

## ABSTRACT

Translocations and amplifications of the mixed lineage leukemia-1 (*MLL1*) gene are associated with aggressive myeloid and lymphocytic leukemias in humans. MLL1 is a member of the SET1 family of histone H3 lysine 4 (H3K4) methyltransferases, which are required for transcription of genes involved in hematopoiesis and development. MLL1 associates with a sub-complex containing WDR5, RbBP5, Ash2L, and DPY-30 (WRAD), which together form the MLL1 core complex that is required for sequential mono- and dimethylation of H3K4. We previously demonstrated that WDR5 binds the conserved WDR5 interaction (*Win*) motif of MLL1 *in vitro*, an interaction that is required for the H3K4 dimethylation activity of the MLL1 core complex. In this dissertation, we demonstrate that arginine 3765 of the MLL1 *Win* motif is required to co-immunoprecipitate WRAD from mammalian cells, suggesting that the WDR5-*Win* motif interaction is important for the assembly of the MLL1 core complex *in vivo*. We also demonstrate that peptides that mimic human SET1 family *Win* motif sequences (MLL1-4, SETd1a and SETd1b) inhibit H3K4 dimethylation by the MLL1 core complex with varying degrees of efficiency. We show that the MLL3 *Win* motif peptide is the best inhibitor of the H3K4 dimethylation activity of MLL1 core complex. To understand the structural basis for these differences, we determined three-dimensional structures of WDR5 bound to six different naturally occurring *Win* motif sequences (MLL1-4, SETd1a and SETd1b). The structural analysis reveal that binding energy differences result from interactions between non-conserved residues C-terminal to the *Win* motif and to a lesser extent from subtle variation of residues within the *Win* motif. Based on the structure-function analysis, we deduce structural rules to facilitate the design of two additional *Win* motif –based inhibitors (Ac-10-mer and six-residue *Win* motif peptides) that bind WDR5 with <10 nM affinities. To understand the structural basis

for this low nanomolar affinity, we determined X-ray three-dimensional structures of the Ac-10-mer *Win* motif peptide bound to WDR5. The structures suggest that the presence of additional intramolecular hydrogen bonds might contribute to the increased affinities for WDR5 possibly through the stabilization of the bound  $3_{10}$ -helical conformation. We extend this structure-function analysis further to identify other peptidomimetics by characterizing peptides identified in a randomized phage display screen, which are also highly specific inhibitors of MLL1 core complex. Crystal structures of these peptidomimetics reveal novel protein structural features that contribute to increased affinity. We also present preliminary evidence suggesting that the MLL3 *Win* motif based peptide that has a cell penetrating sequence is readily taken up by mammalian cells. This MLL3 *Win* motif-based peptide (MLL3-FITC-TAT) is localized to euchromatin regions of cell nuclei, induces nuclear defects and inhibits global levels of H3K4 trimethylation. These results highlight a new class of methylation inhibitors that may be useful for the treatment of MLL1-related malignancies.

**THERMODYNAMIC AND STRUCTURAL  
CHARACTERIZATION OF THE INTERACTION  
BETWEEN WD-40 REPEAT PROTEIN 5 (WDR5)  
AND HUMAN SET1 FAMILY  
METHYLTRANSFERASES**

By

**Venkatasubramanian Dharmarajan**  
B.Tech. Anna University, Chennai, INDIA (2004)

**DISSERTATION**

Submitted in partial fulfillment of the requirements for the degree of Doctor of  
Philosophy in Structural Biology, Biochemistry and Biophysics in the Graduate  
School of Syracuse University

**May 2013**

**Copyright 2013 Venkatasubramanian Dharmarajan**

**All Rights Reserved**

## ACKNOWLEDGEMENTS

I find it very hard to put it in words the feelings I share towards this project. This whole project and my entire graduate school career has been a “memorable” journey for me in several ways. I have come a long way as a scientist and as a person and this graduate school experience has helped shape both aspects of it. There are several people that I would like to thank for their contribution towards the completion of this thesis, especially my advisor Dr. Michael S. Cosgrove. He is the best mentor I could have asked for in several ways. He is an exceptional scientist, a hardworker, passionate towards science, and on top of all a great person. I thank him for his constant support and motivation.

I am also grateful to the members of my committee, Dr. Bruce Hudson, Dr. Liviu Movileanu, and especially Dr. Stephan Wilkens. They all have helped me throughout my graduate school career and I thank them for their support.

My special thanks to Dr. Anamika Patel our lab postdoc. She has always been there for me when I needed something and took care of me like a brother. I thank her for being such a good-hearted person and being a role model for me. I would also like to thank Melody Sanders and Stephen Shinsky, my fellow graduate students, for making our lab environment tremendous fun. I would like to thank my undergrads Michael Hu and Kelsey Monteith for being my friends and sharing all the good times with me.

I would like to thank Dr. Rebecca Oot from Stephan Wilkens lab for her help with ITC. I would like to thank Dr. Laila Kabrossy and Dr. Torsten Wollert for their help with the in vivo

experiments. I thank our collaborators, Dr. David Skalnik's group and Dr. Ziwei Huang's group for doing the Co-immunoprecipitation experiments and synthesizing some of the peptides used in this dissertation. I thank the staff at Cornell Synchrotron facility for their help with X-ray data collection and analysis.

I would like to thank my family, my mom, dad and my two sisters for all their support and love over the several years. I would like to thank all my friends who made this possible. I am blessed to have such a wonderful family and friends in my life and I thank them all.

## TABLE OF CONTENTS

ABSTRACT.....	I
ACKNOWLEDGEMENTS.....	V
TABLE OF CONTENTS.....	VII
LIST OF FIGURES.....	X
LIST OF TABLES.....	XIV
<b>CHAPTER 1: INTRODUCTION.....</b>	<b>1</b>
BACKGROUND AND SIGNIFICANCE.....	1
PURPOSE .....	15
<b>CHAPTER 2: TARGETED THERAPY FOR MIXED LINEAGE LEUKEMIA-WHAT HAVE WE LEARNED FROM THE BIOCHEMISTRY OF MLL1.....</b>	<b>19</b>
<b>CHAPTER 3: THERMODYNAMIC CHARACTERIZATION OF THE INTERACTION BETWEEN WDR5 AND HUMAN SET1 FAMILY <i>WIN</i> MOTIFS.....</b>	<b>72</b>
TITLE PAGE.....	72
INTRODUCTION.....	73
METHODS.....	74
RESULTS.....	79

DISCUSSION.....	107
<b>CHAPTER 4: STRUCTURAL CHARACTERIZATION OF THE INTERACTION BETWEEN WDR5 AND HUMAN SET1 FAMILY <i>WIN</i> MOTIFS.....</b>	<b>110</b>
TITLE PAGE.....	110
INTRODUCTION.....	111
METHODS.....	113
RESULTS.....	115
DISCUSSION.....	152
<b>CHAPTER 5: <i>WIN</i> MOTIF INHIBITORS: A NOVEL CLASS OF METHYLATION INHIBITORS WITH THE POTENTIAL TO TREAT MLL-RELATED LEUKEMIA .....</b>	<b>155</b>
TITLE PAGE.....	155
INTRODUCTION.....	156
METHODS.....	158
RESULTS AND DISCUSSION.....	163



APPENDIX 1.....	212
APPENDIX 2.....	218
APPENDIX 3.....	224
APPENDIX 4.....	227
APPENDIX 5.....	230
APPENDIX 6.....	236
REFERENCES.....	237
BIOGRAPHICAL DATA.....	262

## LIST OF FIGURES

Figure 1.1 Schematic models for the role of WDR5 in the MLL1 core complex.....	8
Figure 1.2 WDR5 binds to a conserved arginine containing sequence in the N-SET region of MLL1.....	11
Figure 2.1: Substrate specificity of mammalian histone lysine-methyltransferases (HKMTs)	26
Figure 2.2 Schematic representations of human SET1 family methyltransferases.....	27
Figure 2.3 Model for multiple lysine methylation catalyzed by MLL1 core complex.....	32
Figure 2.4: Schematic representation of the most common genetic alterations associated with the MLL1 gene .....	37
Figure 2.5: Domain architecture of human menin showing the putative conserved domains...	46
Figure 2.6: Menin Binding Motifs (MBMs) are present in the wild-type, chromosomal translocations and partial tandem duplication mutations in MLL1.....	53
Figure 2.7: Domain representation of human AF4 indicating the putative functional domains including the AF9 interaction region .....	58
Figure 2.8 Potential protein-protein interaction targets for the development of novel therapies for MLL1-associated leukemias.....	71
Figure 3.1 R3765 of MLL1 <i>Win</i> motif is required for the co-immunoprecipitation of MLL1 core complex components.....	81
Figure 3.2 Different length MLL1 <i>Win</i> motif peptides bind WDR5 with similar affinities.....	84
Figure 3.3 Human SET1 family <i>Win</i> motif peptides bind WDR5 with a wide range of affinities.....	87
Figure 3.4 Thermodynamic binding signatures for Human SET1 family <i>Win</i> motif peptides...	91
Figure 3.5 Comparison of inhibition constants for different MLL1 <i>Win</i> motif peptides.....	94
Figure 3.6 Human SET1 family <i>Win</i> motif peptides inhibit the H3K4 dimethylation activity of the MLL1 core complex with varying efficiencies.....	97
Figure 3.7 Pre-incubation with WRAD lowers the inhibition constants for SET1 family <i>Win</i> motif peptides.....	100

