1L phosphorylation. (A) Schematic of experimental design. (B) Representative Coomassie blue stained gel of Flag-DOT1L protein expressed in 293 cells.

The protein sample was stained with Coomassie blue. The major protein band in the sample, with a molecular weight of about 180 kDa, corresponding to DOT1L protein, was cut out and used for trypsin digestion and mass spectrometry analysis (Figure 2.15B).

The peptide corresponding to DOT1L 854-878, was identified (Figure 2.16). No phosphorylation was observed however in S868, which is potentially involved in regulation of the AF9-DOT1L interaction.

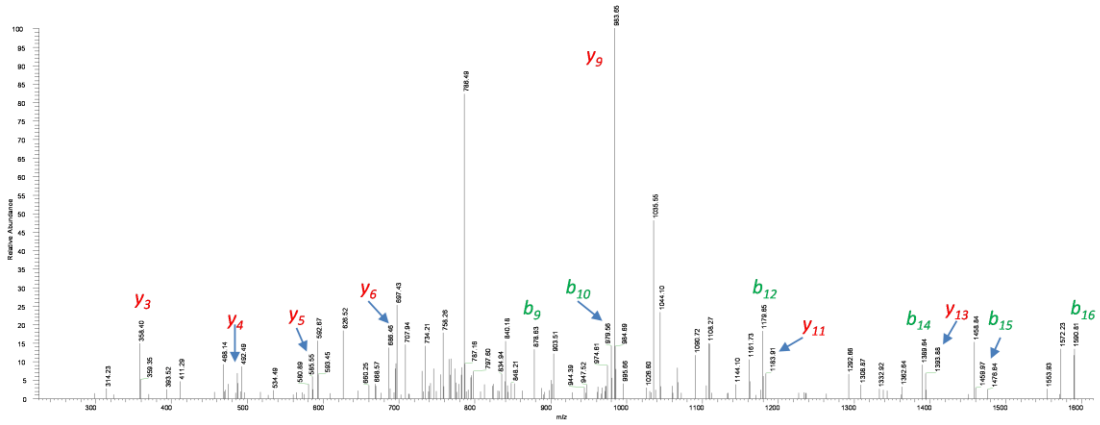


Figure 2.16 A representative MS/MS spectrum of the DOT1L tryptic peptides which contain the AF9 interaction region. DOT1L was immunoprecipitated and subjected to LC-MS/MS analysis. A representative MS/MS spectra assigned to ^{854}AYGSSGELITSLPISIPLSTVQPNK^{878} (precursor ion $[MH]^{+3}$= 858.16 m/z) are shown. Observed b- and y-ions are indicated.

2.4.10 Photo-affinity crosslinking approaches to study of the AF9/ENL-DOT1L interaction

2.4.10.1 Design of DOT1L photo-affinity peptides

Covalent cross-linking has been used successfully to study protein-protein interactions and identification of drug targets [37, 38] [39-41]. Benzophenone is one of the most popular photophores used in photo-affinity labeling. A non-natural amino acid bearing a benzophenone moiety (3-benzoylphenylalanine, Bpa), which covalently crosslinks to protein targets upon exposure to ultraviolet (UV) light, was chosen for covalent crosslinking. Bpa has clean photochemistry and high stability in the conditions of peptide synthesis [39]. Benzophenone derivatives react *via* a triplet biradical mechanism [41](Figure 2.17) with a high

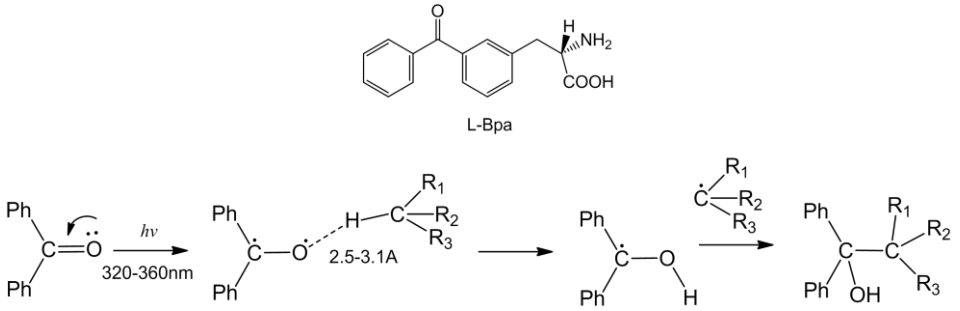


Figure 2.17 Scheme of Formation of covalent adducts from Bap and proteins

preference for reaction with C-H bonds in proximity of 2.5–3.1Å [40]. Thus, the Bpa residue is particularly useful for labeling proteins. To better understand the DOT1L-AF9 interaction, we applied Bpa photo-affinity crosslinking approaches to its study.

DOT1L-Bpa peptides are designed to incorporate Bpa into the peptide. Typically, Bpa was introduced to substitute conservative aromatic residues in the peptide [39, 42, 43] but, since there are no aromatic amino acids in our DOT1L peptide, we chose to introduce Bpa between S872 and T873, a position in close proximity to the core hydrophobic interaction residues (865-871), but not affecting the binding of the peptide to AF9/ENL protein. We designed several DOT1L related peptides, including BL-DOT1L-Bpa, with N-terminal biotin. As a control, phenylalanine was introduced at the same position as Bpa; a tryptic site at the non-interacting site of DOT1L peptide was introduced in order to ensure that the Bpa crosslinking protein was optimal for sequencing by mass spec analysis (Figure 2.18A). Fluorescence polarization based competitive assays performed using Flu-DOT1L peptide and recombinant GB1-AF9 protein confirmed that Bpa incorporation did not affect the target binding (Figure 2.18B).

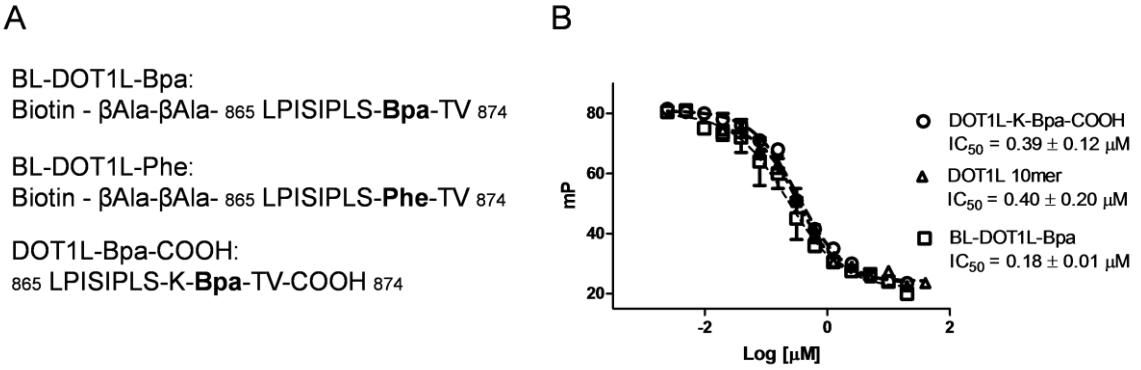


Figure 2.18 Design of DOT1L-Bpa peptide for crosslinking. (A) Sequences of different DOT1L-Bpa peptides. (B) Fluorescence polarization (FP) competitive binding affinities of DOT11L-Bpa to peptides to DOT1L-AF9 interaction.

2.4.10.2 DOT1L-Bpa peptides crosslinking with recombinant GB1-AF9 protein

The designed peptides, including Biotin-DOT1L-Bpa and DOT1L-K-Bpa, were incubated with GB1-AF9 at a ratio of 1:1. Upon exposure to UV light, these two peptides are able to bind to GB1-AF9 and form the crosslinking product, which was detected by Coomassie staining (Figure 2.19A and C).

Western blot using antibodies against biotin indicate the specificity of the formed product. To exclude the possibility that GB1 tag might be involved in the crosslinking, DOT1L-Bpa peptide was incubated with GB1 tag alone. No crosslinking products were observed upon UV exposure, indicating that the covalent capture of GB1-AF9 protein by DOT1L-Bpa is indeed selective (Figure 2.19D).

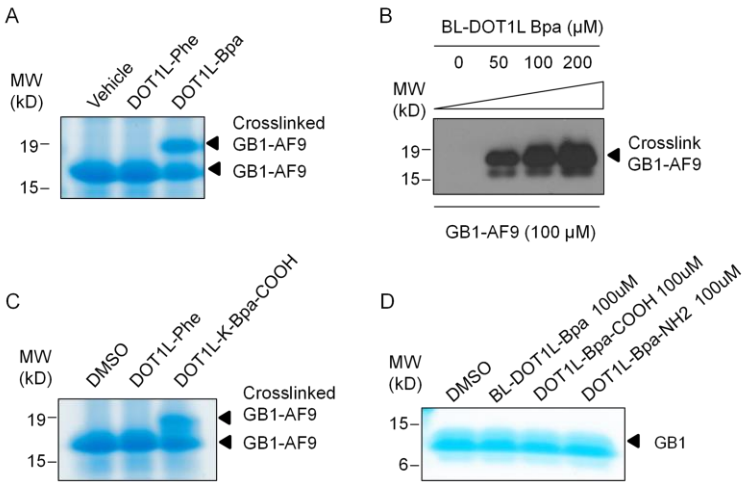


Figure 2.19 (A) Crosslinking product of BL-DOT1L-Bpa peptides with GB1-AF9 proteins by Commassie blue staining. (B) Western blot of BL-DOT1L-Bpa peptides and GB1-AF9 crosslinking product. (C) Crosslinking product of DOT1L-K-Bpa-COOH peptide with GB1-AF9 protein by Commassie blue staining. (D) Crosslinking of DOT1L-Bpa peptides with GB1 tag.

2.4.10.3 DOT1L-Bpa peptide crosslink with overexpressed cellular AF9 protein

Having demonstrated DOT1L-Bpa peptide was selective for its protein target *in vitro*, we next investigated whether the photo-reactive peptide could target and crosslink to its physiologic binding partners. We incubated biotinylated DOT1L-Bpa peptide with a cellular extract from 293 cells overexpressed with Myc-CxxC-AF9 protein in the presence of UV irradiation, followed by SA-based pull-down, stringent washing to eliminate non-covalent binders, and western blot analysis. While UV irradiated cell extracts treated with the DOT1L-Phe control peptide showed no cross-linked product, DOT1L-Bpa peptide crosslinks to endogenous Myc-CxxC-AF9 proteins in a dose-dependent manner (Figure 2.20A and B).

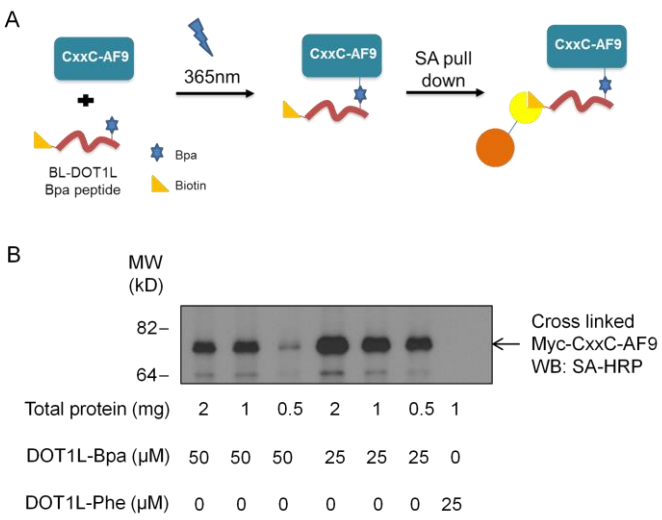


Figure 2.20 Crosslinking and pull down of AF9 protein in 293 cell lysates. (A) Scheme of crosslinking and pull down assay of AF9 proteins by BL-DOT1L-Bpa peptides. (B) Western blot of crosslinked CxxC-AF9 protein.

2.5 Discussion

2.5.1 AF9 or ENL-DOT1L interaction is a promising therapeutic target

We have identified and mapped the protein-protein interaction site between DOT1L and MLL-fusion proteins, AF9 and ENL. Using SPR and FP based binding assays, ITC, HSQC NMR, immunoprecipitation and pull-down experiments, we determined the binding affinity of these PPIs and defined the AF9/ENL interaction site as the region corresponding to DOT1L $_{865}$LPISIPLSTV$_{874}$. The importance of this binding region is indicated by its conservation in different species. Our *in vitro* and *in vivo* findings demonstrated that the corresponding synthetic DOT1L 10mer peptide can competitively interfere with the interaction between MLL-fusion proteins, AF9 and ENL, and DOT1L. The alanine scanning mutagenesis studies showed the critical importance of several conserved residues, L865, I867, I869 and L871, for binding to AF9 and ENL. Binding studies confirmed a direct interaction between AF9/ENL and DOT1L, and showed that an intact C-terminal domain in MLL-fusion proteins is critical for optimal interaction. Furthermore, ENL mutant, L550E, which was reported to block the transforming capacity of MLL-ENL fusion protein [23], completely lost its ability to interact with DOT1L 10mer peptide. These results together with reported functional studies further demonstrate a high correlation between the DOT1L recruitment and the transforming potential of MLL-ENL. Several groups have demonstrated that DOT1L is required for MLL-AF4 and

MLL-AF9 fusions-mediated leukemic transformation [25, 44-47], establishing the crucial role of DOT1L in MLL-rearranged leukemia [48] and, consistent with these findings, our functional studies using a colony forming unit assay provide strong evidence that the identified $DOT1L_{865-874}$ fragment is essential for the DOT1L recruitment and MLL-AF9 leukemic transformation. The colony forming potential of the MLL-AF9 immortalized cells was completely abolished by introduction of a DOT1L construct lacking the 10 amino acid AF9 interacting residues. In addition to H3K79 methylation, these results highlight the importance of the DOT1L recruitment and its interaction with MLL-AF9 for the transforming activity of this MLL fusion oncogene.

We thus validated the PPIs between DOT1L and AF9 or ENL, respectively, as promising therapeutic targets and the basis of a potential strategy for pharmacological targeting of DOT1L. Our results indicate that disruption of AF9-DOT1L interaction abolishes MLL-AF9 leukemia transformation, without affecting the global level of H3K79 methylation level. More importantly, they also suggest that selective disruption of this PPI is a promising therapeutic strategy with potentially fewer adverse effects than enzymatic inhibition of DOT1L for MLL-fusion protein associated leukemia.

Our binding studies clearly demonstrate that DOT1L and AF4 proteins, as well as the DOT1L 10mer and AF4 14mer peptide, bind to the same C-terminal hydrophobic domain of AF9 and ENL (Figure 2.20A), consistent with biochemical and functional analyses of protein complexes associated with MLL-fusion proteins. This shows that AF9 and ENL exist in two mutually exclusive complexes, AF9/ENL-DOT1L and AF9/ENL-AF4-pTEFb [14, 23].

Figure 8

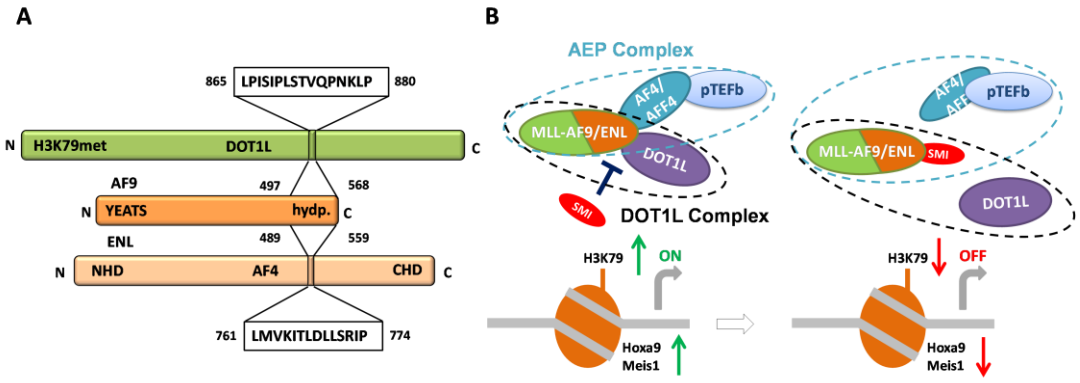


Figure 2.21 Schematic of proposed model for targeting DOT1L and MLL-fusion protein-protein interactions. A. AF9 binding sites mapped in DOT1L and AF4 proteins (Srinivasan et al, 2004). B. A small-molecule inhibitor (SMI) that binds to C-terminal domain of AF9/ENL will disrupt the MLL-fusion protein complexes involved in mixed lineage leukemia, the DOT1L and AEP complex.

Based on our results, we propose that small molecules that bind to the conserved hydrophobic domain of AF9 and ENL should abolish the AF9 interactions with both AF4 and DOT1L, disrupt the higher order complex AEP and the complex with DOT1L, and thereby blocking the MLL fusion-mediated leukemogenesis (Figure 2.20B). Taken together, this study supports the

hypothesis that DOT1L recruitment is a potential therapeutic target in MLL-rearranged leukemia and discovery and development of chemical probes that can block these PPI is warranted.

2.5.2 Signal transduction as potential mechanism for regulation of DOT1L recruitment

As an important post-transcription modification, phosphorylation is widely involved in the regulation of protein-protein interactions by affecting the stability, kinetics and specificity of interactions [49]. Our results show that phosphorylation of S868 of DOT1L abolishes the AF9-DOT1L interaction, indicating that phosphorylation might be involved in regulation of DOT1L recruitment by AF9 and ENL. Recently it has been reported that phosphorylation of T766 of AF4 also resulted in a ~30-fold reduction in the affinity for interaction with AF9 [33]. Together with our findings, these results indicate that phosphorylation sites within the consensus AF9 binding sequences may allow transient and reversible regulation of the interaction. Although phosphorylation of S868 was not identified in our mass spectrometry analysis, several considerations should be taken into account before reaching a definite conclusion. The exogenous tagged DOT1L protein overexpressed in 293 cells may not be an optimized system for studying DOT1L phosphorylation since some phosphorylation events are specific to cell type and cell cycle [50], and several approaches can be applied to further explore the role of phosphorylation in regulating DOT1L recruitment. Production

of phosphorylated S868 specific antibodies and profiling the panel of cell lines with the specific antibody might give some insights into the phosphorylation status. Predictions according to Human Protein Resource data base and based on the sequence of the DOT1L 10mer peptide indicated that this region might be a possible substrate motif of several Ser/Thr kinases, including GSK-3, MAPPK, PTK. Another approach might be use of the DOT1L peptides as substrates and screening a panel of Ser/Thr kinases, to determine if the Ser868 might be the substrate of certain known kinases.

2.5.3 DOT1L photo affinity peptide may serve as a useful research tool for MLL

Many critical biological interactions are transient or are of insufficient affinity to withstand the experimental conditions required to prepare cellular extracts or isolate complexes by affinity chromatography. Our results indicate that DOT1L-Bpa peptide was able to trap the transitory interactions as well as static binding events and can be applied to probe the mechanisms of dynamic MLL fusion protein complexes (Figure 2.20). In addition, photo-crosslinking approaches have been applied to locate protein interaction sites in combination with MS/MS and computational docking strategy. We have also developed DOT1L-Bpa peptides optimal for mass spectrometry analysis (Figure 2.18A). The DOT1L-K-Bpa peptide showed similar binding affinities as the DOT1L 10mer peptide (Figure 2.18B), and was able to crosslink with GB1-AF9 efficiently

(Figure 2.19C). A unique and compelling feature of this approach is the ability to conduct rigorous validation studies for the identified targets and their sites of interaction. DOT1L-Bpa peptides might serve as research tools that efficiently, selectively, and irreversibly crosslink both static and dynamic protein targets for proteomic and mechanistic analyses in MLL. Thus, the development and application of photo-reactive peptides represents a potentially powerful approach to identifying, distinguishing, and modulating protein targets relevant to pathologic cellular processes.

2.6 References

1. Huret, J.L., P. Dessen, and A. Bernheim, *An atlas of chromosomes in hematological malignancies. Example: 11q23 and MLL partners.* Leukemia, 2001. **15**(6): p. 987-9.
2. Zhang, W., et al., *Aldosterone-induced Sgk1 relieves Dot1a-Af9-mediated transcriptional repression of epithelial Na+ channel alpha.* The Journal of clinical investigation, 2007. **117**(3): p. 773-83.
3. Schulze, J.M., A.Y. Wang, and M.S. Kobor, *YEATS domain proteins: a diverse family with many links to chromatin modification and transcription.* Biochemistry and cell biology = Biochimie et biologie cellulaire, 2009. **87**(1): p. 65-75.
4. Schulze, J.M., A.Y. Wang, and M.S. Kobor, *Reading chromatin: insights from yeast into YEATS domain structure and function.* Epigenetics : official journal of the DNA Methylation Society, 2010. **5**(7): p. 573-7.
5. Slany, R.K., C. Lavau, and M.L. Cleary, *The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX.* Molecular and cellular biology, 1998. **18**(1): p. 122-9.
6. Hemenway, C.S., A.C. de Erkenez, and G.C. Gould, *The polycomb protein MPc3 interacts with AF9, an MLL fusion partner in t(9;11)(p22;q23) acute leukemias.* Oncogene, 2001. **20**(29): p. 3798-805.
7. Zeisig, D.T., et al., *The eleven-nineteen-leukemia protein ENL connects nuclear MLL fusion partners with chromatin.* Oncogene, 2005. **24**(35): p. 5525-32.
8. Garcia-Cuellar, M.P., et al., *The ENL moiety of the childhood leukemia-associated MLL-ENL oncoprotein recruits human Polycomb 3.* Oncogene, 2001. **20**(4): p. 411-9.
9. Tan, J., et al., *CBX8, a polycomb group protein, is essential for MLL-AF9-induced leukemogenesis.* Cancer Cell, 2011. **20**(5): p. 563-75.
10. Monroe, S.C., et al., *MLL-AF9 and MLL-ENL alter the dynamic association of transcriptional regulators with genes critical for leukemia.* Exp Hematol, 2011. **39**(1): p. 77-86 e1-5.
11. Mueller, D., et al., *A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification.* Blood, 2007. **110**(13): p. 4445-54.
12. Srinivasan, R.S., A.C. de Erkenez, and C.S. Hemenway, *The mixed lineage leukemia fusion partner AF9 binds specific isoforms of the BCL-6 corepressor.* Oncogene, 2003. **22**(22): p. 3395-406.
13. Zhang, W., et al., *Dot1a-AF9 complex mediates histone H3 Lys-79 hypermethylation and repression of ENaCalpha in an aldosterone-*

sensitive manner. The Journal of biological chemistry, 2006. **281**(26): p. 18059-68.
14. Biswas, D., et al., *Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes.* Proc Natl Acad Sci U S A, 2011. **108**(38): p. 15751-6.
15. Garcia-Cuellar, M.P., et al., *ENL, the MLL fusion partner in t(11;19), binds to the c-Abl interactor protein 1 (ABI1) that is fused to MLL in t(10;11)+.* Oncogene, 2000. **19**(14): p. 1744-51.
16. Nie, Z., et al., *Novel SWI/SNF chromatin-remodeling complexes contain a mixed-lineage leukemia chromosomal translocation partner.* Molecular and cellular biology, 2003. **23**(8): p. 2942-52.
17. Srinivasan, R.S., et al., *The synthetic peptide PFWT disrupts AF4-AF9 protein complexes and induces apoptosis in t(4;11) leukemia cells.* Leukemia, 2004. **18**(8): p. 1364-72.
18. Feng, Q., et al., *Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain.* Current biology : CB, 2002. **12**(12): p. 1052-8.
19. Okada, Y., et al., *hDOT1L links histone methylation to leukemogenesis.* Cell, 2005. **121**(2): p. 167-78.
20. Reisenauer, M.R., et al., *AF17 competes with AF9 for binding to Dot1a to up-regulate transcription of epithelial Na+ channel alpha.* The Journal of biological chemistry, 2009. **284**(51): p. 35659-69.
21. Mohan, M., et al., *Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom).* Genes Dev, 2010. **24**(6): p. 574-89.
22. Kim, S.K., et al., *Human histone H3K79 methyltransferase DOT1L protein [corrected] binds actively transcribing RNA polymerase II to regulate gene expression.* The Journal of biological chemistry, 2012. **287**(47): p. 39698-709.
23. Yokoyama, A., et al., *A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription.* Cancer Cell, 2010. **17**(2): p. 198-212.
24. DelProposto, J., et al., *Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications.* Protein Expr Purif, 2009. **63**(1): p. 40-9.
25. Jo, S.Y., et al., *Requirement for Dot1l in murine postnatal hematopoiesis and leukemogenesis by MLL translocation.* Blood, 2011. **117**(18): p. 4759-68.
26. Lewis, I.A., S.C. Schommer, and J.L. Markley, *rNMR: open source software for identifying and quantifying metabolites in NMR spectra.* Magn Reson Chem, 2009. **47 Suppl 1**: p. S123-6.

27. Goddard, T.D. and D.G. Kneller, *SPARKY 3*, in *University of California, San Francisco*.
28. Keller, A., et al., *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.* Anal Chem, 2002. **74**(20): p. 5383-92.
29. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry.* Anal Chem, 2003. **75**(17): p. 4646-58.
30. Zhang, W., et al., *Dot1a-AF9 complex mediates histone H3 Lys-79 hypermethylation and repression of ENaCalpha in an aldosterone-sensitive manner.* J Biol Chem, 2006. **281**(26): p. 18059-68.
31. Zhang, W., et al., *Aldosterone-induced Sgk1 relieves Dot1a-Af9-mediated transcriptional repression of epithelial Na+ channel alpha.* J Clin Invest, 2007. **117**(3): p. 773-83.
32. Slany, R.K., C. Lavau, and M.L. Cleary, *The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX.* Mol Cell Biol, 1998. **18**(1): p. 122-9.
33. Leach, B.I., et al., *Leukemia fusion target AF9 is an intrinsically disordered transcriptional regulator that recruits multiple partners via coupled folding and binding.* Structure, 2013. **21**(1): p. 176-83.
34. Min, J., et al., *Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase.* Cell, 2003. **112**(5): p. 711-23.
35. McAvoy, T., et al., *Regulation of neurabin I interaction with protein phosphatase 1 by phosphorylation.* Biochemistry, 1999. **38**(39): p. 12943-9.
36. Ku, N.O., J. Liao, and M.B. Omary, *Phosphorylation of human keratin 18 serine 33 regulates binding to 14-3-3 proteins.* The EMBO journal, 1998. **17**(7): p. 1892-906.
37. Majmudar, C.Y., et al., *A high-resolution interaction map of three transcriptional activation domains with a key coactivator from photo-cross-linking and multiplexed mass spectrometry.* Angew Chem Int Ed Engl, 2009. **48**(38): p. 7021-4.
38. Braun, C.R., et al., *Photoreactive stapled BH3 peptides to dissect the BCL-2 family interactome.* Chem Biol, 2010. **17**(12): p. 1325-33.
39. Dorman, G. and G.D. Prestwich, *Benzophenone photophores in biochemistry.* Biochemistry, 1994. **33**(19): p. 5661-73.
40. Prestwich, G.D., et al., *Benzophenone photoprobes for phosphoinositides, peptides and drugs.* Photochem Photobiol, 1997. **65**(2): p. 222-34.
41. Dorman, G. and G.D. Prestwich, *Using photolabile ligands in drug discovery and development.* Trends Biotechnol, 2000. **18**(2): p. 64-77.

42. Saghatelian, A., et al., *Activity-based probes for the proteomic profiling of metalloproteases.* Proc Natl Acad Sci U S A, 2004. **101**(27): p. 10000-5.
43. Vodovozova, E.L., *Photoaffinity labeling and its application in structural biology.* Biochemistry (Mosc), 2007. **72**(1): p. 1-20.
44. Bernt, K.M., et al., *MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L.* Cancer Cell, 2011. **20**(1): p. 66-78.
45. Daigle, S.R., et al., *Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor.* Cancer Cell, 2011. **20**(1): p. 53-65.
46. Nguyen, A.T., et al., *DOT1L, the H3K79 methyltransferase, is required for MLL-AF9-mediated leukemogenesis.* Blood, 2011. **117**(25): p. 6912-22.
47. Chang, M.J., et al., *Histone H3 lysine 79 methyltransferase Dot1 is required for immortalization by MLL oncogenes.* Cancer research, 2010. **70**(24): p. 10234-42.
48. Bernt, K.M. and S.A. Armstrong, *A role for DOT1L in MLL-rearranged leukemias.* Epigenomics, 2011. **3**(6): p. 667-70.
49. Seet, B.T., et al., *Reading protein modifications with interaction domains.* Nat Rev Mol Cell Biol, 2006. **7**(7): p. 473-83.
50. Bagowski, C.P., et al., *Cell-type specific phosphorylation of threonines T654 and T669 by PKD defines the signal capacity of the EGF receptor.* The EMBO journal, 1999. **18**(20): p. 5567-76.

Chapter 3

Identification of small molecular inhibitor against AF9/ENL-DOT1L Protein-Protein Interaction by High-throughput Screening

3.1 Abstract

The MLL fusion proteins AF9 and ENL, activate target genes in part *via* recruitment of the histone methyltransferase DOT1L (Disruptor Of Telomeric silencing 1-Like). In this Section, we report characterization of the protein-protein interaction between DOT1L and AF9/ENL. Binding studies demonstrate that only 10 amino acids in DOT1L are involved in the interaction with AF9/ENL and four of these are conserved hydrophobic residues essential for this PPI. Significantly, a DOT1L construct lacking the residues that interact with AF9/ENL completely abolishes the colony forming potential of the MLL-AF9 immortalized cells. This result demonstrates that the PPI of DOT1L and AF9/ENL is required for transformation of MLL-AF9 and suggests that disruption of this PPI is a promising therapeutic strategy for MLL-fusion protein-associated leukemia. We employed a high-throughput screening approach for the discovery of inhibitors targeting the

recruitment of DOT1L by MLL-fusion proteins. For this purpose we developed and optimized a fluorescence polarization (FP) based assay in a 384-well format using AF9 protein and fluorescein-labeled DOT1L 10mer peptide with a Z' factor of 0.75. Approximately 100,000 compounds were screened at a concentration of 20 μM for their ability to inhibit binding of the DOT1L peptide to AF9. A total of 1191 compounds with inhibitory activity larger than three standard deviations in primary screening were selected for a confirmatory assay in order to identify fluorescent compounds, quenchers and to confirm their binding. A total of 164 compounds were selected and tested in a dose-response screen and 39 active, structurally diverse compounds were identified. Eight of these compounds, whose selection was based on their potency, chemical structure and availability, were purchased and retested in the FP competitive assay. The activity of four compounds was confirmed in both assays with IC_{50} values from 10 - 60 μM in the FP assay, and 20 – 70 μM in the SPR assay. These compounds served as a starting point for the development of small molecule inhibitors of the AF9-DOT1L interaction.

3.2 Background

3.2.1 Challenges of targeting protein-protein interactions

Targeting the interactions between proteins has significant therapeutic potential [1-3], but discovery of compounds that can interfere with PPIs is not a trivial task. Protein-protein interfaces are likely to be wider and shallower than canonical small molecule binding sites on proteins [4, 5] and the contact surfaces involved in a typical PPI are from 1,500–3,000Å^2 [6, 7] compared with ~300–1,000Å^2 surfaces involved in protein-small-molecule interaction[8, 9]. X-ray structures of protein-protein complexes show that protein-protein interfaces are generally flat and often lack the grooves and pockets that characterize the surfaces of proteins that bind to small molecules [10, 11]. Secondly, unlike classical protein targets such as enzymes and G-protein-coupled receptors, which have natural substrate or ligand as a starting point for inhibitor design, protein-protein interactions do not have natural small-molecule partners that can bind to the PPI interface.

Several lines of evidence however provide encouragement to the search for small molecules that target protein-protein interfaces. Although these interfaces are large, mutational studies have shown that only a few of the residues involved in the protein-protein binding contribute to most of the free energy of binding [12-15]. These residues, known as “hotspots”, account for a

much reduced contact surface of the protein involved in the protein-protein interaction and are usually found at the center of the contact interface. The fact that all protein-protein interfaces that have been studied in detail contain hot spots of binding energy has important implications for the development of small molecules that disrupt PPIs. Hot spots, which generally account for less than 10 % of the whole protein-protein interface [15] effectively allow reduction of the target area in rational design studies from thousands of square Ångströms to a handful of key contacts in a specific area [12, 16-18] and targeting the “hot spots” may be a feasible strategy [5, 19].

3.2.2 Strategies of targeting protein-protein interactions

Considerable progress has been made in the past decades in finding inhibitors that disrupt protein-protein interfaces and successful examples of the use of hot spots to drive rational design studies are beginning to appear in the literature [20]. There are several dominant strategies for the discovery of inhibitors that bind at protein-protein interfaces: use of protein peptide mimetics, screening of large chemical libraries *via* high-throughput screening, fragment based screening, and *in silico* screening. These approaches all have distinct benefits and drawbacks.

Peptidomimetics

Peptidomimetics are compounds whose essential elements or pharmacophore, mimic a natural peptide or protein in 3D space and which retain the ability to interact with the biological target and produce the biological effects [21]. Peptidomimetics have been applied to the development of PPI inhibitors. A small-molecule peptidomimetic inhibitor of the MLL1/WDR5 protein-protein interaction was developed based on the minimum binding motif derived from MLL1 [22]. The designed peptidomimetic binds to the target protein WDR5, with high affinity (K_i <1 nM) and is a potent antagonist of MLL1 activity.

High-throughput screening

High-throughput screening (HTS) refers to the process by which large numbers of compounds, generally on the order of tens of thousands, are examined for their ability to bind to a target of interest. HTS has played an integral role in the identification of small molecule leads in drug discovery and since its first use in the early 1990s, it has become a well-established process in drug discovery research [23]. The development of automation technology and commercially available chemical libraries has allowed laboratories not specializing in chemistry to actively attempt to discover small molecules that modulate a macromolecular target of interest and a number of inhibitors of PPI have been identified by HTS. An early example is the screening of inhibitors

modulating the Bax/Bcl-xL interaction [24]. In this work, the authors used an initial *in vitro* fluorescence displacement screen with the fluorescently labeled BH3 domain of Bak and Bcl-xL. From an initial screen of about 16,000 compounds, several hits with K_d in the low micromolar range for Bcl-xL were identified. These lead compounds were further tested in cells and were found to successfully induce apoptosis by interfering with the Bak/Bcl-xL interaction. Another example is the identification from an HTS of nutlins as p53-MDM2 interaction inhibitors. Co-crystallization of nutlins with the target protein MDM2 showed the nutlin effectively binding to and blocking the pocket that would have been occupied by the transactivation domain of p53. Further experiments revealed that these compounds bind to MDM2 and activate the p53 pathway in cancer cells leading to cell cycle arrest, apoptosis and inhibition of tumor growth in mice[25].

Fragment based discovery

The fragment-based discovery technique screens small components for binding to a protein of interest and subsequently links suitable motifs together. Generally, fragments are small organic compounds, typically less than 200 Da in mass, weak in binding affinity but with better ligand efficiency, which is a measurement of the binding energy per atom of a ligand to its binding partner.. Initially, fragment libraries are screened for binding to a protein target judged by biophysical or biochemical data. The fragments identified in this way can be

linked together through links spanning distant sites on the protein-protein interaction surface. The discovery of ABT-737, which is a Bcl-2 family protein inhibitor [26] is an example. Two fragments with binding affinities between 300 and 4,000 μM to the Bcl-xL protein were initially identified by a high-throughput NMR-based method. These two fragments bind to distinct, proximal subsites within the BH3 peptide binding cleft. The linkage of these two proximal fragments and further chemical modification achieved the highly potent Bcl-2 family protein inhibitor ABT-737. One of the analogues of this compound, ABT-263, is now in clinical trials for the treatment of cancer [26].