Figure 3.8 MLL3 <i>Win</i> motif peptide inhibits MLL1 core complex mediated nucleosomal methylation.....	103
Figure 3.10 Free WDR5 acts as an inhibitor of MLL1 core complex mediated H3K4 dimethylation activity.....	105
Figure 4.1: Amino acid sequence of the human SET1 family <i>Win</i> motif peptides used in crystallization.....	116
Figure 4.2 Pair-wise superposition of peptide monomers in the asymmetric unit for MLL1, MLL2 and MLL4 <i>Win</i> motif structures.....	120
Figure 4.3 Simulated Annealing omit maps showing the different human SET1 family <i>Win</i> motif peptides bound to WDR5.....	121
Figure 4.4a Structure of MLL1 <i>Win</i> motif peptide bound to WDR5.....	126
Figure 4.4b Structure of MLL2 <i>Win</i> motif peptide bound to WDR5.....	127
Figure 4.4c Structure of MLL3 <i>Win</i> motif peptide bound to WDR5.....	128
Figure 4.4d Structure of MLL4 <i>Win</i> motif peptide bound to WDR5.....	129
Figure 4.4e Structure of SETd1a <i>Win</i> motif peptide bound to WDR5.....	130
Figure 4.4f Structure of SETd1b <i>Win</i> motif peptide bound to WDR5.....	131
Figure 4.5 Intramolecular hydrogen bonds within human SET1 family <i>Win</i> motif peptides...	132
Figure 4.6 Intermolecular hydrogen bonds between WDR5 and different SET1 family <i>Win</i> motif peptides.....	133
Figure 4.7 Intra and intermolecular van der Waals contacts present in different human SET1 family <i>Win</i> motif peptides.....	134
Figure 4.8 Human SET1 family <i>Win</i> motif peptides bind WDR5 using a similar $3_{10}$ helical conformation.....	136
Figure 4.9 Conserved <i>Win</i> motif residues adopt structurally similar conformations.....	138
Figure 4.10 Structural differences at the N-terminus.....	143
Figure 4.11 Structural differences at the C-terminus.....	146
Figure 4.12 The -2 and +4 amino acids of SET1 family <i>Win</i> motifs contribute to high affinity binding to WDR5.....	151

Figure 5.1 Ac-10-mer peptide ( ${}_{\text{Ac}}\text{ARTEVHLRKS}_{\text{NH}_2}$ ) binds WDR5 with the strongest affinity.....	166
Figure 5.2: Ac-10-mer peptide specifically inhibits the H3K4 dimethylation activity of MLL1 core complex.....	167
Figure 5.3 Structure of Ac-10-mer peptide bound to WDR5.....	174
Figure 5.4 Crystal structure of Ac-10-mer peptide bound to WDR5.....	175
Figure 5.5 Summary of interactions observed in peptide 3-WDR5 co-crystal structure.....	176
Figure 5.6 Structural comparisons between Ac-10-mer and MLL1/MLL3 <i>Win</i> motif peptides.....	177
Figure 5.7 Six residue <i>Win</i> motif-based peptide ( ${}_{\text{Ac}}\text{ARTEVY}_{\text{NH}_2}$ ) binds WDR5 with low nanomolar affinity.....	180
Figure 5.8 MALDI-TOF based inhibition data for the peptides identified from phage display screen.....	184
Figure 5.9 Structure of peptide 3 bound to WDR5.....	187
Figure 5.10 Crystal structure of peptide 3 bound to WDR5.....	188
Figure 5.11 Summary of interactions observed in peptide 3-WDR5 co-crystal structure....	191
Figure 5.12 Structure of peptide 6 bound to WDR5.....	193
Figure 5.13 Crystal structure of peptide 6 bound to WDR5.....	194
Figure 5.14 Summary of interactions observed in peptide 6-WDR5 co-crystal structure....	196
Figure 5.15: Structural superposition of Peptide-3 and MLL3 <i>Win</i> motif peptide structures	197
Figure 5.16: Structural superposition of Peptide-6 and MLL3 <i>Win</i> motif peptide structures	198
Figure 5.17 Chemical structure of compounds identified in virtual screen.....	201
Figure 5.18 MALDI-TOF based inhibition assays for the chemical compounds identified from virtual screen.....	202
Figure 5.19 MALDI-TOF based inhibition data for MLL3-FITC-TAT peptide.....	204
Figure 5.20 MLL3-FITC-TAT peptide is readily taken up by P19 cells.....	206
Figure 5.21 MLL3-FITC-TAT peptide induces nuclear defects in P19 cells.....	207

Figure 5.22 Global levels of H3K4 trimethylation is decreased in the presence of MLL3-FITC-TAT.....	210
---	-----

## LIST OF TABLES

Table 3.1 Thermodynamic binding parameters for the different MLL1 <i>Win</i> motif peptides...	85
Table 3.2 Summary of thermodynamic binding parameters for the binding of human SET1 family <i>Win</i> motif peptides to WDR5.....	88
Table 3.3 Isoelectric points (pI) for human SET1 family <i>Win</i> motif peptides at pH 7.0.....	106
Table 4.1 Summary of the X-ray data collection statistics for the different human SET1 family <i>Win</i> motif peptides.....	117
Table 4.2 Summary of the refinement statistics for the different human SET1 family <i>Win</i> motif peptides.....	118
Table 4.3 Pair-wise RMSD values for the superimposition of WDR5 molecules only from <i>Win</i> motif peptide bound structures.....	148
Table 5.1 Thermodynamic binding parameters for the Ac-10-mer and six residue <i>Win</i> motif-based peptides.....	168
Table 5.2 Summary of the X-ray data collection statistics for Ac-10mer, peptide 3 and peptide 6.....	172
Table 5.3 Summary of the refinement statistics for Ac-10mer, peptide 3 and peptide 6.....	173
Table 5.4 Sequence and IC <sub>50</sub> values for peptides identified using Phage display screen.....	182

# CHAPTER 1: INTRODUCTION

## BACKGROUND AND SIGNIFICANCE

Mixed Lineage Leukemia constitutes a heterogeneous category of rare acute leukemias that are characterized by a mixed population of poorly differentiated lymphoid and myeloid progenitor cells. The mixed lineage leukemia-1 (*MLL1*) gene, also known as HRX or ALL-1, is a frequent site of genetic rearrangements in infant acute leukemias and therapy-related malignancies (1) and since its discovery (2-4), significant progress has been made in understanding its role in leukemogenesis (5, 6). Chromosomal abnormalities involving the *MLL1* gene include reciprocal chromosomal translocations, internal partial tandem duplications (PTD), and amplifications of un-rearranged *MLL1* (7). Recurrent *MLL1* translocations account for >70% of infant acute leukemias (both ALL and AML) and are also observed in approximately 10% of de novo AML in adults (8), and in therapy-related leukemias that develop in patients treated with topoisomerase II inhibitors (8-10). The second common *MLL1*-rearrangement, internal partial tandem duplication (*MLL1*-PTD) was first observed in de novo AML patients with a normal karyotype or trisomy 11 (11). *MLL1*-PTDs are found in 4-7% of the cases of AML and present a cytogenetically normal rearrangement that is associated with poor prognosis (12-15). The third common aberration, amplifications of the *MLL1* gene, including trisomy 11 and intrachromosomal amplifications, are found with less frequency in AML and other myelodysplastic syndromes and are associated with a complex karyotype and poor prognosis (16-22). All these chromosomal aberrations disrupt the normal function of the *MLL1* gene, the protein product of which is a histone H3 lysine 4 (H3K4) specific methyltransferase that is essential for the regulated expression of HOX genes during embryonic development and

hematopoiesis. Despite the growing body of literature on the important biological roles of MLL1 and its involvement in human leukemias, development of MLL1 targeted therapeutics has been lacking. Therefore, there is an immediate need for a structural and functional characterization of MLL1, an approach that would facilitate the development of novel therapeutic strategies for the treatment of MLL1-associated leukemias. Although in recent years the introduction of broadly based chemotherapeutic interventions such as all-trans retinoic acid, cytosine arabinoside, histone deacetylase and DNA methyltransferase inhibitors has increased the survival times in some leukemia patients (23-25), molecular therapies that specifically target MLL1 are still lacking. The research carried out as a part of dissertation begins to address this knowledge gap by identifying protein structural features responsible for the H3K4 methyltransferase activity of MLL1. As a result of this research, we expect to identify novel therapeutic inhibitors that have the potential to target MLL1's function in cancer cells and epigenetically reprogram them to normal phenotypes.