In silico screening

In silico screening is an approach using the structural knowledge of a protein-protein interface in combination with computational modeling. It has been employed to identify PPI inhibitors when the structure of the interaction site is known. For instance, a virtual screen was used as the starting point in the search for a compound that targets the SH3 binding surface of the HIV type 1 Nef protein. High-throughput docking and application of a pharmacophore filter identified compounds that were shown to possess micromolar affinity and compete with the Nef-SH3 interaction in a cell-based assay [27].

With the exception of HTS, these approaches require additional pre-existing knowledge of the target protein, specifically the structure of the desired target and knowledge of its binding hot-spots.

3.2.3 Biochemical assays for high-throughput screening

Unlike the traditional HTS with enzymes as targets, the screening assays for protein-protein interactions monitor the binding of two or more proteins, and can be used to measure inhibition of the PPIs by small molecules or other modulators, such as antibodies or peptides. A growing number of biochemical assays have been used for detection and screening of protein-protein interactions in HTS. These include fluorescence polarization (FP), Fluorescent/Forster resonance energy transfer (FRET), Enzyme-linked immunosorbant assays (ELISA) and AlphaLISA [28]. An overview of these assay formats is provided in Table 3.1.

To measure a PPI by ELISA, one of the proteins is attached to a stationary surface and the second protein is then allowed to bind to it. The second protein is detected by binding of an antibody that is linked to an enzyme. ELISA analyses can be very sensitive, because the readout is amplified by the use of an enzyme. Further amplification can be achieved by use of multiple layers, such as secondary antibodies. Assays of this sort are unlikely to suffer from compound interference since the compound is not present in the processing

step. Most ELISAs however require multiple incubation and washing steps, which are time-consuming for in automated and bench-top assays. Washing can also disrupt weak (K_d > 1 μM) interactions.

Table 3.1 Overview of different biochemical assays applied in high-throughput screening

Method	FP	ELISA/DELFIA	FRET	AlphaScreen
Size of protein and/or complex	Labeldd ligand need to be >1500Da and protein need to be > 10kDa	No limit	Distance between donor and acceptor need to be <9 nm	Distance between donor and acceptor need to be <9 nm
Sensitivity	[FP probe] ~nM [Protein] determined by Kd of PPI	Can be as low as fg/well	Depends on Kd of PPI, typically 1-100 nM	Can be as low as fg/well
Washing steps	no	yes	no	no
Signal range (high/low)	3-5 fold	4-5 log	<10 fold	2-3 log
Cost	Low - moderate	Low - moderate	Moderate - high	High
Advantages	Low standard deviation, direct PPI measurement	Low spectroscopic interference from compounds	Radiometric, readily controlled for spectral artifacts	Monitor the widest range of complex size
Disadvantages	Interference of fluorescence compounds	Sensitive to off rate, laborious	Signal window dependent on distance and dye orientation	Sensitive to ambient light

The other three assays, FP, FRET and AlphaLISA, are all mix-and-read in nature. There are no additional washing steps required and this leads to a wider

dynamic range and higher plate throughput. FRET and AlphaScreen are proximity measurements, in that they rely on the protein partners being within 10 – 90 Å of one another.

Fluorescent Polarization (FP) measures the tumbling time, which is related to its molecular mass, experienced by a fluorophore. It is a sensitive nonradioactive method for the study of molecular interactions in solution [29-32] and can be used to measure association and dissociation of two molecules if one of the molecules is relatively small and fluorescently labeled. When a fluorescently labeled molecule is excited by polarized light, it emits light whose degree of polarization is inversely proportional to the rate of molecular rotation which in turn is largely dependent on molecular mass, with larger masses showing slower rotation. Thus, when a small, fluorescent biomolecule, such as a fluorescently labeled small peptide or ligand (typically < 1500 Da), is free in solution, it will emit low fluorescence polarization, but when bound to a larger (e.g. > 10,000 Da) molecule, such as a protein, the rotational movement of the fluorophore decelerates and the fluorescence polarization will be highly polarized. Thus, the binding of a fluorescently labeled small molecule or peptide to a protein can be monitored by the change in polarization. FP-based technology has a number of key advantages for monitoring bimolecular interactions, especially for HTS applications. It does not involve radioisotopes and employs a homogenous

“mix-and-read” format with no wash steps, multiple incubations or separations. FP measurement is carried out directly in solution; no perturbation of the sample is required, making the measurement faster and more direct than immobilization-based methods like ELISA and it is readily adaptable to low volumes (30 µl for a 384-well plate or 5 µl for a 1536-well plate). As a fluorescence-based technology, FP is subject to optical interference from compounds that absorb at the excitation or emission wavelengths of the fluorescent probe. The method is also sensitive to the presence of fluorescence from test compounds but the use of red-shifted probes can minimize background fluorescence interference. Since FP assay is sensitive, homogeneous and simple, it has been used in primary HTS and successfully for identification of small molecules as inhibitors of protein-protein interactions [33-36].

3.2.4 Multiple approaches for validation of HTS hits

The hits identified by primary HTS have to be validated further and it must be decided if they should be pursued further. A high degree of validation is recommended for protein-protein interactions, because the targets are typically novel in small-molecule discovery research and the initial compounds are often hydrophobic with weak activities and having nonspecific binding. Many approaches have applied to identification of new compounds that can serve as

starting points for further study. Steps which can be used to validate and characterize new PPI inhibitors are listed in Figure 3.1.

Compounds are first identified as hits when they show activity in a primary screen (Figure 3.1). To rule out assay-specific artifacts, usually a secondary functional assay which uses a different method of detection is employed (Figure 3.1). Aggregation or irreversible binding of compounds are also issues that must be addressed during the validation. Qualitative biophysical assays can be used in a screening mode for the direct identification of high-stoichiometry binding or unusual binding kinetics. Surface plasmon resonance (SPR) is a very commonly used method for qualitative and quantitative examination of protein-compound interactions [37, 38] (Figure 3.1).

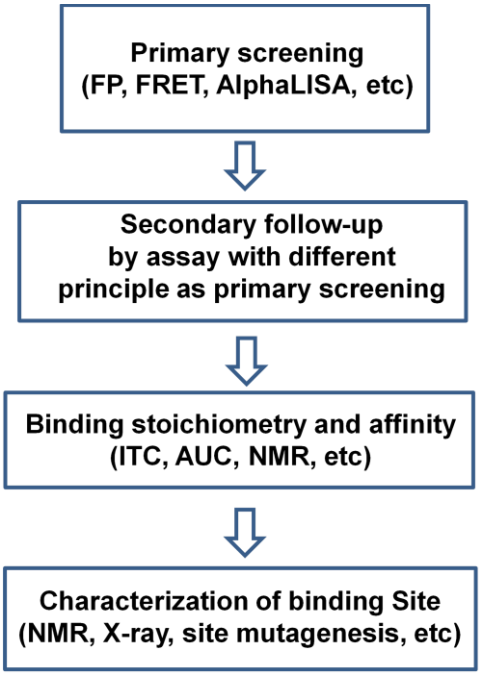


Figure 3.1 Multiple levels of assays for validation and characterization of protein-protein interaction inhibitors.

For new biological systems, it is also useful to measure the stoichiometry of the protein-ligand interaction and to demonstrate that the observed inhibition is related to the binding interaction by relating the IC_{50} and K_d values. A number of biophysical methods are available to address these issues. Isothermal calorimetry (ITC) for

example, provides a wealth of valuable data, including the K_d, the enthalpy and entropy of the interaction and the binding stoichiometry[39].

NMR or X-ray crystallographic approaches provide a highly detailed level of characterization of the protein-compound interaction site. In addition to validating the binding of the small molecule, the binding site characterization can have a major impact on compound optimization and often leads to an understanding of the mechanisms at play.

A compound or series of compounds can be considered to be reasonably well validated when assay-dependent artifacts, aggregation and irreversible binding are ruled out. Biophysical methods, including NMR and X-ray crystallography, further allow the researcher to view the binding event itself, providing strong, positive evidence that a compound is a *bona fide* hit.

3.2.5 Summary

We have characterized the AF9/ENL-DOT1L protein-protein interaction extensively and identified the DOT1L 865-874 10mer peptide as the minimum region essential for the AF9/ENL-DOT1L interaction. It was reasoned that the FP assay might be particularly appropriate to the study of AF9/ENL-DOT1L protein-protein interactions because the interaction of the AF9/ENL protein with DOT1L is mediated by the well-defined short peptide motifs in DOT1L, which can be synthesized and labeled for development using an FP assay.

In this study, we first developed a novel FP-based competitive assay, which is robust, reliable, cost efficient and reproducible. We performed HTS with about 101,000-compounds for small molecule inhibitors of the AF9-DOT1L interaction. Of the 101,000 compounds, 4 were found to inhibit the AF9-DOT1L PPI in both FP and SPR competitive assays and with further synthetic chemistry follow-up, these compounds may lead to inhibitors of the AF9-DOT1L interaction useful for both research and therapeutic purposes.

3.3 Experimental procedures

3.3.1 Protein purification and chemical reagents

The GB1-AF9 (497-568) was expressed in *Escherichia coli* strain BL21 (DE3) (Invitrogen). The medium for bacterial growth in all binding studies was Luria Broth. The proteins were induced with 200 μM IPTG at 16 °C. The cells were harvested after 20 h and resuspended in cold lysis buffer consisting of 50 mM Tris HCl, pH 7.5, 150 mM NaCl, 0.01 % β-mercaptoethanol and purified by affinity chromatography employing Ni-agarose (Qiagen). The proteins were further purified with size exclusion chromatography in 50 mM Tris HCl, pH 7.5, 150 mM NaCl, 3 mM DTT. All purified recombinant proteins were stored at -80 °C for further experiments.

The Triton X-100 was from Sigma (St. Louis, MO). The 10,000 structurally diverse chemical compounds were obtained from ChemBridge (San Diego, CA) as part of the collection of the University of Michigan Center for Chemical Genomics (CCG).

3.3.1.1 Synthesis of fluorescein labeled DOT1L peptides

The fluorescein-labeled peptides N-Flu and C-Flu were synthesized according to the standard Fmoc protocols described in Chapter 2 (2.3.3). For fluorescein labeling, resin-bound peptides were treated with 1 % TFA/DCM to selectively remove the Mtt protective group from the lysine side chain. The

peptide was then treated with solution of 5-carboxyfluorescein (5-FAM) *N*-hydroxysuccinimidyl ester, HATU, HOAt, and DIEA in NMP, and then washed with DCM and MeOH. The reaction mixture was covered in aluminum foil and mixed by being rocked on a Clay Adams Nutator for 2 h. Finally, the peptide was removed from the resin and simultaneously the side-chain was deprotected with 87.5 % TFA, 5 % DTT, 5 % H_2O, and 2.5% TIS at room temperature for 2 h. The crude peptides were purified by semi-preparative reverse-phase high-performance liquid chromatography (RP-HPLC model of HPLC). Peak fractions were concentrated and lyophilized for 16−24 h. The pure peptides were characterized by analytical RP-HPLC and electrospray ionization mass spectrometry (ESI-MS). The purity of all fluorescent peptides was over 98 %, based on their HPLC and mass spectral data.

3.3.2 Determination of the equilibrium dissociation constant (K_d) of Flu-DOT1L peptides with GB1-AF9

Fluorescence polarization experiments were performed in Dynex 96-well, black, round-bottom plates (Fisher Scientific) using a plate reader (Biotek H1 hybrid). To each well, fluorescein-labeled DOT1L peptides (10 nM final) and increasing concentrations of GB1-AF9 protein (from 0 to 24 μM) were added to a final volume of 125 μL in the assay buffer (100 mM Na_2HPO_4, pH 7.5, 150 mM NaCl, 0.01 % Triton X-100). The plate was mixed on a shaker and incubated at room temperature for 3 h to reach equilibrium. The polarization values in milli-

polarization units (mP) were measured at an excitation wavelength at 485 nm and an emission wavelength at 530 nm. To test the assay stability, a plate was measured at different times over a 24 h period. To determine the effect of DMSO on the assay, binding experiments were performed under conditions with or without 4 % DMSO in the assay buffer described above. For the effects of the reducing reagent, binding experiments were performed with or without 0.1 % β-mercaptoethanol. Experimental data were analyzed with Prism 5.0 software (GraphPad Software, San Diego, CA).

3.3.3 Competitive binding experiments

The corresponding unlabeled DOT1L peptides, including DOT1L-16mer, -10mer and -7mer, were tested for their ability to displace the N-Flu DOT1L fluorescent probe from GB1-AF9. Negative controls containing GB1-AF9 and probe (equivalent to 0 % inhibition), and positive controls containing only free N-Flu DOT1L peptide (equivalent to 100 % inhibition), were included on each assay plate. The dose-dependent binding experiments were carried out with serial dilutions of peptides, prepared in DMSO. A 5 µL sample of the tested samples and pre-incubated GB1-AF9 protein (500 nM) and N-Flu DOT1L peptide (10 nM) in the assay buffer were added in 96-well plates to produce a final volume of 125 µL. The polarization values were measured after 3 h incubation and IC_{50} values were determined from the plot by nonlinear least-squares analysis.

The K_i values of competitive inhibitors were calculated using the equation described previously [40] based upon the measured IC_{50} values, the K_d value of the probe and GB1-AF9 complex, and the concentrations of the protein and probe in the competitive assay.

The SPR based competitive assay is described in Chapter 2.

3.3.4 High-Throughput Screening of AF9-DOT1L interaction

Using a Multidrop 384 (Thermo Scientific, Waltham, MA), assay buffer was added to each well of a low-volume black 384-well plate (Cat. 3676, Corning, NY). The assay buffer (10 µL) was added first to the wells then, using a 384-well pin tool on a Biomed FX liquid-handling workstation (Beckman Coulter, Fullerton, CA) 200 nL of compounds (stock concentration 2 mM) or DMSO for control wells, were added, producing a final compound concentration of 20 µM and 1 % DMSO. 10 µL of 2X GB1-AF9 and Flu-DOT1L complex, producing final concentrations of 500 nM GB1-AF9 and 10 nM Flu-DOT1 were added to the wells by Multidrop. Plates were incubated at room temperature for 1 h, then read for fluorescence polarization with a PHERAstar (BMG LabTech, Offenburg, Germany) plate reader with wavelengths of 485 nm for excitation, and 525 nm for emission. For follow-up dose-response studies of the primary hits, the same method as for the primary screen was performed, except that compounds were first added to the wells of the 384-well plate using serial of 1.67-fold dilution over a range of

concentrations between 150 µM and 4.2 µM, and the buffer and 2X complex were added to the wells. The chemical structures of confirmed hits were analyzed and clustered for which a 65% or greater similarity was used for their clustering by the software Tripos BenchWare DataMiner.

3.3.5 Data calculations

To evaluate the quality and suitability of the AF9-Dot1L fluorescence polarization assay for high-throughput screening, we determined the Z' factor, which is an indicator of the viability of the assay for screening, by incorporating the precision of the assay [41]:

$$Z' = 1 - \frac{(3SDf + 3SDd)}{(\mu b - \mu f)}$$

where SD_f and SD_d are the standard deviation of the signal (mP) for free and bound probe, respectively, μ_b represents the mean of the signal obtained for the bound probe in the absence of a competitive inhibitor, and μ_f is the mean of the free probe in the absence of the GB1-AF9 protein.

3.4 Results

3.4.1 Assay principles

Following previous studies, we devised a homogeneous fluorescence polarization assay based on binding of recombinant soluble GB1-AF9 protein to a synthetic DOT1L 10mer peptide conjugated with a fluorescein (Flu-DOT1L) (Figure 3.2A).

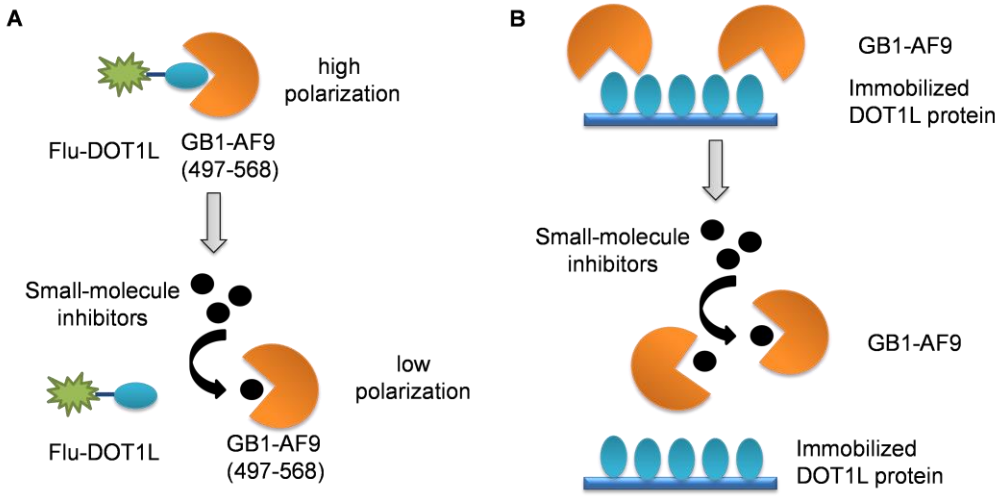


Figure 3.2 Schematic illustration of the Fluorescence Polarization and Surface plasmon Resonance (SPR) based competitive assay. (A) FP competitive assay (B). SPR competitive assay.

In this FP assay, the rotation of Flu-DOT1L peptide slows down upon binding to the GB1-AF9 protein, thus affecting the polarization of the Flu-DOT1L peptide. Compounds that displace this Flu-DOT1L peptide from the AF9 protein cause a change in the FP of the sample, and this change is used to quantify the binding. The FP assay can be used in a high-throughput format to identify small molecule leads that inhibit the AF9-DOT1L interaction.

However, fluorescence from the compounds library can cause artifacts and accordingly, surface plasmon resonance (SPR) assay was established as a secondary approach to confirm specificity of the interruption of AF9-DOT1L interaction (Figure 3.2B). The DOT1L protein was immobilized on a sensor chip and AF9 protein flowing in the mobile phase over the protein immobilized protein gives a high refractive signal (Figure 3.2B). When small molecules bind to the AF9 protein and block the binding of AF9 protein with DOT1L, the refractive signal will be reduced. It had been shown that the DOT1L 10mer peptide competes readily with the immobilized DOT1L protein for binding to AF9 protein with IC_{50} = 0.49 μM (Figure 2.6A), and this indicates that the SPR-based competitive assay will be an effective method to gain independent confirmation of the specificity of compounds identified by the fluorescence polarization assay.

3.4.2 Design, synthesis and characterization of fluorescein DOT1L peptides

We synthesized two different fluorescent probes using fluorescein as a fluorophore (Figure 3.3): the N-terminal fluorescein-labeled DOT1L peptide with the sequence Ac-K(FAM)(β)A(β)ALPISIPLSTV-NH_2 termed N-Flu-DOT1Ll; the C-terminal DOT1L fluorescein labeled DOT1L peptide with the sequence Ac-LPISIPLSTV(β)A(β)AK(FAM)-NH_2. These two fluorescent probes were first tested in a saturation binding experiment to determine their binding affinity to AF9 C-terminal domain (Figure 3.4).

The dissociation constant (K_d) for each protein/peptide pair was determined using a constant concentration of probe and titrating with the GB1-AF9 (497-568) protein. The two tested probes showed binding and the FP values of the peptides increased as a function of GB1-AF9 protein concentration. The K_d of binding between the N-Flu-DOT1Ll probe and GB1-AF9 was determined to be 46 nM with a dynamic range of 128 mP. The C-Flu-DOT1L peptide showed a slightly decreased binding affinity as the N-Flu-DOT1L peptide, with a K_d of 141 nM, and a much narrower dynamic range of 80 mP. Based on the information obtained in saturation experiments, the K_d values and dynamic binding ranges, we chose to use the

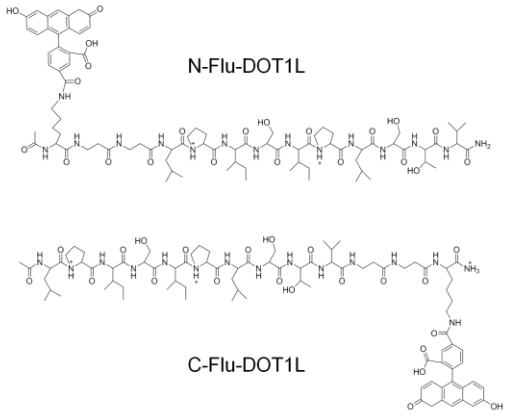


Figure 3.3 Structure of two fluorescein labeled DOT1L peptides.

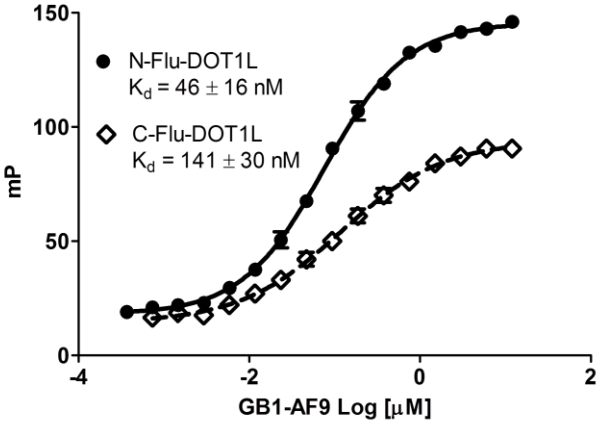


Figure 3.4 Saturation of GB1-AF9 with different fluorescent probes. The fluorescein DOT1L peptides is 10 nM and GB1-AF9 is in the range of 0 to 24 μM.

N-Flu-DOT1L peptide as the probe for further development and optimization of the FP-based assay for GB1-AF9.

To demonstrate that the FP is independent of the total fluorescence, N-Flu-DOT1L (2.5 – 80 nM) was titrated against a constant GB1-AF9 concentration (500 nM) (Figure 3.5A).

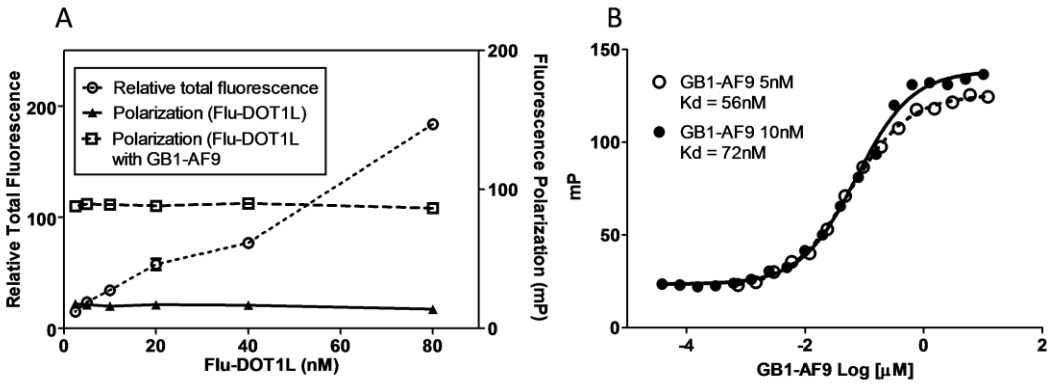


Figure 3.5 Effect of probe peptide concentration on assay performance. (A) Fluorescence polarization and total fluorescence of different concentration of Flu-DOT1L probe (2.5 – 80 nM) alone or in complex with GB1-AF9 (500 nM). (B) Saturation of GB1-AF9 protein with different concentration of Flu-DOT1L probe (5 nM and 10 nM).

The data summarized in Figure 3.5A show a linear increase in total fluorescence with increasing concentration of N-Flu-DOT1L, whereas the FP signal with N-Flu-DOT1L peptide alone, or N-Flu-DOT1L peptide in complex with the GB1-AF9 protein, remain constant. Additionally, saturation of the GB1-AF9 protein with different concentrations of the N-Flu-DOT1L peptide (5 nM and 10 nM) gives similar K_d values (56 nM and 72 nM respectively) (Figure 3.5B). It was concluded that the FP signal is independent of the probe concentration and a 10

nM concentration of the Flu-DOT1L peptide was selected for the FP assay optimization.

3.4.3 Stability and specificity of the FP assay

The stability of the AF9 FP assay, critical for high-throughput screening, was tested by incubating the plate at room temperature over a 24 h period, reading the plate several times, and analyzing the data. The results obtained showed that the assay is stable, as evidenced by the binding curves (Figure 3.6A). The obtained K_d values and the binding ranges remained unchanged over a 24-hour time period. We also tested the influence of DMSO, a commonly used solvent, in high-throughput screening. The results obtained showed that the binding affinity and dynamic range of N-Flu-DOT1L probe are unchanged in the presence of 4 % DMSO (Figure 3.6B) The DMSO tolerance renders this assay suitable for high-concentration (*i.e.* 100 μM) screening campaigns where 4 % DMSO is necessary to help minimize solubility issues of library compounds. The effects of reducing reagent were also tested, as shown in Figure 3.6C. In the presence of 0.1 % β-mercaptoethanol, the binding affinities of N-Flu-DOT1L peptide to GB1-AF9 are not affected. We also test the specificity of the assay by titrating GB1 tag protein with N-Flu-DOT1L peptide (Figure 3.6D). The GB1 tag failed to bind to the DOT1L peptide, indicating that the FP assay is specific and unaffected by the fusion tag.

All the above tests indicate that fluorescence polarization assay of GB1-AF9 and the N-Flu-DOT1L peptide is stable, specific and robust, and adequate for high-throughput screening.

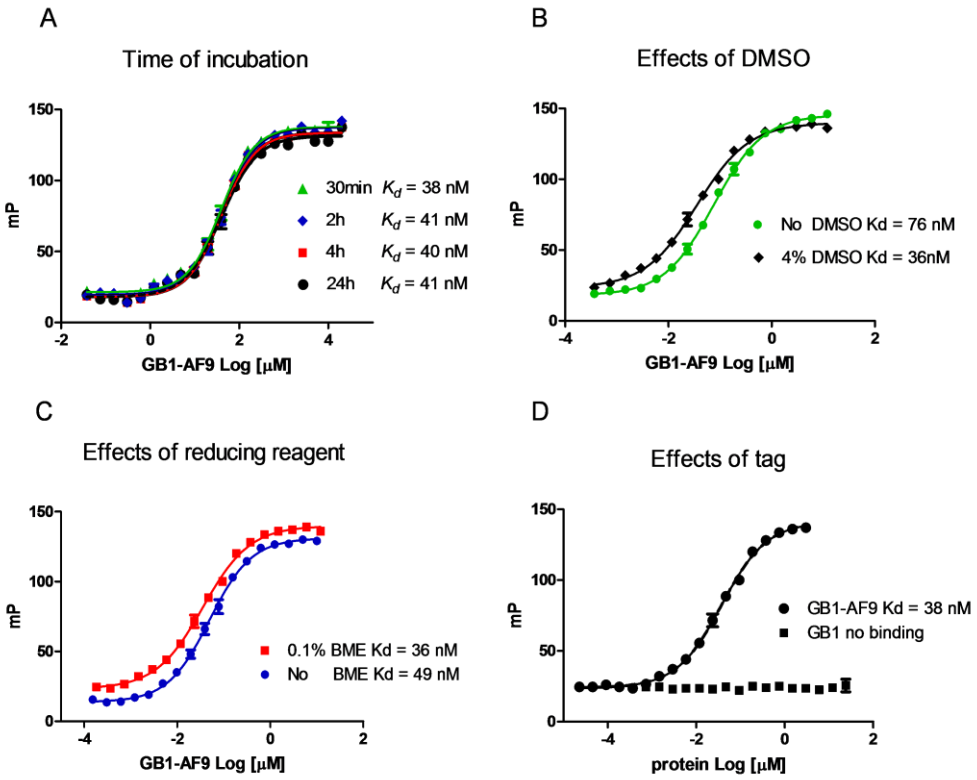


Figure 3.6 Stability and specificity of GB1-AF9 FP assay. (A) Saturation experiment with N-Flu-DOT1L to monitor the Kd and dynamic range over an extended time period. N-Flu-DOT1L (5 nM) was titrated with GB1-AF9 (0.04 nM 24 μM) in assay buffer (100 mM NaH2PO4, pH 7.4, 150mM NaCl, 0.01% Triton X-100) containing 4 % DMSO and the plate was read at the indicated time points. (B) Saturation experiments with N-Flu-DOT1L to determine the effect of DMSO. Experiments were performed as in (A) in the presence of 0 or 4 % DMSO. (C) Saturation experiments with N-Flu-DOT1L to determine the effect of adding detergent to the assay buffer. Saturation experiments were performed as in (A) using assay buffers containing 0, 0.1 % β-ME in the presence of 4 % DMSO. (D) Saturation experiments of N-Flu-DOT1L with GB1-AF9 or GB1 tag to determine the effect fusion tag on the binding assay. Saturation experiments were performed as in (A) using assay buffer containing 4 % DMSO.

3.4.4 Validation of the FP assay with known DOT1L peptide antagonists

An FP-based competitive assay is used to screen compounds and typically uses standard concentrations of protein and fluorescence probes, with various concentrations of inhibitors. The concentrations of the fluorescence tracer and the protein used in the FP-based assay should be chosen carefully with a view to maximizing the difference between the highest and lowest polarization values and increasing the sensitivity of the assay. FP competition experiments are best designed such that the [receptor]/K_d ratio is at least 1, and the starting polarization value represents approximately 50 % of the maximal FP change observed in the saturation experiment. We chose to use a concentration of 10 nM for the probe. This is a concentration which has high fluorescence intensity and can overcome the potential interference of any weakly fluorescent compounds. To determine the optimal concentration of the protein, we evaluated three different concentrations (250, 500 and 1000 nM) of the GB1-AF9 with a fixed concentration (10 nM) of N-Flu-DOT1L in the competitive binding assay to determine the IC_{50} value and K_i of the DOT1L 10mer peptide (Figure 3.7A).

Using 250 nM, 500 nM and 1000 nM (5-, 10- or 20- fold the K_d), as the protein concentration, the dynamic ranges of the assay were 75ΔmP, 87ΔmP and 100ΔmP, respectively (Figure 3.7A). With higher protein concentrations, the IC_{50} values obtained for the competitors were increased. There is a good correlation between the observed IC_{50} values and concentrations of the protein used in the

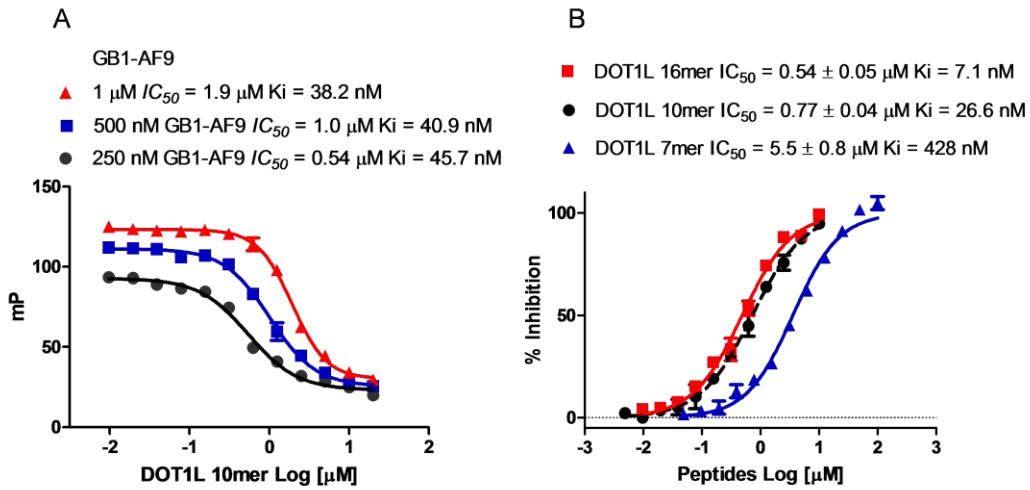


Figure 3.7 Optimizaion of FP competitive assays. (A) Displacement of N-Flu Dot1l probe from GB1-AF9 by DOT1L 16mer peptide using three protein concentrations. Increasing protein concentration increases the dynamic range of the assay but also increases the IC50. (B) Test of different DOT1L peptides by competitve assay with 500nM GB1-AF9 protein and 10 nM Flu-DOT1L peptide and various concentrations of peptides. The IC50 and Ki values of each peptides are indicated.

assay (Figure 3.7A). The IC_{50} values obtained for these peptides are significantly higher than the K_d values of corresponding fluorescent labeled peptide tracers, but the rank order is the same under all conditions. The K_i values for the DOT1L 10mer under different concentrations of GB1-AF9 proteins were calculated as reported previously [40]. As expected, K_i values of the DOT1L 10mer peptides

are similar with these different GB1-AF9 protein concentrations, and are close to the K_d of the Flu-DOT1L with the protein (Figure 3.7A). We chose 500 nM as the concentration of the GB1-AF9 protein for the competitive binding assay because this is about 10-fold of the actual K_d value (Figure 3.4) and at this concentration a majority of the fluorescent probe will be bound to the protein.

To further validate the optimized FP-based competitive binding assay conditions, we tested three corresponding unlabeled peptides: DOT1L 16mer, DOT1L 10mer and DOT1L 7mer, with different binding affinities to the AF9 protein (Figure3.7B). The K_i values for these peptides were calculated and as expected, DOT1L 16mer and DOT1L 10mer peptide have similar binding affinities (K_i = 7.6 nM for DOT1L 16mer and 26.6 nM for DOT1L 10mer). DOT1L 7mer has weak binding affinity (K_i = 428 nM).

These validation experiments provide evidence that the FP competitive binding assay will quantitatively and accurately determine the binding affinities of small-molecule inhibitors with different binding affinities to the GB1-AF9 protein. The optimized 96-well format assay is robust and sensitive (Z' factor of 0.79) and suitable for using it in the high-throughput screening.

3.4.5 High-throughput format development and miniaturization

The simple one-step “mix-and-measure” format of the AF9 and DOT1L FP assay and its demonstrated sensitivity allow this method to be used for automated operation in HTS approach for identifying small-molecule inhibitors of this PPI. In order to use the FP based optimized assay, the reaction volume has to be scaled down for performing the screening in low volume 384-well plates, a total volume of 20 µL. To test the utility of this HTS format, we examined the precision and robustness of the assay (Figure 3.8A).

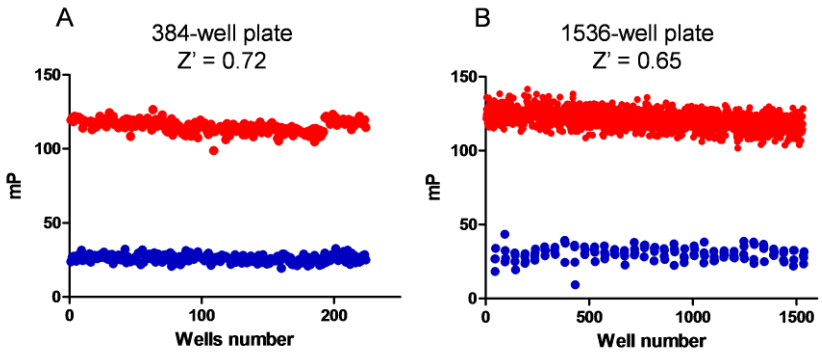


Figure 3.8 Calculated Z’ factor in (A) 384-well and (B) 1536-well plate. The polarization signal of the free Flu-DOT1L peptide (10 nM) is in blue and the bound peptide (500 nM GB1-AF9 and 10 nM Flu-DOT1L) is in red.