### ***1.1 MLL1 belongs to the SET1 family of lysine methyltransferases***

The Mixed Lineage Leukemia protein-1 (MLL1) is a member of the SET1 family of lysine methyltransferases that regulate the degree of histone H3 lysine 4 (H3K4) methylation in eukaryotes (26). H3K4 methylation is an evolutionarily conserved epigenetic mark required for the recruitment of enzymes that heritably maintain transcriptionally permissible states of chromatin (27-31). Recent studies have shown that the Suppressor of Variegation, Enhancer of Zeste and Trithorax (SET) domain of MLL1 catalyzes monomethylation of H3K4 (H3K4me1) (32, 33), a mark associated with nucleosomes in distal enhancer sequences and silenced genes (34-37). However, since H3K4 di- and trimethylation (H3K4me<sub>2,3</sub>) are associated with



transcriptional competence (38, 39), the activity of MLL1 alone is insufficient to promote transcription. MLL1 interacts with an evolutionarily conserved sub-complex that includes WD40-repeat protein-5 (WDR5), retinoblastoma binding protein-5 (RbBP5), absent-small-homeotic-2-like protein (Ash2L), and dumpy-30 (DPY-30) (40-48). This complex (called WRAD) catalyzes H3K4 methylation on its own (32, 33, 49, 50), and has been shown to catalyze H3K4me2 when in complex with MLL1 (33). MLL1 when in complex with WRAD is known as the MLL1 core complex that is required for the di/trimethylation of H3K4 (32, 33, 51). These results suggest that the degree of H3K4 methylation is regulated in a sequential fashion by different enzymes within a multi-subunit complex- together called the MLL1 core complex (32, 33).

MLL1 is one of six human SET1 family members that include MLL2, MLL3, MLL4, SETd1a and SETd1b (41, 42, 47, 48, 51-56). Each of these enzymes regulate H3K4 methylation levels and are assembled into large multi-subunit complexes that include WRAD (41, 42, 51, 57). The presence of WRAD within each of these complexes suggests a common mechanism of complex assembly and enzymatic activity regulation among different SET1 family members. Indeed, while deletion of MLL1 affects methylation at only a subset of genes (58), deletion of WDR5 affects global H3K4me2,3 levels (51, 59), consistent with a more global role for WRAD in regulating H3K4 methylation.

### ***1.2 WDR5 is central for the assembly and H3K4 methylation activity of MLL1 core complex***

The WD-40 repeat protein WDR5 is a conserved component of MLL1 family complexes ranging from yeast to humans and has been shown to be crucial for the assembly and H3K4 dimethylation activity of MLL1 core complex (51, 60, 61). Consistent with the role of WDR5 in

the regulation of the degree of H3K4 methylation, siRNA mediated knock-down of WDR5 in mammalian cells results in a global decrease in the levels of H3K4 di- and trimethylation and exhibit down-regulated expression of HOXA9 and HOXC8 genes that are also correlated with defects in hematopoiesis and development (51, 59). WDR5 knock down in *X. laevis* tadpoles and MLL1 $\Delta$ SET mice show similar developmental defects suggesting that MLL1 and WDR5 function together to regulate H3K4 di/trimethylation and HOX gene expression *in vivo* ((59, 62). Furthermore, growing evidence suggest that WDR5 interacts directly with MLL1 or other SET1 family members and functions as a scaffold to bridge the interactions between MLL1 and rest of the components of MLL1 core complex (51, 59, 61, 63). Indeed, MLL1 and WRAD enzymatic complexes fail to associate in the absence of WDR5 (51, 61), suggesting a crucial role for WDR5 in the assembly and H3K4 methylation activity of the MLL1 core complex.

### ***1.3 Proposed roles for WDR5 in MLL1 core complex***

Several models have been proposed to explain the role of WDR5 within the MLL1 core complex (63). While the earlier studies have suggested that WDR5 functions within the MLL1 core complex as a histone “effector” or “presenter” domain (Figure 1.1a) (59, 63-67), more recent studies by Patel et al., (60, 61) and Song and Kingston., (68) suggest an alternate role for WDR5. According to this new model, WDR5 recognizes a conserved arginine containing motif in the N-SET region of MLL1 called the WDR5 interaction (*Win*) motif (61), which is required for the assembly and the H3K4 dimethylation activity of the MLL1 core complex (Figure 1.1b) (60, 61, 68).

In the histone effector domain model, WDR5 functions within the MLL1 core complex to specifically recognize dimethylated H3K4 (59) and is required for the subsequent conversion of di- to trimethylation at H3K4 by the MLL1 core complex. Based on this study, the final product-H3K4 trimethylated peptide has a weaker affinity to WDR5 compared to that of H3K4 dimethylated peptide. Because of this weaker affinity, the MLL1 core complex is released from the nucleosome bearing H3K4 trimethylation and binds to a neighboring nucleosome that has dimethylated H3K4. WDR5 therefore functions as a histone effector that helps in the propagation of H3K4 trimethylation signal to the neighboring nucleosomes (59).

In contrast to the preferential binding of WDR5 to dimethylated H3K4 suggested by the histone effector model, the histone presenter model suggests that WDR5 binds the N-terminus of histone H3 and presents the lysine 4 side chain for further methylation by the MLL1 SET domain (Figure 1.1a) (51, 67). Evidence for this hypothesis comes from the binding studies that demonstrate that WDR5 does not discriminate H3K4 peptides based on the absence or presence of methyl groups (51, 67). In support of the binding studies, the crystal structures of WDR5 in complex with histone H3 peptides (unmodified, mono-, di-, or trimethylated at H3K4) presented in that investigation reveal that the first three N-terminal residues (ART) of the histone H3 sequence in each of the structures make extensive contacts with WDR5 while the lysine 4 (H3K4) side chain does not make any significant interactions with WDR5 (67). Therefore, the substrate lysine (H3K4) is not a part of the WDR5 binding site and presence or absence of methyl groups on the lysine side chain will not affect its WDR5 binding affinity. Furthermore, mutations in WDR5 that disrupt its histone binding affinity also decrease the histone

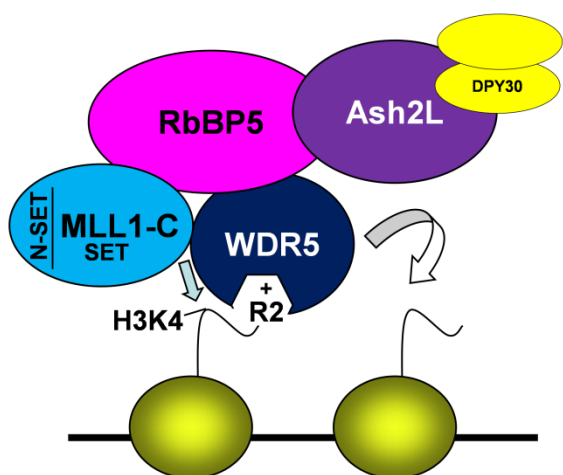
methyltransferase activity when assembled into the MLL1 core complex in place of wild-type WDR5 (67) suggesting that WDR5 functions as a histone presenter domain.

Based on these models, it has been predicted that both histone H3 and MLL1 bind WDR5 simultaneously but at distinct sites (Figure 1.1a) to catalyze H3K4 methylation. Evidence for the histone binding role for WDR5 within the MLL1 core complex have come from the studies that have shown that peptides derived from histone H3 can bind free WDR5 (59) and hence expected to bind WDR5 within the context of MLL1 core complex. While these studies that have shown that histone H3 peptides can bind free WDR5 (59, 64-67), a direct interaction between histone H3 and WDR5 when in complex with MLL1 was not demonstrated in these studies. Therefore, it is unclear if WDR5 can bind histone H3 when it is a part of the MLL1 core complex.

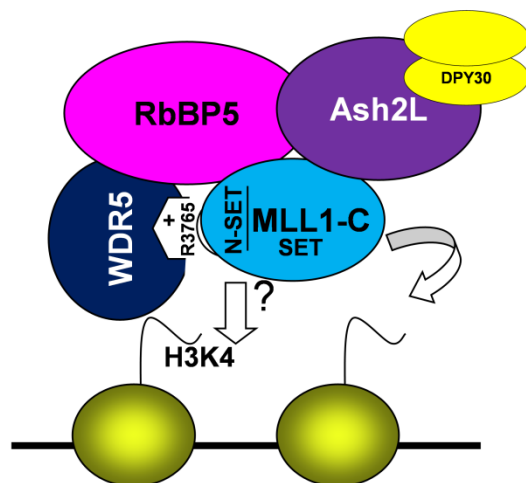
In contrast to the models that suggest a role for WDR5 in binding histone H3, several previous observations have suggested that WDR5's function within the MLL1 core complex could be different. Using a recombinantly purified 180-kDa C-terminal fragment of MLL1 (MLL1-C) that contains the SET domain, Dou *et al.* (51) demonstrated that WDR5 is essential for the assembly of the minimal MLL1 core complex, which includes MLL1-C, WDR5, RbBP5, and Ash2L. Loss of WDR5 disrupts the ability of MLL1-C fragment to co-immunoprecipitate with the other core complex components, RbBP5 and Ash2L (51). Interestingly, Dou *et al.* (51) also observed that point mutations in WDR5 (S91K and F133A) that were previously thought to be important for binding the arginine 2 of histone H3, also disrupted the ability of MLL1-C fragment to co-immunoprecipitate with the rest of the core complex components. While these unexpected results were attributed to local conformational changes in WDR5 that prevents its

association with MLL1-C at a distinct site (51), recent studies by Patel et al (60, 61) and Song and Kingston (68) suggest an alternate possibility. Using a combination of domain-mapping, analytical ultracentrifugation and crystal structures, Patel et al (60, 61) and Song and Kingston (68) have independently demonstrated that WDR5 recognizes an arginine containing histone H3 like sequence present in the N-SET region of MLL1-C (Figure 1.1b). Therefore it is possible that the inability of WDR5 histone binding mutants to co-immunoprecipitate MLL1-C observed in the study by Dou et al (51) could be due to the disruption of MLL1 interaction surface in WDR5.