For this purpose, each microplate contained an identical set of control reactions as the following: 1) Positive control represented by free Flu-DOT1L peptide (10 nM) which give minimal mP signal that corresponds to 100% inhibition, 2) Flu-DOT1L peptide (10 nM) in complex with GB1-AF9 (500 nM) giving the maximum signal and represents negative control for 0 % inhibition.

This test also provides an experimental basis for setting the operational threshold to identify potential positives in the primary screenings. One of the parameters for determining the quality of a high-throughput assay is the Z' factor, a statistical parameter that assesses the performance of HTS assays [42]. The Z' factor is reflective of both the assay dynamic range and the data variation. Assays with small Z' factors are not suitable for high-throughput screening and require further optimization, while assays with Z' factors close to the maximum value of 1 are of high quality. The calculated Z′ factor of the FP assay in 384-well format is 0.72, demonstrating a robust and consistent assay with minimal variability (Figure 3.8A). Thus, the FP assay for AF9 binding based on Flu-DOT1L peptide is of excellent quality for HTS.

The ability to perform primary screening assays in high-density micro-well plates at low volumes (1-5 µL) is important to accelerate the early stages of drug discovery and reduce costs. The AF9-DOT1L FP assay was successfully miniaturized to 1536-well format using 4 µL total volumes, significantly reducing reagent consumption, without compromising assay reliability. Z' factor value determined for the FP assay in 1536-well format was found to be 0.65 (Figure 3.8B), confirming the assay to be robust, reliable, and suitable for HTS purposes.

3.4.6 HTS for identification inhibitors of the AF9/DOT1L protein-protein interaction

Using our optimized AF9-competitive FP binding assay and DOT1L-10mer peptide as the fluorescent probe, in-house library of 101,000 structurally diverse small molecules were screened at the Center for Chemical Genomics at the University of Michigan. Primary screening of these compounds at a single concentration of 20 µM identified 1,191 potential active compounds with inhibition larger than 3 standard deviations above the negative control giving a hit rate of 1.2 % (Figure 3.9A and C).

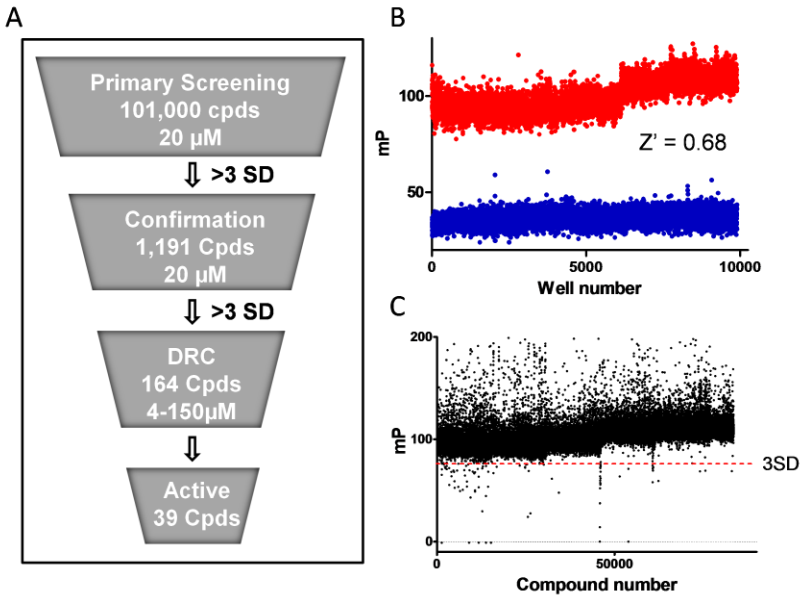


Figure 3.9 Fluorescence polarization screen for inhibitors of AF9-DOT1L interaction. (A) Schematic of the HTS process.(B) Control samples run on each plate include 10 nM Flu-DOT1L and 500nM GB1-AF9 complex as a negative (0 % inhibition) control (•) and Flu-DOT1L only as a positive (100 % inhibition) control (•). The statistical Z' Factor measurement for the HTS was calculated to be 0.68 for this entire screen indicating suitable reproducibility. (C) Data from the 101,000 compound HTS are shown (•) with a 3 standard deviation line indicated in red.

The Z' factor for the primary screen was determined to be 0.68 (Figure 3.9B), the signal-to-noise was 19 and the coefficient of variation value was 0.08, indicating the quality and reproducibility of the used FP assay.

The 1191 compounds were further tested using the same conditions as in the primary screen and 164 compounds were confirmed, after removing the autofluorescent compounds, quenchers and compounds with reactive functional groups (14 % confirmation rate, 0.16 % overall). These 164 compounds were chosen for dose-response confirmation using concentration range from 4-150 µM with compound stocks directly from the mother plates. Thirty nine compounds demonstrated dose-dependent inhibition in the FP assay.

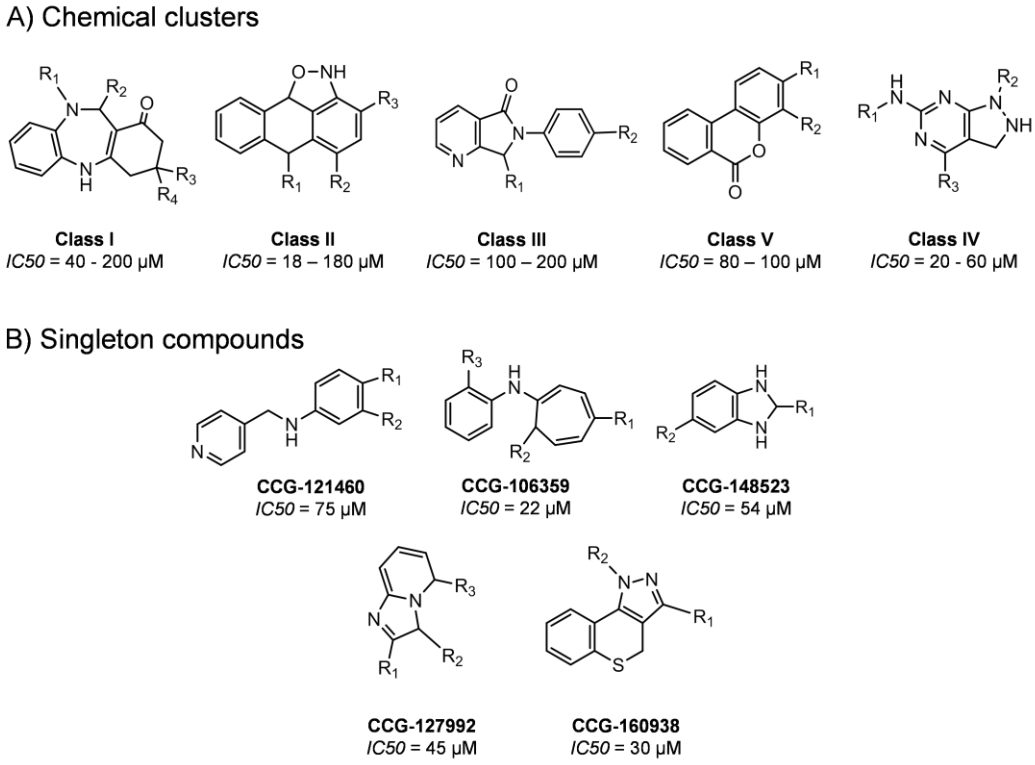


Figure 3.10 HTS Identified compounds structures. (A) Five classes of compounds with different scaffolds. (B) Structures of identified singleton compounds.

The analysis of the chemical structures of these 39 compounds showed that they can be classified into five clusters by Tripos BenchWare DataMiner with different chemical scaffolds and the rest of the compounds have unique structures (Figure 3.10).

3.4.7 Validation of HTS identified compounds

One of the most important steps after performing HTS approach for identifying small-molecule modulator of protein function is the follow-up of the obtained hits and their characterization and validation into high-content lead series.

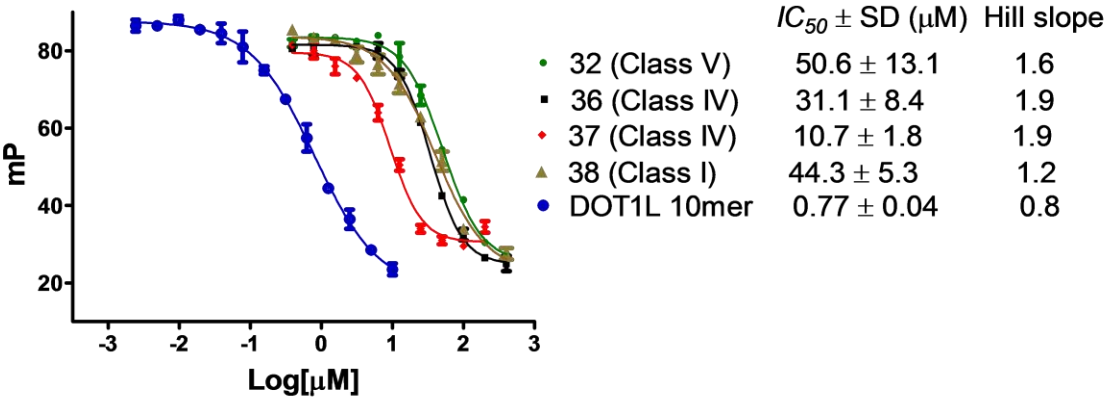


Figure 3.11 Dose-response FP competitive binding curves of active compounds.

Based on the potency, chemical structure and availability of these 39 compounds, 8 compounds were purchased and new stock solutions were prepared. Their chemical structures were confirmed by NMR or LC-MS and re-tested using the FP based binding assay. Four of these compounds, belonging to

the chemical clusters I, IV and V, demonstrated reproducible dose-dependent inhibition against AF9-DOT1L interaction with IC_{50} from 10 – 50 μM (Figure 3.11), confirming the results obtained in HTS. The Hill slopes of competitive curve for these four compounds are between 1-2.

One compound did not show any binding and was not confirmed as a binder. The other three compounds showed higher fluorescence of the Flu-DOT1L-10mer peptide used in the assay, making them not appropriate to be tested and determined their binding affinity with FP assay because of the potential interference with the assay.

3.4.8 Orthogonal testing with biophysical methods

In order to further validate the binding of the four confirmed compounds and identify potential promiscuous inhibitors, several secondary assays have been developed to support hit confirmation and the hit-to-probe optimization process. Using the FP based assay, we showed that the compounds were able to compete off the fluorescent-labeled DOT1L peptide from AF9 protein. In order to determine that the compounds bind effectively to the target and disrupt protein-protein interaction, SPR competitive binding assay was used. This assay use different platform that the FP based assay where DOT1L protein is immobilized on CM5 chip and allows testing the inhibitors for their ability to disrupt AF9-DOT1L interaction. The four compounds confirmed in FP were

further validated and showed similar potency in disrupting the PPI with IC_{50} from 15 to 79 µM (Figure 3.12) as in the FP based assay.

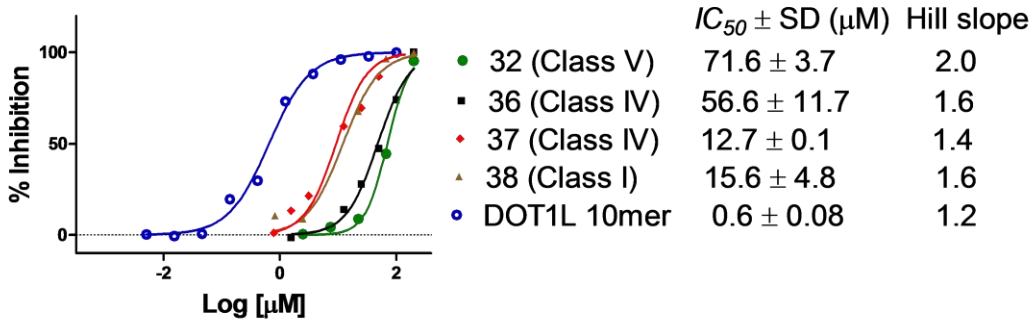


Figure 3.12 Dose-response SPR competitive binding curves of active compounds

NMR spectroscopy has been successfully utilized to detect binding of small-molecule compounds to large biomolecules and in the last decade many new NMR-based screening methods have been established to characterize the binding processes. Therefore, confirmed hit compounds in two different binding assays, FP and SPR, were further tested using Saturation Transfer Difference (STD) NMR experiments [43]. The purpose of this assay is to provide further validation of the hit selection in addition to the biochemical assays by verification of direct binding of the compounds to AF9 protein. This technique is the most direct measure of inhibiting the AF9-DOT1L interaction by a small molecule and is allowing characterization of the binding interactions at an atom level, termed group epitope mapping (GEM). A strong STD effect was observed for compound

38 and weaker STD signals for compound 36, indicative of their direct binding to AF9 (Figure 3.13). Compounds **32** and **37** were also tested but because of their limited solubility we were not able to perform this experiment.

Overall these data demonstrate the feasibility of development of small molecules targeting C-terminal hydrophobic domain in AF9, indicating it is a “druggable” target. These data demonstrate that our FP-based assay is suitable for high-throughput screening and for the first time, we have shown that it is possible to discover non-peptide small-molecules that can block the interaction between DOT1L and AF9. This is an important initial proof-of-concept for the development of small molecules targeting C-terminal hydrophobic domain in AF9, indicating it is a “druggable” target.

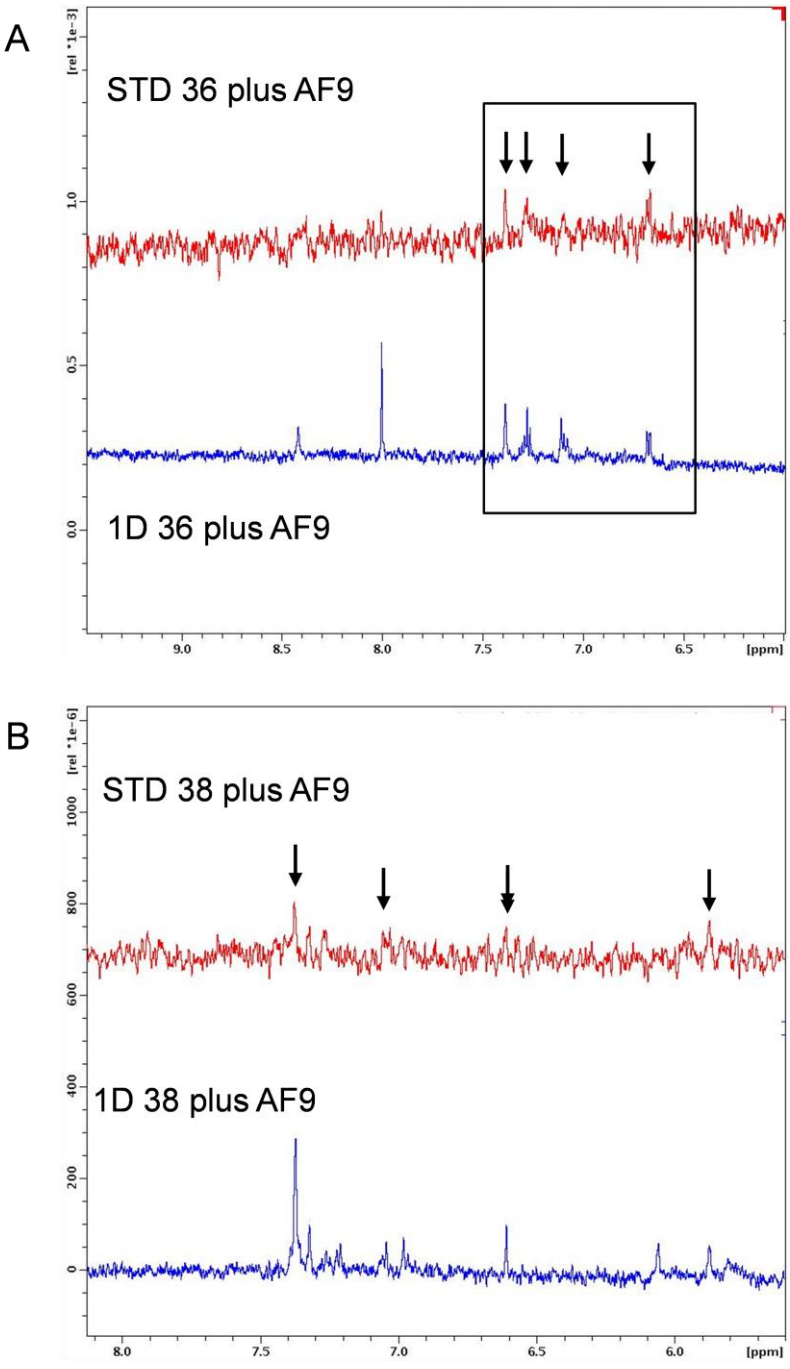


Figure 3.13 STD NMR experiments with compounds 36 (A) and 38 (B). Top spectrum represents 1D STD spectrum of tested compound (200 μM) in complex with AF9 (5 μM) and the bottom spectrum is 1D NMR spectrum.

3.5 Discussion

First, we synthesized two fluorescein labeled DOT1L peptides and tested the binding of these peptides with GB1-AF9 protein (Figure 3.4). The N-terminal fluorescein labeled DOT1L peptide showed better dynamic range and binding affinity compared to the C-terminal fluorescein labeled peptide, and was used for the FP assay development. The binding affinity for the GB1-AF9 and Flu-DOT1L peptide was 42 ± 16 nM (Figure 3.4), which is consistent with the Kd between AF9 and DOT1L protein determined by SPR (Table 2.1). These results demonstrate that the AF9-DOT1L protein interaction could accurately be detected and quantitated by FP assay. Second, the competition binding experiment results demonstrated that the Flu-DOT1L peptide could be displaced by the unlabeled DOT1L peptides (Figure 3.7). This was important as it demonstrates specificity of the Flu-DOT1L interaction with GB1-AF9 and also showed that the assay can be used to search for compounds that disrupt AF9-DOTL1L interaction. Third, because of its excellent stability, DMSO tolerance, and simplicity, in 384-well plate format, the developed FP assay has a Z' factor of 0.72 (Figure 3.8A), making this FP assay is particularly suited for HTS. Furthermore, the assay was miniaturized to a volume of 5 µL with excellent performance (Figure 3.8B), suggesting a possibility of adapting it to a 1536-well plate format for ultra-HTS of small-molecule AF9-DOT1L interaction inhibitors.

Using our HTS optimized FP-competitive assay, we screened of 10,1000 compounds at the Center for Chemical Genomics at the University of Michigan. Our first round of screening identified 1191 potential active compounds. After confirmation screening and dose-response experiments we were left with 39 compounds that required further follow up. Eight compounds were obtained from commercial sources and four showed confirmed dose-dependent inhibition in our FP and SPR competitive assay (Figure 3.11 and 3.12). Of these four compounds, CCG-186912 (IC_{50} = 12.1 μM) and CCG-186911 (IC_{50} = 31.1 μM) belong to the same class and both demonstrated dose-dependent inhibition in both our FP assay and SPR-based secondary assay (Figure 3.11 and 3.12).

3.6 References

1. Arkin, M., *Protein-protein interactions and cancer: small molecules going in for the kill.* Curr Opin Chem Biol, 2005. **9**(3): p. 317-24.
2. White, A.W., A.D. Westwell, and G. Brahemi, *Protein-protein interactions as targets for small-molecule therapeutics in cancer.* Expert Rev Mol Med, 2008. **10**: p. e8.
3. Blazer, L.L. and R.R. Neubig, *Small molecule protein-protein interaction inhibitors as CNS therapeutic agents: current progress and future hurdles.* Neuropsychopharmacology, 2009. **34**(1): p. 126-41.
4. Whitty, A. and G. Kumaravel, *Between a rock and a hard place?* Nat Chem Biol, 2006. **2**(3): p. 112-8.
5. Wells, J.A. and C.L. McClendon, *Reaching for high-hanging fruit in drug discovery at protein-protein interfaces.* Nature, 2007. **450**(7172): p. 1001-9.
6. Jones, S. and J.M. Thornton, *Principles of protein-protein interactions.* Proc Natl Acad Sci U S A, 1996. **93**(1): p. 13-20.
7. Lo Conte, L., C. Chothia, and J. Janin, *The atomic structure of protein-protein recognition sites.* J Mol Biol, 1999. **285**(5): p. 2177-98.
8. Cheng, A.C., et al., *Structure-based maximal affinity model predicts small-molecule druggability.* Nat Biotechnol, 2007. **25**(1): p. 71-5.
9. Smith, R.D., et al., *Exploring protein-ligand recognition with Binding MOAD.* J Mol Graph Model, 2006. **24**(6): p. 414-25.
10. Hopkins, A.L. and C.R. Groom, *The druggable genome.* Nat Rev Drug Discov, 2002. **1**(9): p. 727-30.
11. Thompson, A.D., et al., *Fine-tuning multiprotein complexes using small molecules.* ACS Chem Biol, 2012. **7**(8): p. 1311-20.
12. Clackson, T. and J.A. Wells, *A hot spot of binding energy in a hormone-receptor interface.* Science, 1995. **267**(5196): p. 383-6.
13. Clackson, T., et al., *Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity.* J Mol Biol, 1998. **277**(5): p. 1111-28.
14. Thanos, C.D., W.L. DeLano, and J.A. Wells, *Hot-spot mimicry of a cytokine receptor by a small molecule.* Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15422-7.
15. Moreira, I.S., P.A. Fernandes, and M.J. Ramos, *Hot spots--a review of the protein-protein interface determinant amino-acid residues.* Proteins, 2007. **68**(4): p. 803-12.
16. Bogan, A.A. and K.S. Thorn, *Anatomy of hot spots in protein interfaces.* J Mol Biol, 1998. **280**(1): p. 1-9.
17. DeLano, W.L., *Unraveling hot spots in binding interfaces: progress and challenges.* Curr Opin Struct Biol, 2002. **12**(1): p. 14-20.

18. Ma, B., et al., *Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces.* Proc Natl Acad Sci U S A, 2003. **100**(10): p. 5772-7.
19. Arkin, M.R. and J.A. Wells, *Small-molecule inhibitors of protein-protein interactions: progressing towards the dream.* Nat Rev Drug Discov, 2004. **3**(4): p. 301-17.
20. Wilson, A.J., *Inhibition of protein-protein interactions using designed molecules.* Chem Soc Rev, 2009. **38**(12): p. 3289-300.
21. Vagner, J., H. Qu, and V.J. Hruby, *Peptidomimetics, a synthetic tool of drug discovery.* Curr Opin Chem Biol, 2008. **12**(3): p. 292-6.
22. Karatas, H., et al., *High-affinity, small-molecule peptidomimetic inhibitors of MLL1/WDR5 protein-protein interaction.* J Am Chem Soc, 2013. **135**(2): p. 669-82.
23. Mayr, L.M. and P. Fuerst, *The future of high-throughput screening.* J Biomol Screen, 2008. **13**(6): p. 443-8.
24. Degterev, A., et al., *Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-xL.* Nat Cell Biol, 2001. **3**(2): p. 173-82.
25. Vassilev, L.T., et al., *In vivo activation of the p53 pathway by small-molecule antagonists of MDM2.* Science, 2004. **303**(5659): p. 844-8.
26. Oltersdorf, T., et al., *An inhibitor of Bcl-2 family proteins induces regression of solid tumours.* Nature, 2005. **435**(7042): p. 677-81.
27. Betzi, S., et al., *Protein protein interaction inhibition (2P2I) combining high throughput and virtual screening: Application to the HIV-1 Nef protein.* Proc Natl Acad Sci U S A, 2007. **104**(49): p. 19256-61.
28. Arkin, M.R., et al., *Inhibition of Protein-Protein Interactions: Non-Cellular Assay Formats*, in *Assay Guidance Manual*, G.S. Sittampalam, et al., Editors. 2004: Bethesda (MD).
29. Checovich, W.J., R.E. Bolger, and T. Burke, *Fluorescence polarization--a new tool for cell and molecular biology.* Nature, 1995. **375**(6528): p. 254-6.
30. Jameson, D.M. and W.H. Sawyer, *Fluorescence anisotropy applied to biomolecular interactions.* Methods Enzymol, 1995. **246**: p. 283-300.
31. Heyduk, T., et al., *Fluorescence anisotropy: rapid, quantitative assay for protein-DNA and protein-protein interaction.* Methods Enzymol, 1996. **274**: p. 492-503.
32. Lundblad, J.R., M. Laurance, and R.H. Goodman, *Fluorescence polarization analysis of protein-DNA and protein-protein interactions.* Mol Endocrinol, 1996. **10**(6): p. 607-12.
33. Knight, S.M., et al., *A fluorescence polarization assay for the identification of inhibitors of the p53-DM2 protein-protein interaction.* Anal Biochem, 2002. **300**(2): p. 230-6.

34. Glover, C.J., et al., *A high-throughput screen for identification of molecular mimics of Smac/DIABLO utilizing a fluorescence polarization assay.* Anal Biochem, 2003. **320**(2): p. 157-69.
35. Kenny, C.H., et al., *Development of a fluorescence polarization assay to screen for inhibitors of the FtsZ/ZipA interaction.* Anal Biochem, 2003. **323**(2): p. 224-33.
36. Roehrl, M.H., J.Y. Wang, and G. Wagner, *Discovery of small-molecule inhibitors of the NFAT--calcineurin interaction by competitive high-throughput fluorescence polarization screening.* Biochemistry, 2004. **43**(51): p. 16067-75.
37. Rich, R.L. and D.G. Myszka, *Survey of the year 2000 commercial optical biosensor literature.* J Mol Recognit, 2001. **14**(5): p. 273-94.
38. Day, Y.S. and D.G. Myszka, *Characterizing a drug's primary binding site on albumin.* J Pharm Sci, 2003. **92**(2): p. 333-43.
39. Leavitt, S. and E. Freire, *Direct measurement of protein binding energetics by isothermal titration calorimetry.* Curr Opin Struct Biol, 2001. **11**(5): p. 560-6.
40. Nikolovska-Coleska, Z., et al., *Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization.* Anal Biochem, 2004. **332**(2): p. 261-73.
41. Blevitt, J.M., et al., *A Fluorescence-Based High Throughput Screen for the Transporter Associated with Antigen Processing.* J Biomol Screen, 1999. **4**(2): p. 87-91.
42. Pope, A.J., U.M. Haupts, and K.J. Moore, *Homogeneous fluorescence readouts for miniaturized high-throughput screening: theory and practice.* Drug Discov Today, 1999. **4**(8): p. 350-362.
43. Mayer, M. and B. Meyer, *Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor.* J Am Chem Soc, 2001. **123**(25): p. 6108-17.

Chapter 4

Conclusions and Future Directions

4.1 MLL fusion oncogenes and DOT1L

Over the past 20 years, MLL has received much attention and has been extensively studied for its pathological mechanism in leukemia development. Leukemias bearing rearrangements of the *MLL* gene are of particular interest because of their unique clinical and biological characteristics and poor prognostic. MLL translocations found in leukemia generate fusions that encode an N-terminal fragment of MLL linked to a C-terminal fragment of various fusion partners [1]. The translocation partners are highly diverse and more than 60 known fusion proteins of MLL have been identified. The most frequent translocations generate the MLL-AF9, MLL-ENL, MLL-AF4, MLL-AF10 and MLL-AF6 fusion proteins in leukemias. The extensive studies in past several years have clearly demonstrated that fusions of MLL induce widespread epigenetic dysregulation that may mediate their transforming activity. The histone methyltransferase DOT1L, which methylates histone 3 on lysine 79 (H3K79), is one of the identified binding partners of MLL-fusion oncogenes [2-5]. It was demonstrated that

transformation in MLL-rearranged leukemias is driven by a DOT1L dependent, aberrant epigenetic program. [6]. MLL fusion proteins, including AF9, ENL, AF10 and AF17 have been shown to physically interact with DOT1L and recruit DOT1L to MLL target downstream genes. A suggested general mechanism by which DOT1L contributes to the leukemogenesis process is shown in Figure 1.3 and can be described as interaction between DOT1L and MLL fusion partners leading to mistargeting of DOT1L to the gene targets of MLL fusion proteins, such as the *Hoxa* cluster and Meis1. Aberrant hypermethylation of H3K79 leads to constitutive transcriptional activation of these genes, which in turn results in leukemic transformation.

DOT1L is a validated therapeutic target and its crucial role in *MLL*-rearranged leukemias is now well established [7-9]. Proof-of-concept studies have recently been reported, using small-molecule inhibitors targeting the catalytic domain of DOT1L [9]. EPZ004777, a potent selective inhibitor of DOT1L, blocks cellular H3K79 methylation, inhibits leukemogenic gene expression, and selectively kills cultured cells bearing *MLL* translocations. However global inhibition of DOT1L function could have major deleterious effects in normal hematopoietic stem cells (HSCs) and other tissues, in particular cardiac function.

Based on the reported evidences that DOT1L recruitment by fusion proteins is pivotal for leukemogenesis induced by MLL fusion oncogenes, we hypothesize that targeting the AF9/ENL-DOT1L protein-protein interaction represents an attractive and promising therapeutic strategy for several reasons. First, the interactions between MLL fusion protein and DOT1L were shown to be crucial for MLL transformation, since disruption of the interaction by mutations or deletion abolished transformation [2, 3, 10]. Second, since DOT1L has been involved in multiple biological processes [6], targeting the enzymatic activity of DOT1L might cause hematopoietic and other toxicities. However, targeting AF9/ENL-DOT1L PPI is providing selective inhibition of the DOT1L activity and might avoid the side effects caused by global inhibition of DOT1L enzymatic activity. Third, as demonstrated by our work, DOT1L binds to AF9/ENL with a binding affinity of nanomolar range (Figure 2.4). AF9/ENL-DOT1L interaction is primarily mediated by a 10 amino acids region in DOT1L (Figure 2.6), and four hydrophobic residues are most crucial for the interaction (Figure 2.8), making the design of small molecule inhibitors a plausible reality. Development of small molecule inhibitors of the PPI between DOT1L and MLL-fusion proteins, AF9 and ENL, will help to address some of the mechanistic questions about the molecular interactions between MLL fusions, DOT1L and chromatin modifications.

4.2 Summary and conclusions of our work

4.2.1 Characterization of AF9/ENL-DOT1L protein-protein interaction

In Chapter 2, we have initiated the efforts of characterization of the protein-protein interaction between DOT1L and MLL-fusion proteins, AF9 and ENL. We have determined that the binding affinities between the full length DOT1L and AF9/ENL are 33 nM and 206 nM respectively. The binding site in DOT1L was mapped as the region corresponding to DOT1L $_{865}$LPISIPLSTV$_{874}$. Our *in vitro* and *in vivo* findings demonstrated that the corresponding synthetic DOT1L 10mer peptide can competitively interfere with the interaction between MLL-fusion proteins, AF9 and ENL, and DOT1L. The alanine scanning mutagenesis studies showed the critical importance of several conserved hydrophobic residues, L865**,** I867, I869 and L871, for binding to AF9 and ENL. Deletion binding studies and HSQC NMR showed that an intact C-terminal domain in MLL-fusion proteins is critical for optimal interaction and binding of DOT1L 10-mer peptide to ENL protein induced significant conformation change of the protein. Furthermore**,** ENL mutant, L550E, which was reported to block the transforming capacity of MLL-ENL fusion protein [3], completely lost its ability to interact with DOT1L 10mer peptide.

4.2.2 Validation of MLL fusion protein and DOT1L PPI as a therapeutic target

Our functional studies using a colony forming unit assay provide strong evidence that the identified DOT1L$_{865\text{-}874}$ fragment is essential for the DOT1L

recruitment and MLL-AF9 leukemic transformation. The colony forming potential of the MLL-AF9 immortalized cells was completely abolished by introduction of a DOT1L construct lacking the 10 amino acid AF9 interacting residues**.** In addition to the importance of the H3K79 methylation, these results highlight the the DOT1L recruitment and its interaction with MLL-AF9 is necessary for the transforming activity of this MLL fusion oncogene and validate that MLL fusion protein and DOT1L interaction as a potential therapeutic target.

1.2.3 Phoshorylation of Ser 868 disrupts PPI between DOT1L and MLL-fusion protein

As an important post-transcription modification, phosphorylation is widely involved in the regulation of protein-protein interactions by affecting the stability, kinetics and specificity of interactions [11]. Phosphorylation and acetylation have been reported in DOT1L protein (http://www.phosphosite.org). However there are no reports of any PTMs in AF9 binding region of DOT1L.

The identified AF9 binding region in DOT1L has two Serines and one Threonine which can be post translational modified. Therefore we investigated if phosphorylation of these residues will affect the binding of DOT1L to AF9 and ENL. Our results showed that only phosphorylation of DOT1L Ser 868 abolished the AF9-DOT1L interaction, while phosphorylation of Ser 872 and Thr 873 did not have a significant impact on the binding affinity. Knowing that the

phosphorylation has an important role in creating docking sites for interaction with other proteins, our findings indicate that phosphorylation of Ser 868 may allow transient and reversible regulation of the PPI between DOT1L and AF9/ENL, thus controlling the DOT1L recruitment.

4.2.3 Development and optimization of FP assay and perform high-throughput screening

We designed a FP-based competitive assay, using fluorescein labeled DOT1L peptide including the 10 amino acids crucial for interaction with AF9 and ENL and recombinant GB1-AF9 protein. Various factors, including concentration of the probe and protein, percent of DMSO, time window for the assay reading, were tested to optimize the performance of the assay. The FP-based competitive assay was optimized and miniaturize to 384-well and 1536-well plate format with Z' factor of 0.75 and 0.65 respectively. To complement our FP assay, we also developed a SPR-based competitive assay and STD NMR as secondary assays to confirm and validate identified compounds from the HTS for their binding affinity and direct interactions with the fusion protein. In addition the SPR based binding assay allowed us to further validate if the compounds can disrupt PPI since in our experimental setting DOT1L protein (826-1095) was immobilized on the chip and AF9 pre-incubated with tested compounds was injected over DOT1L.

Using the optimized FP assay, we applied high throughput screening approach to identify inhibitors of the PPI between DOT1L and AF9. Approximately 101,000 compounds were screened at single concentration of 20 µM in 384-well plates. Total of 1191 compounds with inhibition larger than 3 fold of standard deviation (3SD) in primary screening were selected for confirmation assay in order to confirm their specific binding and eliminate fluorescent compounds, quenchers and compounds with reactive functional groups. Total of 164 compounds were selected and tested in dose-response screen. Thirty nine compounds demonstrated dose-dependent inhibition in the FP assay. The analysis of the chemical structures of these 39 compounds showed that they can be classified into five clusters with different chemical scaffolds and the rest of the compounds have unique structures. Based on the potency, chemical structure and availability, eight compounds were purchased and retested by the FP competitive assay. Four compounds were confirmed in both FP and SPR competitive assay, with IC_{50} from 10 - 60 µM in FP assay, and 20 - 70µM in SPR assay respectively. The direct binding of two compounds, belonging to two different chemical classes, was further confirmed by STD NMR spectroscopy. These identified compounds merit further characterization and would serve as starting point for chemical optimizations and development of small molecule inhibitors targeting the AF9-DOT1L interaction.