## a. Existing Model



## b. New Model



**Figure 1.1 Schematic models for the role of WDR5 in the MLL1 core complex.**

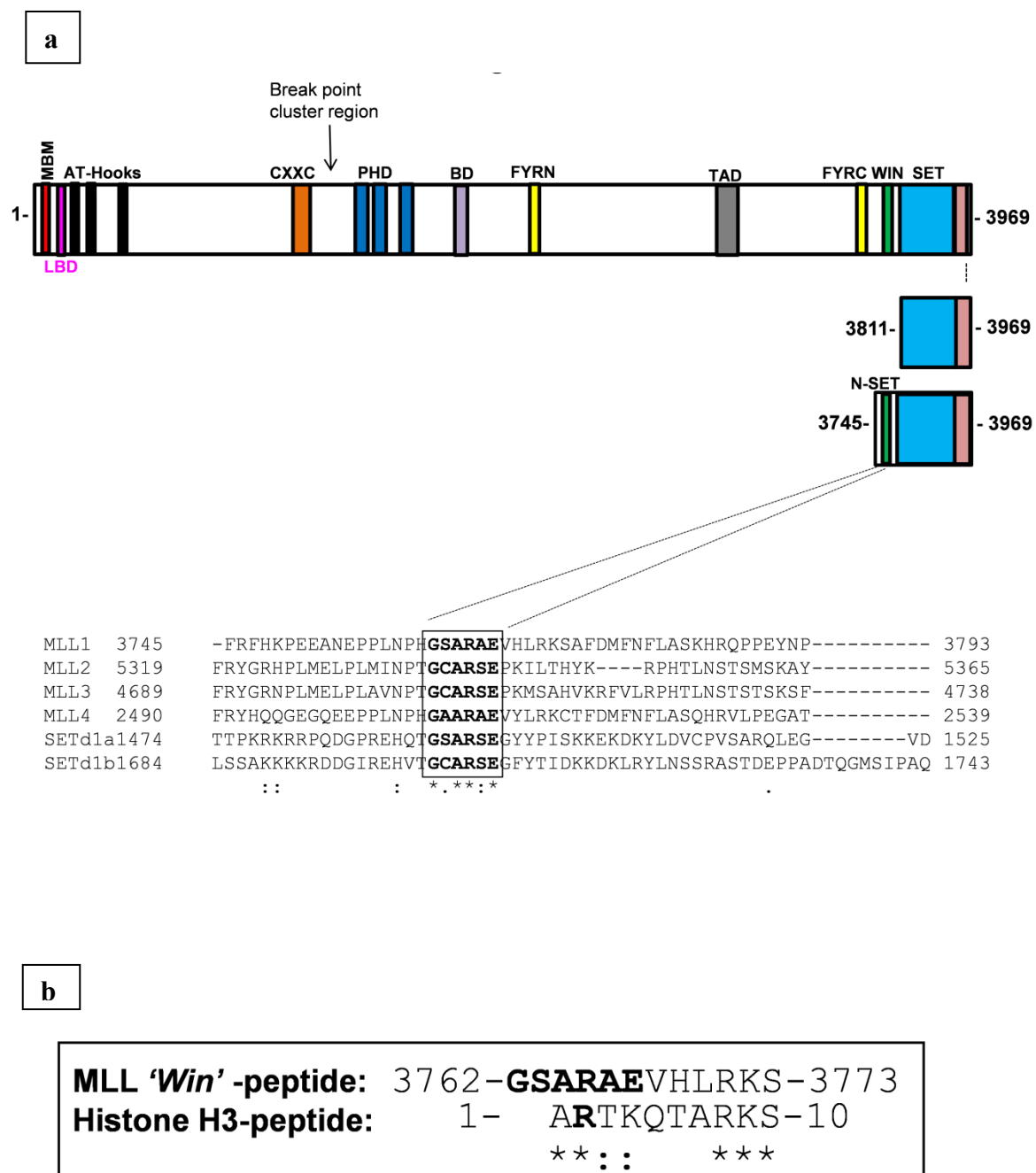
Nucleosomes are shown in *yellow* with the histone H3 N-terminal tail indicated. *a*, histone binding model where WDR5 recognizes Arg-2 of histone H3 and facilitates H3 methylation by presenting the Lys-4 side chain to the SET domain of MLL1. *b*, a new model based on the present data in which WDR5's recognition of Arg-3765 of the MLL1 *Win* motif is required for the assembly and H3K4 dimethylation activity of the MLL1 core complex. MLL1 SET domain and the N-SET regions are indicated. The figure was adapted from Patel et al (61) and modified accordingly.

#### ***1.4 WDR5 recognizes a conserved arginine containing sequence in the N-SET region of MLL1***

Previous studies that were focused on mapping the interaction region between WDR5 and MLL1 led to the identification of three distinct regions in MLL1 that may be required for its interaction with WDR5 (40, 69). In the study by Dou et al, the minimum MLL1 fragment (residues 3301–3969) that was shown to interact with WDR5 included the N-SET region, the evolutionarily conserved SET domain and the Post-SET domain (40). Similarly, Yokoyama et al (48) in their efforts to map the WDR5 binding region in MLL1 showed that the deletion of C-terminal 149 residues (3821-3969), which harbors the SET domain resulted in the loss of co-immunoprecipitation of WDR5, RbBP5 and Ash2L with MLL1. In contrast to these studies that suggest that SET domain is required for the interaction with WDR5, Lee and Skalnik (69) demonstrated that a region corresponding to the N-SET region (amino acid residues 1450-1537) of SETD1a protein, a human MLL1 paralog, is required for the co-immunoprecipitation of WDR5, RbBP5, and Ash2L. While these studies suggest that both the N-SET region and the SET domain may harbor binding sites for WDR5, the precise location of the WDR5 binding site in MLL1 or its family members were previously uncharacterized. To address this knowledge gap and to map the precise location of the WDR5 binding site, Patel et al (61) generated two constructs in MLL1, namely the MLL1<sup>3745</sup> (encompassing residues 3745-3969 and contains a 66 amino acid N-SET region that precedes the catalytic SET domain and the Post-SET domain ) and MLL<sup>3811</sup> (encompassing residues 3811-3969 and contain only the SET domain and the Post-SET domain) (Figure 1.2) and tested the ability of these recombinantly expressed proteins to bind WDR5 in sedimentation velocity analytical ultracentrifugation experiments. Using these constructs, Patel et al (61) demonstrated that the MLL1<sup>3745</sup> construct containing the N-SET region binds to WDR5 as a 1:1 complex with a dissociation constant of 120 nM. The MLL1<sup>3745</sup>

construct was also shown to be the minimal construct that is required for the *in vitro* reconstitution of MLL1 core complex (MLL<sup>3745</sup>, WDR5, RbBBP5, Ash2L, and DPY30) (32, 61). This minimal MLL1 core complex can efficiently catalyze mono- and dimethylation of H3K4, an activity that has been attributed to the association of the two independent methyltransferases MLL1-SET and WRAD (32, 61). In contrast to MLL1<sup>3745</sup>, the shorter construct MLL1<sup>3811</sup> that lacks the N-SET region fails to associate with WDR5 and the rest of the core complex components in sedimentation velocity analytical ultracentrifugation experiments (61). MALDI-TOF mass spectrometry based methylation assays show that while the MLL1<sup>3811</sup> construct can efficiently catalyze H3K4 monomethylation on its own, it is defective for H3K4 dimethylation due to its inability to associate with the WRAD sub-complex (61). Collectively, these studies demonstrate that MLL1 harbors a WDR5 interaction motif within in the N-SET region and suggest that WDR5 functions as a structural platform to bridge the interactions between the MLL1-SET domain and rest of the core complex components- RbBP5, Ash2L and DPY30 (61).





**Figure 1.2 WDR5 binds to a conserved arginine containing sequence in the N-SET region of MLL1**

## Figure 1.2 WDR5 binds to a conserved arginine containing sequence in the N-SET region of MLL1

*Top*, conserved features in MLL1 are indicated as described in Figure 2.2. Menin binding motif MBM (red), LEDGF binding domain or LBD (pink), DNA binding AT-hooks (black), zinc finger containing CXXC motifs (orange), plant homeodomain (PHD) fingers (blue), bromodomain (BD) (purple), phenylalanine-tyrosine rich regions (FYR-N and -C) (yellow), transcriptional activation domain (grey) WDR5 interaction (*Win*) motif (light green), histone methyltransferase SET domain (light blue) and the Post-SET domain (beige) are highlighted.

*Middle*, MLL1 constructs MLL1<sup>3811</sup> (amino acid residues 3811-3969) and MLL1<sup>3745</sup> (amino acid residues 3745-3969) used in the study by Patel et al (61) are shown. Refer to text for more detail.

*Bottom*, ClustalW2 multiple sequence alignment (70) for the *Win* motif peptides from human SET1 family members MLL1 (UniProtKB/Swiss-Prot accession number Q03164), MLL2 (UniProtKB/Swiss-Prot accession number O14686), MLL3 (UniProtKB/Swiss-Prot number Q8NEZ4), MLL4 (UniProtKB/Swiss-Prot accession number Q9UMN6), SETd1a (UniProtKB/Swiss-Prot accession number O15047), and SETd1b (UniProtKB/Swiss-Prot number Q9UPS6). BLOSUM matrix (71) was used for the alignment. Conserved residues are denoted underneath the alignment by an *asterisk* (\*), conservative substitutions are denoted by a *colon* (:), and semiconservative substitutions are denoted by a *period* (.). The amino acid sequences of conserved *Win* motifs are *boxed*. In b, sequence alignment between the MLL1 *Win* motif and the first ten residues of the histone H3 peptides are shown.

In order to identify amino acid residues within the N-SET region that are required for the interaction between MLL1 and WDR5, Patel et al (61) performed an amino acid sequence alignment of mammalian SET1 family members that have been experimentally shown to interact with WDR5 (40-42, 48, 51, 57). Extensive amino acid sequence analysis of the N-SET region among human MLL1 family members revealed that while the amino acid sequence of the N-SET region is generally highly variable across the paralogs, a six residue sequence (containing MLL1 residues GSARAE) alone is highly conserved among the different MLL1 family proteins (Figure 1.2a). This sequence designated the WDR5 interaction (*Win*) motif closely resembles the sequence surrounding arginine 2 of histone H3 (Figure 1.2b) and is highly conserved among metazoan MLL1 orthologs and other SET1 family members (Appendix Figures 1.1-1.6) (61, 68). Using a combination of analytical ultracentrifugation and MALDI-TOF mass spectrometry, it was further demonstrated that the conserved arginine (R3765) of the MLL1 *Win* motif is crucial for the interaction with WDR5 and that substitution of R3765 of MLL1<sup>3745</sup> with alanine abolishes the interaction. MLL1 (R3765A) substitution also results in the loss of the H3K4 dimethylation activity of the MLL1 core complex (61). However, replacement of other residues such as serine 3763 and glutamate 3767 in the conserved *Win* motif sequence only modestly weakens MLL1's interaction with the WRAD sub-complex (61). In addition, Patel et al (60) also demonstrated that a 12-residue peptide derived from the *Win* motif in MLL1 (encompassing residues 3762-3773) can bind WDR5 with strong affinity ( $K_d = 1.7 \mu\text{M}$ ). Addition of the excess of this MLL1 *Win* motif peptide disrupts the interaction between MLL1<sup>3745</sup> and WRAD and results in the loss of H3K4 dimethylation activity of the MLL1 core complex (61). While the *Win* motif peptide represents the major interaction surface between WDR5 and MLL1 (as predicted by the R3765A mutation), the difference in WDR5 binding affinities observed between

MLL1<sup>3745</sup> (120 nM) and the MLL1 *Win* motif peptide (1.7 μM) suggests that the SET domain also contributes to additional interaction affinities (48, 61).

Previously published crystal structures of WDR5 in complex with histone H3 peptides reveal that arginine 2 (R2) of histone H3 inserts into a central cleft in WDR5 and is essential for the histone peptide interaction with WDR5 (67). Since the sequence of the MLL1 *Win* motif closely resembles the sequence of histone H3 peptide, Patel et al (60, 61) hypothesized that MLL1 interacts with WDR5 using the same arginine binding pocket. Interestingly, crystal structures of WDR5 in complex with the MLL1 *Win* motif peptide (60, 68) reveal that R3765 from the MLL1 *Win* motif is present in the same arginine binding pocket of WDR5 previously thought to bind histone H3 peptides (51, 64-67). The crystal structures also revealed that amino acid residues surrounding the *Win* motif arginine participate in more favorable interactions with WDR5 when compared to that of histone H3 peptides (60, 68). WDR5 mutants (S91K and F133A) that line the arginine binding pocket failed to associate with MLL1<sup>3745</sup> in sedimentation velocity experiments suggesting that WDR5 utilizes the same pocket to bind to the MLL1 *Win* motif and histone H3 (61). Collectively, these results suggest a new role for WDR5 within SET1 family complexes, which is its ability to bind to the *Win* motif and also argue against the role of WDR5 as a histone presenter domain, at least when it is in complex with MLL1. While these studies do not rule out the possibility that WDR5 can bind histone H3 in the absence of MLL1, the high interaction affinity between WDR5 and MLL1-*Win* motif suggests that the existing models that describe WDR5's role within the MLL1 core complex as a histone binding module needs to be revised (Figure 1.1).

The significance of the interaction between the *Win* motif and WDR5 in cells is highlighted by the evolutionarily conserved nature of the *Win* motif sequence among metazoan SET1 family members (Refer to Figure 1.2 and appendix Figures 1a-1f), and by the recent demonstration that the conserved arginine (R5340) in the MLL2 *Win* motif has been found to be among a spectrum of MLL2 missense amino acid mutations associated with Kabuki syndrome (72, 73). It has also been demonstrated that peptides derived from the MLL1 *Win* motif disrupts the formation of the MLL1 core complex *in vitro* and abolishes the H3K4 dimethylation activity of the MLL1 core complex (61). These results suggest that *Win* motif peptides or related compounds may be useful for targeted therapies for treatment of malignancies that result from gain-of-function mutations in human SET1 family members (33). For example, amplifications of the MLL4 gene (previously known as MLL2) is associated with solid tumors (74, 75). In addition, a cytogenetically normal rearrangement of the MLL1 gene found in ~10% of acute myeloid leukemias results in a partial tandem duplication of N-terminal MLL1 sequences that retains the conserved SET domain (11, 33, 76, 77). These rearrangements display increased H3K4 methylation, lysine acetylation and HOX gene expression and may be responsive to targeted inhibition (78-80). An understanding of how different human *Win* motif sequences interact with WDR5 will increase our knowledge of how SET1 family complexes are assembled and regulated, and will facilitate the rational design of novel targeted therapies for MLL1 related malignancies.

**Purpose:**

The main goals of the experiments described in this dissertation are to carry out a detailed thermodynamic and structural analysis of the interaction between WDR5 and human SET1

family *Win* motifs. These experiments are focused on identifying novel protein structural features which enable WDR5 to differentially recognize different human SET1 family members. Isothermal titration calorimetry (ITC) is initially used to compare the binding affinities of different SET1 family *Win* motifs to WDR5. Three-dimensional structures of WDR5 bound to SET1 family *Win* motif peptides are then used to generate hypotheses to explain the observed differences in binding affinities. These structure-function hypotheses are then tested using synthetic peptides harboring individual amino acid substitutions.

The data presented in this dissertation are divided into two main Chapters (3-4) that deal with thermodynamic and structural characterization of the interaction between WDR5 and six different human SET1 family *Win* motif based peptides (MLL1-4, SETd1a and SETd1b). In Chapter 2, we have reviewed the three most common chromosomal abnormalities associated with MLL1 and introduce three novel targeted therapies that have emerged from the biochemical characterization of the MLL1 protein. In Chapter 3, experiments are described that show that the conserved arginine 3765 of the MLL1 *Win* motif is required for co-immunoprecipitation of WDR5, RbBP5 and Ash2L from mammalian cells, confirming its critical role in the assembly of the MLL1 core complex. In addition, ITC experiments that measure the binding affinities of different SET1 family *Win* motif peptides to WDR5 are described. In the first set of ITC experiments, MLL1 *Win* motif peptides of varying lengths are compared. A representative MLL1 *Win* motif peptide is then chosen to compare with other SET1 family *Win* motif peptides of similar length in the second set of ITC experiments. Differences in binding affinities can be the result of amino differences within or outside the conserved *Win* motif sequence and these possibilities are tested using experiments described in Chapter 4. The final set of experiments in

Chapter 3 describes MALDI-TOF based methylation assays that test the ability of different *Win* motif peptides to inhibit the H3K4 dimethylation activity of the MLL1 core complex. MALDI-TOF assays were grouped into two categories that are classified as with or without the pre-incubation of MLL1 with the WRAD sub-complex. In Chapter 4, three-dimensional structures of WDR5 bound to six different human SET1 family *Win* motif peptides are reported. Based on this structural analysis, additional *Win* motif peptides bearing single amino acid substitutions are synthesized and their inhibition constants are compared to wild-type *Win* motif peptides. In Chapter 5, the results of the experiments described in Chapters 3-4 are summarized and discussed in a broader context of identifying novel *Win* motif-based therapeutic approaches to treat MLL1-associated leukemias. Preliminary data for the identification and characterization of two additional *Win* motif-like inhibitors (Ac-10-mer and six-residue *Win* motif peptides) derived from structure-function studies is presented. This analysis is extended further to identify additional peptidomimetics that target MLL1-WDR5 interaction using a randomized phage display library screen. The results of additional experiments using chemical library screening approach to identify small molecule inhibitors that target MLL1 core complex are also described in Chapter 5. Based on the experiments described in Chapters 3-4, the MLL3 *Win* motif peptide was identified as the best inhibitor for the H3K4 dimethylation activity of MLL1 core complex and hence was chosen for the preliminary *in vivo* experiments described in Chapter 5. *In vivo* experiments describing intracellular localization, nuclear morphology and changes in global levels of H3K4 methylation are reported in Chapter 5. Collectively, the experiments described in this dissertation suggest that *Win* motif based peptides represent an excellent starting point for the design of inhibitors that would disrupt the interaction between MLL1 and WDR5 and inhibit the H3K4 dimethylation activity of MLL1 core complex *in vivo*. The knowledge gained from the

structure-function analyses of WDR5-SET1 family *Win* motif interaction will facilitate the design of novel class of “*Win* motif-based” methylation inhibitors that may be useful for the treatment of MLL1-related malignancies.