In summary, several candidate small molecules, with diverse chemical structure, were identified and validated that compete with binding of DOT1L to AF9/ENL. For the first time we have shown that it is possible to discover non-peptide small-molecules that can block the interaction between DOT1L and AF9. This is an important initial proof-of-concept for the development of small molecules targeting C-terminal hydrophobic domain in AF9, indicating it is a “druggable” target.

Overall, in the thesis described here, to gain insights into the unique functions of DOT1L in MLL-driven leukemia we have characterized AF9/ENL and DOT1L interaction and mapped the binding site to a short segment of 10 amino acids in DOT1L. Peptides derived from this region disrupted the interaction between DOT1L and MLL-AF9. DOT1L mutants lacking these 10 residues did not support transformation by MLL-AF9/ENL, while preserving global H3K79 methylation. Based on these results, FP-based competitive assay was optimized and high-throughput screen of a large chemical library (>101,000 compounds) was performed. Several candidate small molecules, with diverse chemical structure, were identified and validated that compete with binding of DOT1L to AF9/ENL. These work generated sufficient data and evidence for further optimization of identified lead compounds based on structural analysis, chemical synthesis and cell based assays towards development of small molecule inhibitors targeting the PPI between MLL fusion proteins and DOT1L.

4.3 Future Directions

The work presented in this thesis opens many potential avenues for future research toward further elucidating and understanding the mechanism of DOT1L recruitment by MLL fusion proteins and development of small molecules inhibitors of this PPI. Thus for the future directions of our work, we propose four major branches of research: 1) Following up the identified compounds from HTS; 2) Development of peptidomimetic inhibitors; 3) Structural studies of AF9/ENL-DOT1L interaction by X-ray crystallography. 4) Investigation the role of signal transduction in the regulation of DOT1L recruitment

4.3.1 HTS hits follow up

Our work identified several chemical scaffolds as inhibitors of the interaction between DOT1L and AF9 and generated sufficient data and evidence for their further optimization. Based on structural analysis of the most promising identified lead compounds, **36** and **38**, in the first step commercially available analogues of these compounds will be ordered and tested by our established assays in order to be established initial structure activity relationship (SAR). The ultimate goal of these studies will be improving their potency and their further characterization in cell-based assays.

According to SciFinder database search, there are 42 and 30 analogues of compounds **36** and **38,** respectively**,** with 80% or higher structure similarity. STD NMR studies of compound **36,** identified class IV**,** indicated that the R_1 substitution is involved in interaction with AF9 protein. Therefore, commercially available compounds with variations on R_1 substitution will be prioritized and ordered for establishment of initial SAR analysis. For compound **38**, which belong to class I, in the initial step we will order commercially available analogues maintain the 1, 4-diazepin core structure, with variations on R substitutions. All purchased analogues will be tested by several binding assays including FP and SPR based competitive assay, ITC or direct SPR based binding assay, followed by STD and HSQC NMR spectroscopy. These assays will provide us with detailed information about their binding affinity to ENL and AF9, ability to disrupt the PPI and structural insights about the interactions between the small-molecule inhibitors and fusion proteins, AF9 and ENL. In addition, the physico-chemical properties, solubility as well as the metabolic stability will be used as criteria to select the most promising lead compound for further development.

Based on the initial SAR analysis of the small set of compounds, additional medicinal chemistry efforts will be put to improve the potency and properties of these compounds. Developed compounds will be used as chemical

probes to understand the biological consequence of inhibiting DOT1L recruitment by MLL-fusion proteins, AF9 and ENL, and its impact on leukemia development, as well as on normal hematopoiesis. Ultimately, this work may result in the development of novel drugs for effective treatment of MLL acute leukemias.

4.3.2 Development of peptidomimetics

In addition to using HTS approach to identify small molecule inhibitors, based on our homology model of ENL/DOT1L 7mer peptide complex (Figure 2.11) we will proceed towards developing peptidomimetics inhibitors as additional approach for targeting AF9/ENL-DOT1L interaction. This approach has been applied successfully in targeting protein-protein interactions [12, 13]. Based on our alanine scanning mutagenesis, four conserved hydrophobic residues in DOT1L were identified crucial for the interaction with AF9/ENL. As a starting point, we will synthesize series of mutated DOT1L peptides where these crucial residues will be substituted with un-natural amino acids in order to improve their binding affinity as well as proteolytic stability. For example, Homoleucine (Hle), Norleucine (Nle) and t-butyl-Leucine (Tle) can be used for Leucine and Isoleucine mutated peptides. In the next step using available structural information and the presence of two prolines in the binding peptide motif we will design constrain peptidomimetics which will have significant improvement of

physicochemical properties. This approach has been successfully used for developing Smac peptidomimetics [14]. Besides improving the hydrophobic interactions, future studies will be also focused on exploring the hydrogen bond interactions which were identified based on our homology model. It is known that hydrogen bonds between polarized atoms play a crucial role in protein interactions and therefore we are expecting that hydrogen bond interactions between the small-molecule inhibitors and the target protein will introduce both specificity and affinity within the hydrophobic ligand pocket.

4.3.3 Protein crystallography studies

We have already initiated protein crystallography studies based on our biochemical characterization of AF9/ENL and DOT1L interactions. A series of AF9/ENL-DOT1L fusion protein constructs were generated (Appendix 1, Figure 1) and these constructs gave good expression of soluble fusion proteins (Appendix 1, Figure 2). Once we optimized the purification of these recombinant fusion proteins, we will set up crystallization screening trials using the following buffer screens: Hampton Index, Hampton Screen, JCSG and Emerald Wizards. Positive hits from the initial screens will be optimized by grid screen to get good quality protein crystals suitable for structure determination. If the initial crystallization screening trials fail to generate positive hits, NMR spectroscopy

can be applied as an alternative approach. These generated constructs can be used to express ^{15}N and ^{13}C labeled proteins suitable for NMR structural studies. With the combination of X-ray crystallography and NMR approaches, we would have better understanding of this protein-protein interaction from the structure perspective.

4.3.4 Role of post-transcriptional modification in regulation of DOT1L recruitment by MLL fusion proteins

The increase in complexity from the level of the genome to the proteome is further facilitated by protein post-translational modifications (PTMs). PTMs are chemical modifications that play a key role in functional proteomics, because they regulate activity, localization and interaction with other cellular molecules such as proteins, nucleic acids, lipids, and cofactors. These modifications include phosphorylation, glycosylation, ubiquitination, nitrosylation, methylation, acetylation, lipidation and proteolysis and influence almost all aspects of normal cell biology and pathogenesis. Therefore, identifying and understanding PTMs is critical in the study of cell biology and disease treatment and prevention.

Our results indicate that phosphorylation within the consensus AF9 binding sequence in DOT1L, Ser 868, abolishes the AF9-DOT1L interaction and therefore this phosphorylation may allow transient and reversible regulation of

the PPI between DOT1L and AF9/ENL. Thus it is very important these studies to be continued towards identification of phosphorylated S868 *in vivo*, followed by determination of the proteins/pathways that mediate this phosphorylation.

Recent technical developments in large–scale mass spectrometry-based phosphoproteomic analysis have demonstrated the powerful potential for rapid discovery of phosphorylation-site in proteins. Metal oxide affinity chromatography (MOAC) followed by immunoprecipitation using phopshoserine antibodies to enrich for phosphoserine peptides followed by standard LC-MS/MS analysis is one of the methods for large scale identification of phosphopeptides and phosphorylation sites.

Generating antibody which can specifically recognize the phosphorylation of S868 can be used for profiling panel of cell lines which will provide insights about the phosphorylation status of endogenous DOT1L protein in different type of normal as well as cancer cell lines.

Using available web based programs and the identified binding sequence of 10 amino acids in DOT1L protein, it was predicted that this region might be a possible substrate motif of several Ser/Thr kinases, including Glycogen synthase kinase 3 (GSK-3), G protein-coupled receptor kinase 1 (GRK-1), MAPKAPK2 kinase and PKA kinase. In order to be determined which Ser/Thr kinase is responsible for Ser 868 modification, different DOT1L synthetic peptides, wild

type and mutated, can be used as a substrate and screen a panel of Ser/Thr kinases.

Apart from changing the intrinsic properties of proteins, an important role of phosphorylation is to create docking sites which mediate the protein-protein interactions. Therefore, to study the potential role of DOT1L phosphorylation for interaction with other proteins besides with MLL-fusion proteins, pull down assay using the S868 phosphorylated peptide can be applied. Cellular proteins which might interact with this peptide and be pulled down can be identified by MS.

The significance of these future studies is the elucidation if the post-translational modifications are involved in regulation of the recruitment of DOT1L by MLL-fusion proteins, which may allow transient and reversible regulation of these interactions. Furthermore, identifying signaling network involved in DOT1L phosphorylation in the AF9 binding motif might open new potential therapeutic intervention.

Overall, our targeted molecular approach to prevent DOT1L recruitment to MLL fusion proteins without blocking normal DOT1L functions represents a major innovation of this study. By taking different approaches we expect to identify and develop both small molecule and peptidomimetics inhibitors of the DOT1L/ MLL-fusion protein interaction. Inhibitors that will achieve this goal could markedly improve the therapeutic window of DOT1L inhibitors in patients with MLL-driven

acute myeloid and lymphoblastic leukemia, which are currently incurable and there is a clinical need of novel therapeutics.

4.4 References

1. Krivtsov, A.V. and S.A. Armstrong, *MLL translocations, histone modifications and leukaemia stem-cell development.* Nat Rev Cancer, 2007. **7**(11): p. 823-33.
2. Mueller, D., et al., *A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification.* Blood, 2007. **110**(13): p. 4445-54.
3. Yokoyama, A., et al., *A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription.* Cancer Cell, 2010. **17**(2): p. 198-212.
4. Okada, Y., et al., *hDOT1L links histone methylation to leukemogenesis.* Cell, 2005. **121**(2): p. 167-78.
5. Mohan, M., et al., *Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom).* Genes Dev, 2010. **24**(6): p. 574-89.
6. Nguyen, A.T. and Y. Zhang, *The diverse functions of Dot1 and H3K79 methylation.* Genes Dev, 2011. **25**(13): p. 1345-58.
7. Bernt, K.M. and S.A. Armstrong, *A role for DOT1L in MLL-rearranged leukemias.* Epigenomics, 2011. **3**(6): p. 667-70.
8. Bernt, K.M., et al., *MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L.* Cancer Cell, 2011. **20**(1): p. 66-78.
9. Daigle, S.R., et al., *Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor.* Cancer Cell, 2011. **20**(1): p. 53-65.
10. Biswas, D., et al., *Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes.* Proc Natl Acad Sci U S A, 2011. **108**(38): p. 15751-6.
11. Seet, B.T., et al., *Reading protein modifications with interaction domains.* Nat Rev Mol Cell Biol, 2006. **7**(7): p. 473-83.
12. Karatas, H., et al., *High-affinity, small-molecule peptidomimetic inhibitors of MLL1/WDR5 protein-protein interaction.* J Am Chem Soc, 2013. **135**(2): p. 669-82.
13. Walensky, L.D., et al., *Activation of apoptosis in vivo by a hydrocarbon-stapled BH3 helix.* Science, 2004. **305**(5689): p. 1466-70.
14. Sun, H., et al., *Design of small-molecule peptidic and nonpeptidic Smac mimetics.* Acc Chem Res, 2008. **41**(10): p. 1264-77.

Appendices

Appendix 1. Design and expression of DOT1L-AF9 and DOT1L-ENL fusion proteins for structural studies

Background

During our initial efforts to express and purify the AF9/ENL C-terminal domain, we found that the untagged AF9 (497-568) or ENL (489-559) expressed in *Escherichia coli* as insoluble inclusion bodies (Figure A.1). Although AF9 or ENL C-terminal domain cleaved from the MBP- or GB1-fusion did retain DOT1L protein and peptide-binding activity and we were able to use the cleaved protein to perform biochemical binding assays, these proteins were of insufficient purity and biochemical homogeneity for use in biophysical studies, especially for crystallography studies. Recently, the NMR structure of AF4-AF9 fusion protein was published [1]. In this study, a construct, which includes 43 amino acids of AF4 fused to the flexible residues immediately preceding the AF9 C-terminal domain, was generated to overcome the poor solubility of AF9 protein. This approach resulted in expression of a monomeric homogeneous complex and yielded high quality spectra suitable for structure determination. According to the

NMR structural studies, the hydrophobic core of AF9 is not sufficient to maintain an independently folded structure. The AF9 CTD need to recruit partners through coupled folding and binding to form a well folded complex. This is also observed in ENL C-terminal domain by our HSQC NMR results. Binding of DOT1L protein or peptides to ENL induced significant conformational change of ENL protein and indicated formation of a folded complex structure (Figure 2.10). Based on these studies, we decided to design DOT1L-AF9 and DOT1L-ENL fusion proteins for the further biophysical studies. The design of the fusion protein was based on these criteria: 1) including the minimum interaction region of DOT1L and AF9/ENL; 2) including a flexible linker between two proteins, to allow for optimum folding and formation of the protein complex.

Material and methods

Molecular cloning

The genes sequences for the fusion proteins were based on the protein sequences and the codons were optimized for E. coli expression. All the genes were synthesized by Integrated DNA Technologies. All the constructs for fusion protein expression were cloned by ligation independent cloning (LIC) methods as described before [2]. The primers for PCR are listed in Table 1 and synthesized by Life Technologies. Different fusion constructs (Figure 1) were cloned into

pMCSG7-LIC vector. All the constructs were sequenced by The University of Michigan Sequencing Core.

Table 1 Primers for molecular cloning

Fusion protein No.		Sequence (5'-3')
1	Sense	TACTTCCAATCCAATGCAGGA AGC AGT GGC
	Antisense	TTATCCACTT CCATTA GCT CGTACCGCTGGTCT
2	Sense	TACTTCCAATCCAATGCAGGATCTTCAGGC
	Antisense	TTATCCACTTCCATTATGAGGTACCGCTGGTCT
3	Sense	TACTTCCAATCCAATGCAGGGAGTAGTGGC
	Antisense	TTATCCACTTCCATTAGGTTGCCACTGCTTCCA

Protein expression analysis

All the recombinant proteins were expressed in *Escherichia coli* strain BL21 (DE3) (Invitrogen) in a small scale (50 mL) as a primary screening. The medium for bacterial growth was Terrific Borth (TB). All the proteins were induced with 200 µM IPTG at 16 °C for 20 h. The cells were harvested after 20h and re-suspended in cold lysis buffer (50 mM Tris HCl, pH 7.5, 150 mM NaCl, 0.01 % β-mercatoethanol) and lysed by sonication. The cells were centrifuged at 13,200 rpm for 15 min. The whole cell lysate before centrifugation, the supernatant and pellet after centrifuge were loaded on SDS-PAGE for analysis.

Results and Discussion

Design of DOT1L-AF9 and DOT1L-ENL fusion proteins

The designed fusion proteins are illustrated in Figure A.1. We chose to the include AF9 (490-568) or ENL (482-559) on the C-terminus of the fusion protein. Based on the published NMR structure of AF4-AF9 fusion protein, two different DOT1L fragments were chosen as the N-terminus of the fusion protein, they are: DOT1L 856-897 and DOT1L 856-883.Both of these two fragments contain the AF9/ENL interaction region. A three residues linker including PAG was introduced between the two proteins.

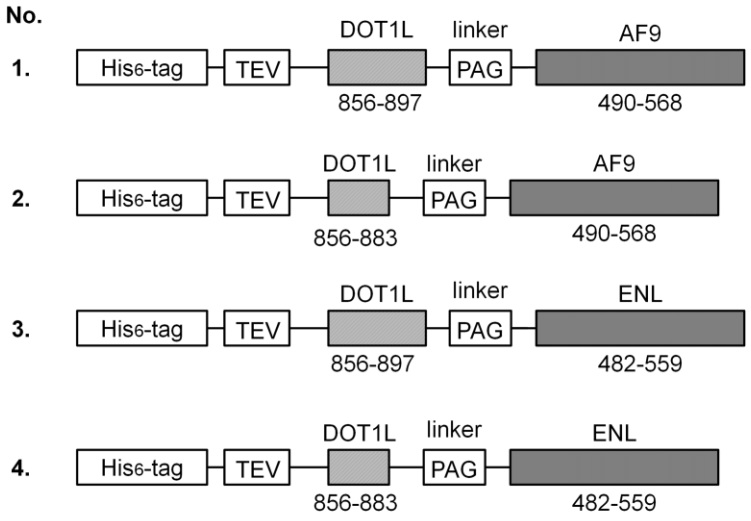


Figure 1 Design of DOT1L-AF9 and DOT1L-ENL fusion proteins. His-tag and TEV cleavage site were on the N-terminus of the protein, following with DOT1L (856-897) or DOT1L (856-883). The AF9 (490-568) or ENL (482-559) was on the C-terminus and there is a three residue (PAG) linker between DOT1 L and AF9 or ENL.

Expression of DOT1L-AF9/ENL fusion proteins

The expression of the designed DOT1L-AF9 or DOT1L-ENL fusion proteins was tested in a small scale. After induction with 200 μM IPTG at 16 °C for 20 h, the cells were analyzed by SDS-PAGE. As shown Figure 2A, His-ENL

(489-559) protein was not soluble when expressed alone (Figure A.2A, lane 5). However, the DOT1L-AF9 and DOT1L-ENL fusion proteins (Figure A.2B, lane 3, 5 and 7), were soluble after the IPTG induction.

The expression of the DOT1L-AF9/ENL fusion protein in the soluble fraction of the cell lysate indicates that fusion of DOT1L to AF9 and ENL assist the folding and solubility of the disordered protein. With further expression and purification of the fusion protein, the protein can be applied for biophysical structural studies.