## **Chapter 2: Targeted therapy for Mixed Lineage Leukemia-what have we learned from the biochemistry of MLL1**

This Chapter is a review of the literature for the most common MLL1 aberrations and the pathogenesis with these malignancies. It also includes a review of three unique targets identified from the biochemical studies using MLL1 protein and their potential therapeutic value in MLL1 associated leukemia. The material presented here was published as a part of reference (81).

## **Introduction**

The mixed lineage leukemia (*MLL1*) gene, also known as HRX or ALL-1, is a frequent site of genetic rearrangements in infant acute leukemias and therapy-related malignancies (1). Chromosomal abnormalities involving the *MLL1* gene include reciprocal chromosomal translocations, internal partial tandem duplications (PTD), and amplifications of un-rearranged *MLL1* (7). These chromosomal aberrations are associated with mechanistically distinct gain-of-function phenotypes that may be amenable to targeted therapeutic approaches. However, progress in this area has been impeded by a lack of understanding of the molecular details by which *MLL1* translocations, amplifications and PTDs contribute to leukemogenesis. To date, more than 60 *MLL1* fusion partners have been described (8), and detailed genetic/biochemical studies have identified several functional domains within chimeric *MLL1*-fusion proteins that are essential for leukemic transformation (1, 5, 8, 82-89). Although our understanding of the molecular pathology of *MLL1*-associated leukemias remains incomplete, recent biochemical and structural information is contributing to an evolution of potential treatment strategies from a broadly-based chemotherapeutics approach towards therapies targeted to the underlying molecular pathogenesis of leukemia (23). This Chapter reviews recent advances in our efforts to develop novel *MLL1*-targeted therapies.

## **Mixed Lineage Leukemia -1 (MLL1): a master epigenetic regulator with multiple roles in transcription**

### ***2.11 MLL1 in embryonic development and hematopoiesis***

The mixed lineage leukemia-1 (*MLL1*) gene located at chromosome 11, band q23, encodes a histone H3 lysine 4 (H3K4) methyltransferase that functions to maintain gene expression during development and hematopoiesis (54, 90, 91). The best studied target genes of *MLL1* include the

homeobox transcription factors or HOX genes, which are important for segment identity and cell fate during metazoan development (92, 93). Genetic studies in mice have demonstrated that the homozygous knock out of MLL1 is embryonic lethal and is associated with multiple developmental defects including neural crest patterning and hematopoietic abnormalities (90, 91, 94, 95). Notably, the expression levels of several HOX genes including HOXA4, HOXA7, HOXA9, and HOXA10 are decreased in MLL1<sup>-/-</sup> mice and are associated with defects in fetal liver hematopoiesis (90, 96). MLL1<sup>+/-</sup> mice are not embryonic lethal but are anemic and exhibit homeotic developmental defects that are related to posterior shifts in HOX gene expression patterns (90, 94). Likewise, MLL1 is also required for adult hematopoiesis and stem cell self-renewal (97). In MLL1 conditional knockout mice, fetal hematopoiesis is unaltered; however, adult mice exhibit anemia, a significant reduction in the number of bone marrow hematopoietic stem cell progenitors, and poor survival rates (98). In addition, it has been demonstrated that MLL1 plays a crucial role in self-renewal in cultured fetal liver and adult bone marrow stem cells (97, 99). These studies suggest that MLL1 orchestrates its biological functions at least in part through the regulation of HOX genes. Indeed, HOX dysregulation is a common phenotype that underlies the pathogenesis of acute leukemias associated with alterations in the MLL1 gene (5, 79, 100-102). However, MLL1 is also required for the regulation of cell cycle dependent genes such as cyclins A, B, and E (103); CDK inhibitors p16<sup>Ink4a</sup>, p18, p27 (103, 104); E2F family of transcription factors (E2F2, E2F4 and E2F6) (103), as well as the transcription factor GATA3, which plays an essential role in specifying lymphoid subtype (105). In addition, MLL1 regulates expression of several genes involved in organogenesis and differentiation (106, 107). Therefore, MLL1 is a master regulator that is critical for many gene expression programs required for normal development, hematopoiesis and the cell cycle.

## ***2.12 Chromatin Structure dynamics and Transcription***

In eukaryotes, DNA is condensed into highly ordered structures known as chromatin- the structure of which is dynamically altered according to the needs of the cell. The basic repeating unit of chromatin is the nucleosome, which is composed of ~146 base pairs of DNA wrapped around an octameric disc of histone proteins containing two copies each of histones H2A, H2B, H3 and H4 (108). As a first level of DNA compaction, the nucleosomal units are positioned along the length of the DNA to form 11-nm chromatin fibers, which represents the ‘beads-on-a-string’ structure (109). Further chromatin compaction, either through binding of linker histones (H1) or through inter-nucleosomal interactions packages the nucleosomal DNA into higher order chromatin fibers that are refractory to gene expression, DNA replication and repair machineries (110, 111). Cellular processes that require access to DNA often use enzymes that dynamically regulate the structure of chromatin either through recruitment of adaptor proteins or additional enzymatic machineries that alter the positioning of nucleosomes on DNA (112-115). Key mechanisms employed by these enzymes include incorporation of histone variants, introduction of post-translational modifications on histones and the underlying DNA, and binding of non-histone proteins that can dynamically alter the structure of chromatin. Among these mechanisms, reversible post-translation modification of histones has been more extensively studied owing to the variety of modifications (including lysine and arginine methylation; lysine acetylation, ubiquitination, and sumoylation; and serine phosphorylation) present in the cell and the combinatorial complexity that arises from the co-existence of these modifications as defined by the “Histone code hypothesis” (115). The complexity is further increased by histone lysine residues that can mono-, di-, or trimethylated with each modification correlating to different

transcriptional outcomes (39, 116-121). Histone lysine methyltransferases characterized to date are highly substrate-specific (lysine residue modified) and product-specific (their ability to characterize mono-, di-, or trimethylation), and participate in a variety of cellular processes, including transcriptional regulation, heterochromatin formation, and DNA repair (122). A cell's differential gene expression program is maintained and inherited through different epigenetic states, which are correlated with distinct types of histone post-translational modifications. (123)

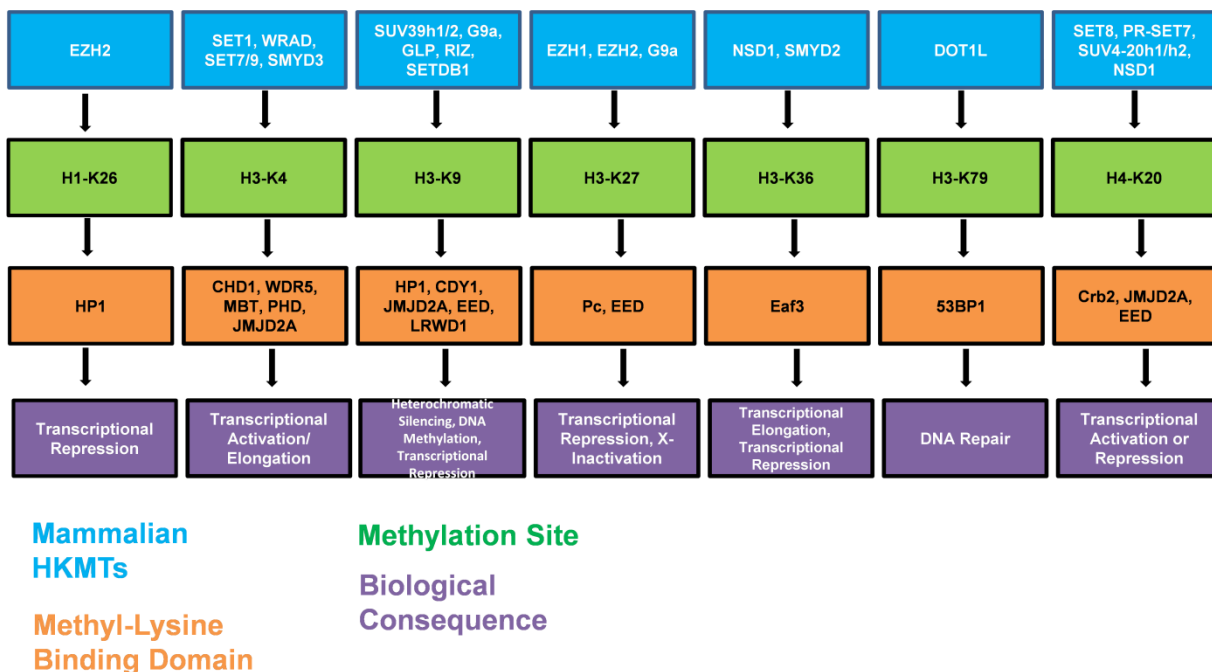
### ***2.13 H3K4 methylation in transcriptional control***

Methylation of lysine 4 of histone H3 (H3K4) is an evolutionarily conserved epigenetic mark predominantly associated with transcriptional activation in eukaryotes (116, 124-127). The epsilon amino group of lysine 4 can be mono-, di-, or trimethylated, with each modification correlating with distinct transcriptional outcomes (39, 117-121). For example, genome-wide chromatin immunoprecipitation studies have demonstrated that high levels of H3K4 trimethylation are present within the 5' regions of actively transcribed genes (39, 119-121). It has been demonstrated that H3K4 trimethylation functions to recruit ATP dependent nucleosome remodeling enzymes that increase promoter DNA accessibility by sliding or displacing nucleosomes (29-31). H3K4 dimethylation is spread more evenly across the coding regions of genes and is thought to be associated with a transcriptionally "poised" state of chromatin (116, 120, 121). In contrast, H3K4 monomethylation is enriched at the 3' ends of the genes and distal enhancer sequences, and is associated with ribosomal DNA (rDNA) and telomeric silencing (34-37, 39, 116, 119-121, 128). These studies suggest that the degree of H3K4 methylation is a highly regulated process. Indeed eukaryotes have evolved a number of highly conserved enzymes whose function appears to precisely regulate the degree of H3K4 methylation.

### **2.14 H3K4 methylation is catalyzed by SET1 family of lysine methyltransferases**

H3K4 methylation is mainly deposited by a group of enzymes that share an evolutionarily conserved SET (SuVar, E(z), Trithorax) domain (26), although a new H3K4 methyltransferase lacking a SET domain has recently been reported (32, 33). While there are several SET domain enzymes that differ with regard to their substrate specificity (26) (refer to Figure 2.1 for SET domain substrate specificity), members of the SET1 family share the properties that they all methylate H3K4 and all interact with an evolutionarily conserved core group of proteins that function to regulate the degree of H3K4 methylation. The budding yeast *S. cerevisiae* enzyme, SET1p, was the first H3K4 methyltransferase to be identified (34, 44, 46, 129) and carries out the bulk of H3K4 methylation in yeast, a function that can be attributed to the presence of a catalytic SET domain (34). In yeast, SET1p is recruited to the promoters of actively transcribing genes through its interaction with the phosphorylated carboxy terminal domain (CTD) of RNA polymerase II (118). In metazoans, SET1 family is more diverse with humans encoding six different SET1 family proteins, including the SET1p orthologs SETd1a/SETd1b, and the Mixed Lineage Leukemia proteins MLL1-4 (41, 42, 47, 48, 51-55, 59). Each of these proteins catalyzes H3K4 methylation of distinct subset of genes, with MLL1 and MLL2 playing crucial roles in the regulated expression of HOX genes (45, 52, 54, 94) that specify segment identity during vertebrate development. In contrast, the H3K4 methylation function of MLL3 and MLL4 are required for nuclear hormone receptor dependent gene expression (130-132) and immunoglobulin class-switching (133). It is becoming increasingly clear from recent studies that SETd1a/b maintains global H3K4 trimethylation levels (134, 135) while the activity of the other metazoan paralogs is required for the regulation of more specific subsets of genes. The non-

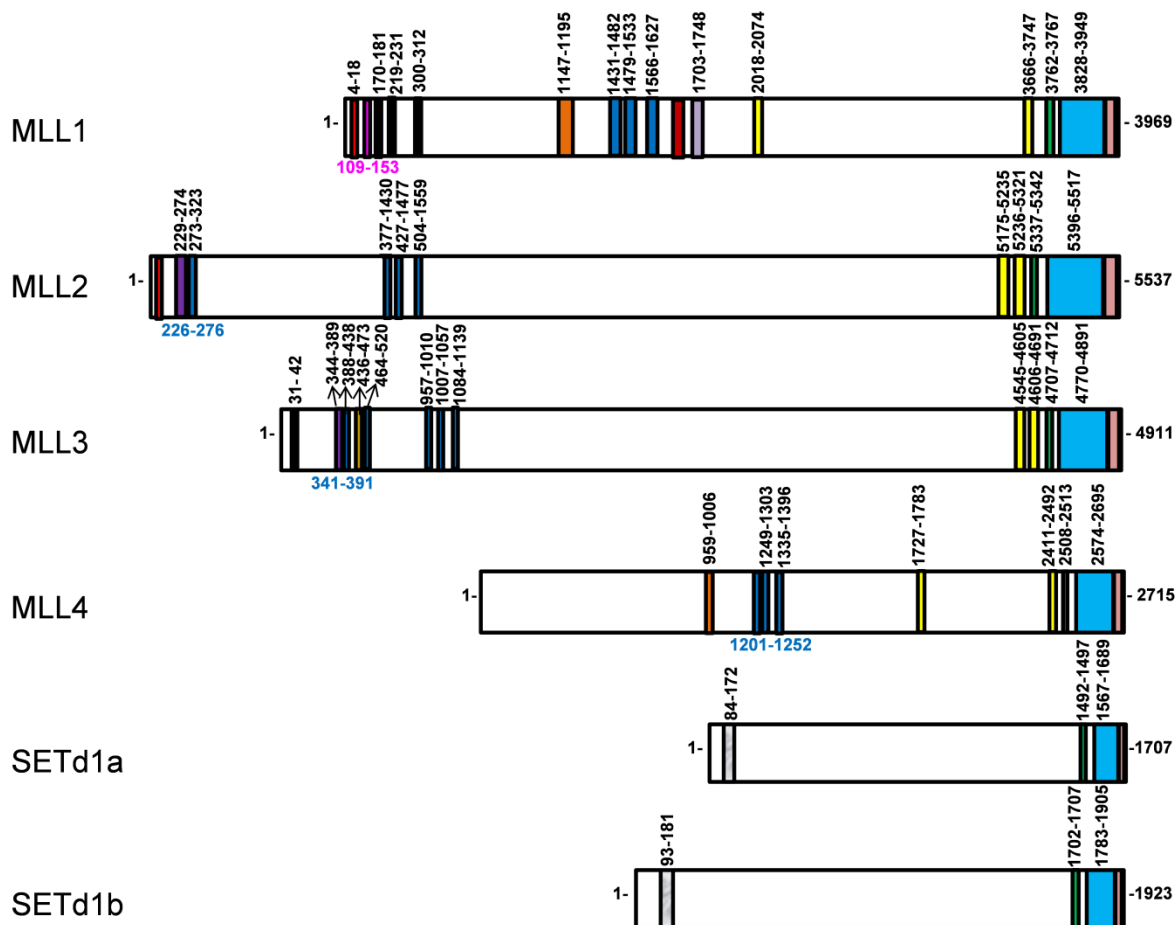
redundant functions of each SET1-family member can likely be attributed to their differential domain architectures outside the conserved catalytic SET domain (refer to Figure 2.2 for the domain architecture of human SET1 family members). These domains differ in their ability to interact with proteins that can regulate its catalytic activity and target gene specificity of SET1 family enzymes. Despite their ability to regulate the expression of only a subset of genes, each SET1 family member is highly essential for development and mutations that disrupt their normal function are associated with a variety of disease phenotypes. For example, recent exome sequencing studies have identified MLL2 genetic mutations as an underlying phenotype in Kabuki Syndrome (72, 73), a rare human multiple malformation disorder. In addition, amplifications of the MLL4 gene (previously known as MLL2) are associated with solid tumors (74, 75).



**Figure 2.1: Substrate specificity of mammalian histone lysine-methyltransferases (HKMTs)**















Seven major lysine sites methylated by human histone methyltransferases and the biological processes regulated by these reversible posttranscriptional modifications are summarized. The methyltransferases are indicated in blue, the substrate lysine is indicated in green, the methyl lysine binding domain is indicated in orange, and the biological process regulated by each specific modification is indicated in purple. The figure was adapted from (136) and (137) and modified to include more recent literature.





**Figure 2.2 Schematic representations of human SET1 family methyltransferases**

The domain architectures of six different human SET1 family methyltransferases are shown. The regions of the various putative functional domains are mapped (residue numbering indicated on top or bottom). ScanProsite (138) was used to annotate the domain regions for MLL1 (UniProtKB/Swiss-Prot accession number Q03164), MLL2 (UniProtKB/Swiss-Prot accession number O14686), MLL3 (UniProtKB/Swiss-Prot number Q8NEZ4), MLL4 (UniProtKB/Swiss-Prot accession number Q9UMN6), SETd1a (UniProtKB/Swiss-Prot accession number O15047), and SETd1b (UniProtKB/Swiss-Prot number Q9UPS6). The various domains are color coded as shown below. The different domains are drawn to scale of the size of the full length protein.