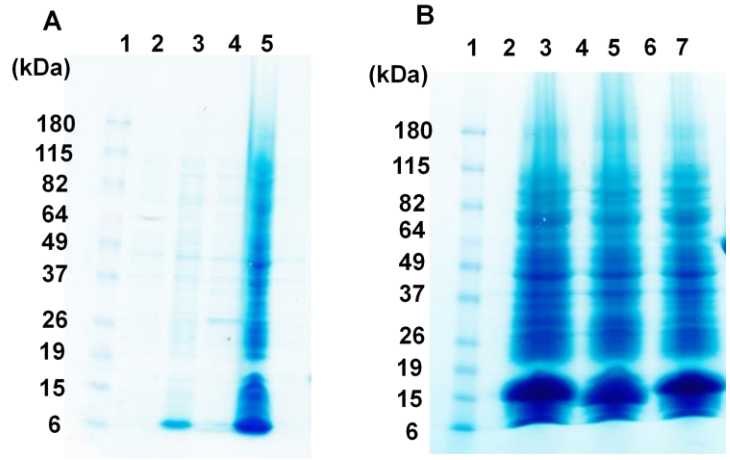
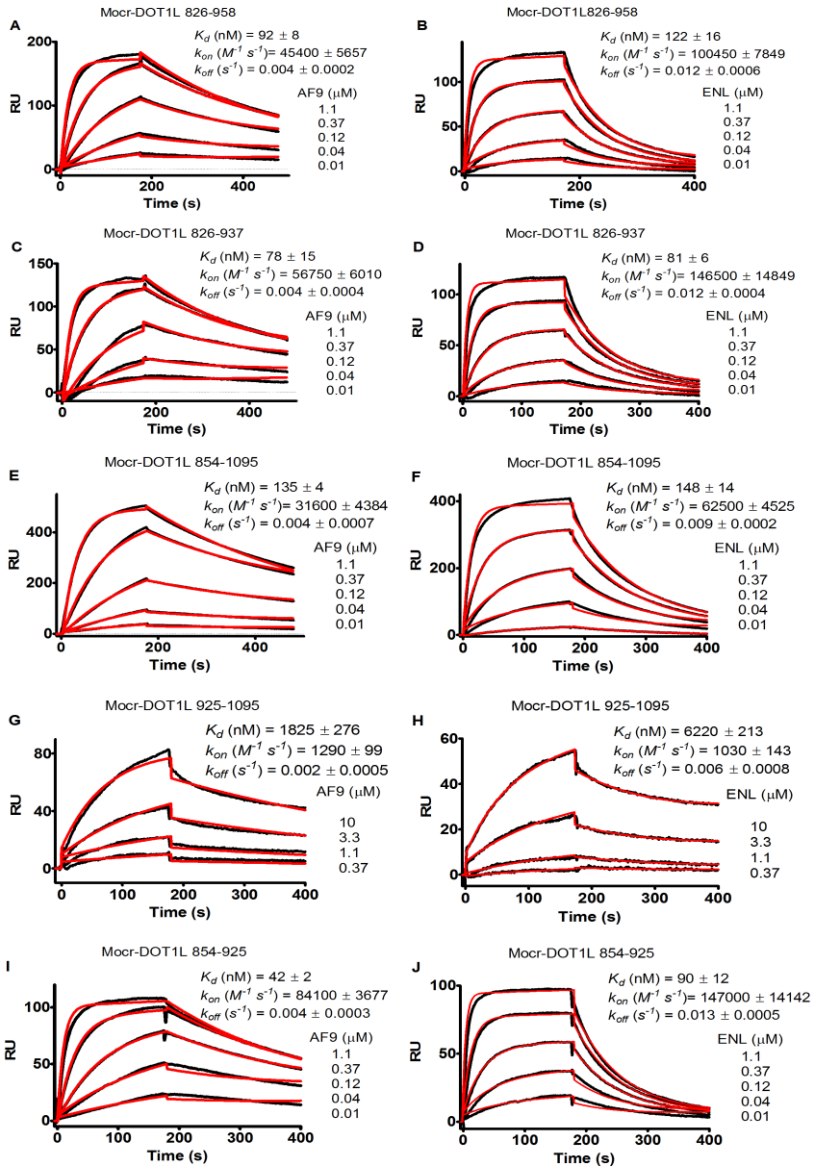


Figure 2 (A) Expression of His-ENL (489-559) in E. coli BL21 (DE3) cells as monitored by Glycine-SDS-polyacrylamide gel electrophoresis of cell extracts. Lane1, protein ladder; lane 2, non-induced; lane 3, induced with IPTG, total extract; lane 4, induced with IPTG, supernatant; lane 5, induced with IPTG, pellet. The calculated molecular weight of His-ENL (489-559) is 11.1 kDa. (B) Expression of DOT1L-AF9/ENL fusion proteins in in E. coli BL21 (DE3) cells. Lane 1, protein ladder; lane 2, empty; lane 3, fusion protein 1, induced with IPTG, supernatant; lane 4, empty; lane 5, fusion protein 2, induced with IPTG, supernatant; lane 6, empty; lane 7, fusion protein 3, induced with IPTG, supernatant. The calculated molecular weight of fusion protein 1, 2 and 3 are: 16.5kDa, 14.6kDa, 16.3kDa respectively.

References

1. Leach, B.I., et al., *Leukemia fusion target AF9 is an intrinsically disordered transcriptional regulator that recruits multiple partners via coupled folding and binding.* Structure, 2013. **21**(1): p. 176-83.
2. DelProposto, J., et al., *Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications.* Protein Expr Purif, 2009. **63**(1): p. 40-9.

Appendix 2 SPR Sensorgrams representing the concentration-dependent binding of AF9 (497 - 568) and ENL (489 - 559) proteins tested with of truncated constructs of DOT1L



Appendix 3 Peptide sequence and IC_{50} values of wild type and alanine mutated DOT1L peptides against AF9 and ENL proteins obtained by competitive SPR based assay using CM5 chip with immobilized DOT1L protein.

Name	**Sequence**	**IC_{50} ± SD [µM]**	
		AF9	**ENL**
DOT1L 10mer	Ac-LPISIPLSTV-NH_2 (a.a. 865-874)	0.49 ± 0.22	1.34 ± 0.34
L865A	Ac-APISIPLSTV-NH_2	98.5 ± 15.1	27.5 ± 6.8
P866A	Ac-LAISIPLSTV-NH_2	3.3 ± 0.6	6.5 ± 1.2
I867A	Ac-LPASIPLSTV-NH_2	57.2 ± 17.3	>200
S868A	Ac-LPIAIPLSTV-NH_2	5.3 ± 1.7	29.1 ± 0.8
I869A	Ac-LPISAPLSTV-NH_2	>200	>200
P870A	Ac-LPISIALSTV-NH_2	2.8 ± 0.1	11.6 ± 1.7
L871A	Ac-LPISIPASTV-NH_2	30.7 ± 10.3	146.8 ± 11.5
S872A	Ac-LPISIPLATV-NH_2	0.55 ± 0.21	2.8 ± 0.6
T873A	Ac-LPISIPLSAV-NH_2	0.058 ± 0.017	0.36 ± 0.02
V874A	Ac-LPISIPLSTA-NH_2	0.7 ± 0.02	4.2 ± 0.6
DOT1L 7mer	Ac-LPISIPL-NH_2	3.9 ± 0.1	7.3 ± 0.3
DOT1L 16mer	Ac-LPISIPLSTVQPNKLP-NH_2 (a.a. 865 - 880)	0.32 ± 0.01	1.56 ± 0.09
AF4 14mer	Ac- LMVKITLDLLSRIP-NH_2 (a.a. 760 - 773)	0.20 ± 0.06	0.43 ± 0.15

Appendix 4 Identified peptides from Flag-DOT1L by mass spectrometry. The phosphorylated Serine, Threonine or Tyrosine residues are highlighted in bold and indicated as **pS, pT or pY.**

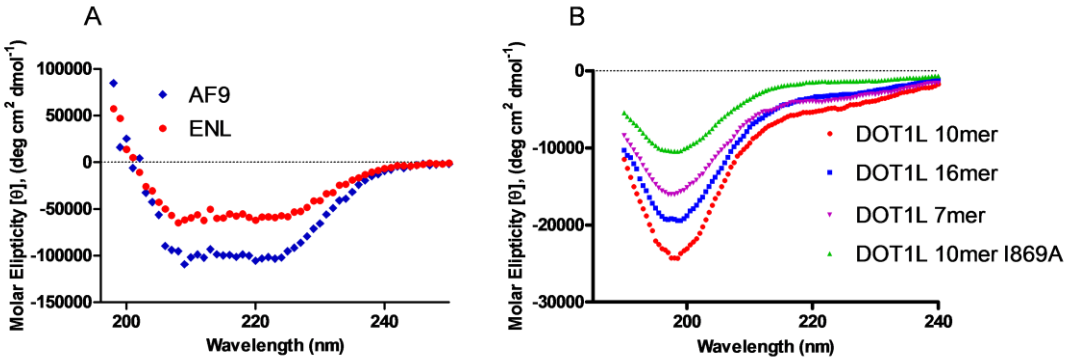
Protein	Peptide sequence	Initial probability	Phosphorylated sites in DOT1L (Q8TEK3)
sp|Q8TEK3	AIDSIHQLWK	0.999	
sp|Q8TEK3	ALTYNDLIQAQK	0.999	
sp|Q8TEK3	CEELQLDWATLSLEK	0.999	
sp|Q8TEK3	EAGEGGLPLCGPTDK	0.999	
sp|Q8TEK3	GDLPSDSGF**pS**DPESEAK	0.999	S1035
sp|Q8TEK3	GDLPSDSGFSDPESEAK	0.999	
sp|Q8TEK3	GPEAAGLS**pS**PLSFPSQR	0.999	S1349
sp|Q8TEK3	GPEAAGLSSPLSFPSQR	0.999	
sp|Q8TEK3	GSVSWTGKPVSYYLHTIDR	0.999	
sp|Q8TEK3	HAEYTLER	0.999	
sp|Q8TEK3	HHDAAHEIIETIR	0.999	
sp|Q8TEK3	HILQQVYNHSVTDPEK	0.999	
sp|Q8TEK3	IANTSVIFVNNFAFGPEVDHQLK	0.999	
sp|Q8TEK3	IQYLQFLAYTK	0.999	
sp|Q8TEK3	IVFTITTGAGSAK	0.999	
sp|Q8TEK3	KIATISLESK	0.999	
sp|Q8TEK3	KNQTALDALHAQTVSQTAASSPQDAYR	0.999	
sp|Q8TEK3	LAMENYVLIDYDTK	0.999	
sp|Q8TEK3	LGGAAQGPLPEASK	0.999	
sp|Q8TEK3	LKSPVGAEPAVYPWPLPVYDK	0.999	
sp|Q8TEK3	LPVSIPLASVVLPSR	0.999	
sp|Q8TEK3	LSPQDPRPL**pS**PGALQLAGEK	0.999	S834
sp|Q8TEK3-2	LTNSHAMGSFSGVAGGTVGGN	0.999	

sp|Q8TEK3	NAQLLGAAQQLLSHCQAQK	0.999	
sp|Q8TEK3	NGHNLFISAAAVPPGSLLSGPGLA PAASSAGGAASSAQTHR	0.999	
sp|Q8TEK3	NSLPA**pS**PAHQLSS**pS**PR	0.999	S1001, S1009
sp|Q8TEK3	NSLPASPAHQLSS**pS**PR	0.999	S1009
sp|Q8TEK3	NSLPASPAHQLSSSPR	0.999	
sp|Q8TEK3	Q[111]IGANAHGAGSR	0.999	
sp|Q8TEK3	**pS**SPVPYQDHDQPPVLK	0.999	S1152
sp|Q8TEK3	SADSPLQASSALSQNSLFTFRPALEEPSADAK	0.999	
sp|Q8TEK3	SCVPPDDALSLHLR	0.999	
sp|Q8TEK3	SPHSPFYQLPPSVQR	0.999	
sp|Q8TEK3	SPVGAEPAVYPWPLPVYDK	0.999	
sp|Q8TEK3	SSPVPYQDHDQPPVLK	0.999	
sp|Q8TEK3	SSTPQHPLLLAQPR	0.999	
sp|Q8TEK3	STFSPISDIGLAK	0.999	
sp|Q8TEK3	VAGPADAPMD**pS**GAEEEK	0.999	S374
sp|Q8TEK3	VAGPADAPMDSGAEEEK	0.999	
sp|Q8TEK3	GKEGSDANPFLSK	0.9989	
sp|Q8TEK3	IVSSKPFAPLNFR	0.9989	
sp|Q8TEK3	LSGLAAPDYTR	0.9989	
sp|Q8TEK3	NLSDIGTIMR	0.9989	
sp|Q8TEK3	NQTALDALHAQTVSQTAASSPQDAYR	0.9989	
sp|Q8TEK3	TILENYFSSLK	0.9989	
sp|Q8TEK3	WVCEEIPDLK	0.9989	
sp|Q8TEK3	ASAGTPSLSAGV**pS**PK	0.9988	S1104
sp|Q8TEK3	LGGAAQGPLPEASKGDLPSDSGFSDPESEAK	0.9988	
sp|Q8TEK3	RASAGTPSLSAGV**pS**PK	0.9988	S1104
sp|Q8TEK3	HLSQDHTVPGRPAASELHSR	0.9987	
sp|Q8TEK3	GDFLSEEWR	0.9986	
sp|Q8TEK3	EQSEQLEQDNR	0.9983	
sp|Q8TEK3	QNTPQYLASPLDQEVVPCTPSHVGRPR	0.9983	

sp|Q8TEK3	ASLQELLGQEK	0.9981	
sp|Q8TEK3	NSLPA**pS**PAHQL**pS**SSPR	0.9981	S1001, S1007
sp|Q8TEK3	KPAPAGEPVNSSK	0.998	
sp|Q8TEK3	QLDGLAGLKGEGSR	0.9979	
sp|Q8TEK3	ASAGTPSLSAGVSPK	0.9978	
sp|Q8TEK3	ENGLPYQSPSVPGSMK	0.9978	
sp|Q8TEK3	ENGLPYQSPSVPGSMK	0.9963	
sp|Q8TEK3	AYGSSGELITSLPISIPLSTVQPNK	0.9962	
sp|Q8TEK3	HCLELQISIVELEK	0.9953	
sp|Q8TEK3	LHLELDCTK	0.9952	
sp|Q8TEK3	TILENYFSSLKNPK	0.993	
sp|Q8TEK3	LNTRPSTGLLR	0.9855	
sp|Q8TEK3	ENGLPYQSPSVPGSMK	0.9833	
sp|Q8TEK3	GKEAGEGGLPLCGPTDK	0.9831	
sp|Q8TEK3	EISAHNQQLR	0.9805	
sp|Q8TEK3	HL**pS**QDHTVPGRPAASELHSR	0.9801	S786
sp|Q8TEK3	LSGLAAPDYTRL**pS**PAK	0.9801	S775
sp|Q8TEK3	ENGLPYQSPSVPGSMK	0.9752	
sp|Q8TEK3	IATISLESK	0.9748	
sp|Q8TEK3	STPSPVLQPR	0.9738	
sp|Q8TEK3	KIATISLESK**pS**PPK	0.9721	S1213
sp|Q8TEK3	KHAEYTLER	0.9689	
sp|Q8TEK3	STP**pS**PVLQPR	0.9669	S902
sp|Q8TEK3	HSPNPLLVAPTPPALQK	0.9605	
sp|Q8TEK3	QLDGLAGLK	0.9497	
sp|Q8TEK3	HSPLTASAR	0.9484	
sp|Q8TEK3	RASAGTPSLSAGVSPK	0.9442	
sp|Q8TEK3	QLDGLAGLK	0.8468	
sp|Q8TEK3	IATISLESK**pS**PPK	0.8465	S1213
sp|Q8TEK3	S**pT**P**pS**PVLQPR	0.8444	T900, S902
sp|Q8TEK3	HLSQDHTVPGRPAA**pS**ELHSR	0.7725	

sp|Q8TEK3	SFESMQR	0.7724	
sp|Q8TEK3	ELEPDASR	0.7695	
sp|Q8TEK3	VVELSPLK	0.7199	
sp|Q8TEK3	VVEL**pS**PLK	0.6456	S297
sp|Q8TEK3	ARS**pT**P**pS**PVLQPR	0.6223	T900, S902
sp|Q8TEK3	LREEQEAAR	0.5131	
sp|Q8TEK3	LSGLAAPD**pY**TRLSPAK	0.441	Y771
sp|Q8TEK3	QQELLQLK	0.3865	
sp|Q8TEK3	TLENGGGLAGR	0.2253	
sp|Q8TEK3	LNNYEPFSPEVYGETSFDLVAQM[147]IDEIK	0.1867	
sp|Q8TEK3	IATISLE**pS**KSPPKTLENGGGLAGR	0.0994	S1211
sp|Q8TEK3	**pS**PPKTLENGGGLAGR	0.0716	S1213
sp|Q8TEK3	AEHTKENGLPYQSPSVPGSMK	0.0683	
sp|Q8TEK3	VAGPADAPMD**pS**GAEEEKAGAATVK	0.0644	S374

Appendix 5 Circular dichroism of proteins and peptides.



A.CD of AF9(497-568) and ENL (489-559) protein. B. CD of different DOT1L peptides. DOT1L peptides were dissolved in phosphate buffer (50mM Na_2HPO_4, pH 7.4, 100mM NaCl). ENL and AF9 protein were dialyzed against the above phosphate buffer overnight. CD measurements were performed at room temperature using a Jasco J-715 and a quartz flow cell with a 1 mm path length. The spectra were averaged over 10 scans and the baseline (buffer scan) was subtracted from each spectrum.