 Menin Binding Motif (MBM)	 WDR5 Interaction motif (Win)
 LEDGF Binding Domain (LBD)	 SET Domain
 AT-Hooks	 Post-SET domain
 CXXC Motif	 RING domain
 Plant Homeo Domain (PHD)	 RNA Recognition Motif (RRM)
 Bromo Domain (BD)	 DHHC motif
 FY Rich Regions	 Gene Internal Promoter

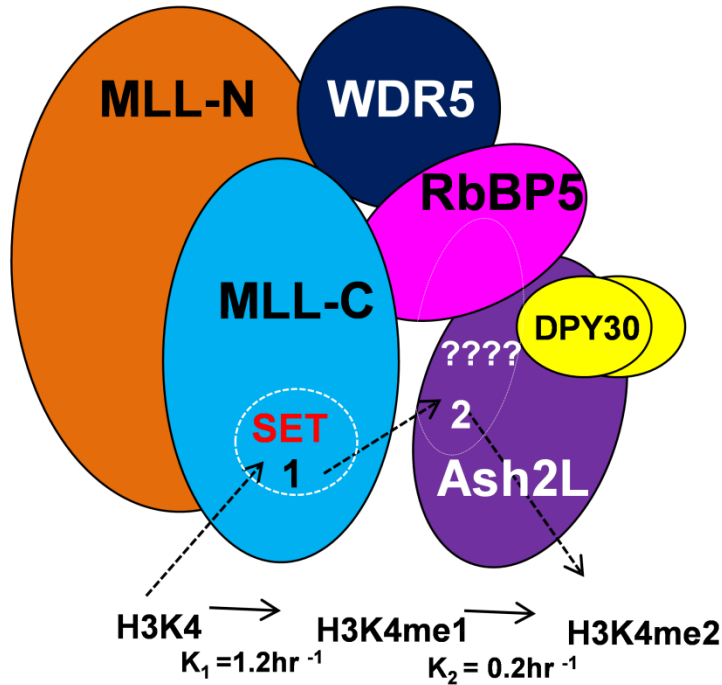
### Legend for Figure 2.2

MLL1 is more extensively characterized than the other homologs owing to its association with a variety of leukemias and evidence suggests that its transcriptional co-activator function is mediated in part by the enzymatic activity of its SET domain and through protein-protein interactions (26, 54). For example, homozygous deletion of the MLL1 SET domain in mice, while not embryonic lethal, exhibits skeletal defects and altered expression of several HOX genes that partially phenocopy the heterozygous knockout of the whole MLL1 gene (62). These changes are correlated with decreased levels of mono-, di- or trimethylation of H3K4 and deregulated DNA methylation patterns at several HOX gene promoters (62). MLL1 functions within a large macromolecular complex with more than 30 subunits that regulate the degree of H3K4 methylation and MLL1's target gene specificity (139). While subunit composition of different SET1 family members varies to some degree, each SET1 family member interacts with the conserved core group of proteins that form the WRAD sub-complex (42, 51, 57) (For a detailed list of interaction partners of human SET1 family proteins refer to appendix Figures 2a-f). It has been shown recently that the WRAD sub-complex possesses an intrinsic histone methyltransferase activity on its own (32, 33, 50). Evidence suggests that both the MLL1 SET domain activity and WRAD's monomethyltransferase activity are required for the methylation of nucleosomal H3K4 (32, 51). These studies have led to a model in which H3K4 methylation is sequentially catalyzed by a complex that contains multiple distinct active sites for the addition of each methyl group (32). The presence of WRAD within each of the SET1 family complexes suggests a common mechanism of complex assembly and enzymatic activity regulation among different SET1 family members.

### ***2.15 Mechanism of multiple lysine methylation catalyzed by MLL1 core complex***

Previously it was thought that mono-, di-, and trimethylation of H3K4 could be attributed to the SET domain dependent methyltransferase activity of MLL1 alone and that the WRAD complex functions merely as an allosteric regulator of MLL1 (47, 51, 65, 67, 140-143). However, since it has more recently been demonstrated that the WRAD complex dimethylates H3K4 in a manner that is independent of the enzymatic activity of the MLL1 SET domain (32), the allosteric model needs to be revised. Indeed, *in vitro* methylation assays demonstrate that the isolated MLL1 SET domain is predominantly a monomethyltransferase, which can be attributed to the presence of a conserved tyrosine residue in the SET domain active site (32). Furthermore, loss of the WDR5, RbBP5 or Ash2L subunits of WRAD results in the loss of di- and trimethylation of H3K4 both *in vivo* and *in vitro* without significant changes in H3K4 monomethylation (32, 51, 59, 61). These results are consistent with a sequential mechanism whereby the MLL1 SET domain catalyzes H3K4 monomethylation and the WRAD enzyme catalyzes H3K4 dimethylation within the MLL1 core complex. However, WRAD lacks the ability to dimethylate H3K4 without MLL1, suggesting that MLL1 amino acid sequences, distinct from the MLL1 SET domain active site, contributes to the WRAD active site within the MLL1 core complex. That a complex between WRAD and MLL1 is required for H3K4 dimethylation is supported by the demonstration that amino acid substitutions that disrupt the interaction between MLL1 and WRAD also disrupt the H3K4 dimethylation activity of the MLL1 core complex (61). Therefore, the completely assembled MLL1 core complex is required for efficient H3K4 dimethylation and for nucleosome methylation (33). The existence of a sequential mechanism utilizing several active sites for multiple lysine methylation suggests that the degree of H3K4 methylation is more highly

regulated than previously appreciated (Refer to Figure 2.3 for the mechanism of H3K4 methylation by MLL1 core complex).



**Figure 2.3 Proposed model for the mechanism of multiple lysine methylation catalyzed by MLL1 core complex**

According to this model, the first methylation (H3K4me1) site is present in MLL1 SET domain (at site 1) and the monomethylated H3 peptide is then transferred to the second active site (site 2) on the WRAD-subcomplex, which catalyzes the dimethylation event (H3K4me2). The catalytic motif within site 2 (indicated with a white question mark) located on WRAD is not characterized yet. The rate constants for mono- ( $K_1$ ) and dimethylation ( $K_2$ ) catalyzed by the fully assembled MLL1 core complex are indicated. The rate constants were reported in (32).

### **2.16 WRAD components are associated with development and oncogenesis**

One of the common features of the greater than 60-MLL fusion proteins associated with leukemias is that they all lose the amino acid sequences required for interaction with WRAD. This raises questions about WRAD's role in MLL1 associated oncogenesis. WRAD subunits are conserved within SET1 family complexes ranging from yeast to humans (41, 42, 47, 51, 57) and have been shown to play essential roles in cellular differentiation (144-146), development (59, 147), dosage compensation (148, 149), and transcription (150). In addition to their role in regulating the H3K4 dimethylation activity of MLL1 core complex (and presumably the other SET1 family members), components of the WRAD complex have also been implicated in recruiting SET1 complexes to chromatin through their ability to associate with other proteins, with DNA and with RNA. For example, the Ash2L component can recruit SET1 family complexes to their genomic targets through its association with sequence-specific transcription factors, such as Mef2d and Ap2delta (150, 151). The N-terminal PHD domain of Ash2L also harbors a putative winged-helix DNA binding motif (152, 153). Similar to Ash2L, the WDR5 component has also been shown to interact with the transcription factor Oct4 (154), which can recruit SET1 complexes to gene targets that play crucial roles in stem cell self-renewal. In addition, WDR5 has been shown to associate with long intergenic noncoding RNA (lincRNA) such as HOTTIP in vitro, an interaction that was required for MLL1/WDR5 recruitment and H3K4me3 at HOXA promoters (155). Collectively, these studies suggest that WRAD components can function to recruit SET1 family enzymes to target genes.

WRAD enzyme lacks sequence homology to known methyltransferase folds, and as a result relatively little is understood about its methyltransferase activity and the role it might play

in oncogenesis. WRAD's potential role in oncogenesis is supported by the observation that the Ash2l component of WRAD is overexpressed at the protein level in many human tumors, and that knockdown of Ash2L inhibits tumor cell proliferation (156). In addition, the RbBP5 component of WRAD has been shown to be amplified in several glioblastomas, suggesting that it may be a novel oncogene (157). Further studies will be required to better understand WRAD's role in MLL1 associated leukemogenesis and other cancers.

In contrast, the MLL1 component of the MLL1 core complex is more extensively characterized and several functional domains implicated in transcriptional regulation have been identified (for recent reviews refer to (106, 139)). Furthermore, MLL1 is a frequent site of chromosomal alterations that sometimes disrupt the functions of these domains.

### ***2.17 MLL1 functional domains implicated in transcriptional regulation***

The MLL1 gene encodes a large protein of 3,969 amino acid residues and contains several functional domains including: menin binding motif (MBM), LEDGF binding domain (LBD), DNA-binding AT hooks, a cysteine-rich CXXC DNA binding motif, plant homeodomain (PHD) fingers, a bromo domain (BD), a transactivation domain (TAD), a WDR5 interaction (Win) motif, and a C-terminal histone methyltransferase SET domain (Figure 1.2) (139). The full-length MLL1 protein, synthesized as a single transcript, is cleaved by taspase I into MLL1-N (320 kDa) and MLL1-C (180 kDa) fragments, which then re-associate non-covalently through the FYRN and FYRC motifs to form the functional MLL1 complex *in vivo* (158-160). The mature MLL1 protein assembles into macromolecular complexes with several regulatory proteins that are essential for MLL1's transcriptional co-activator properties. Biochemical and



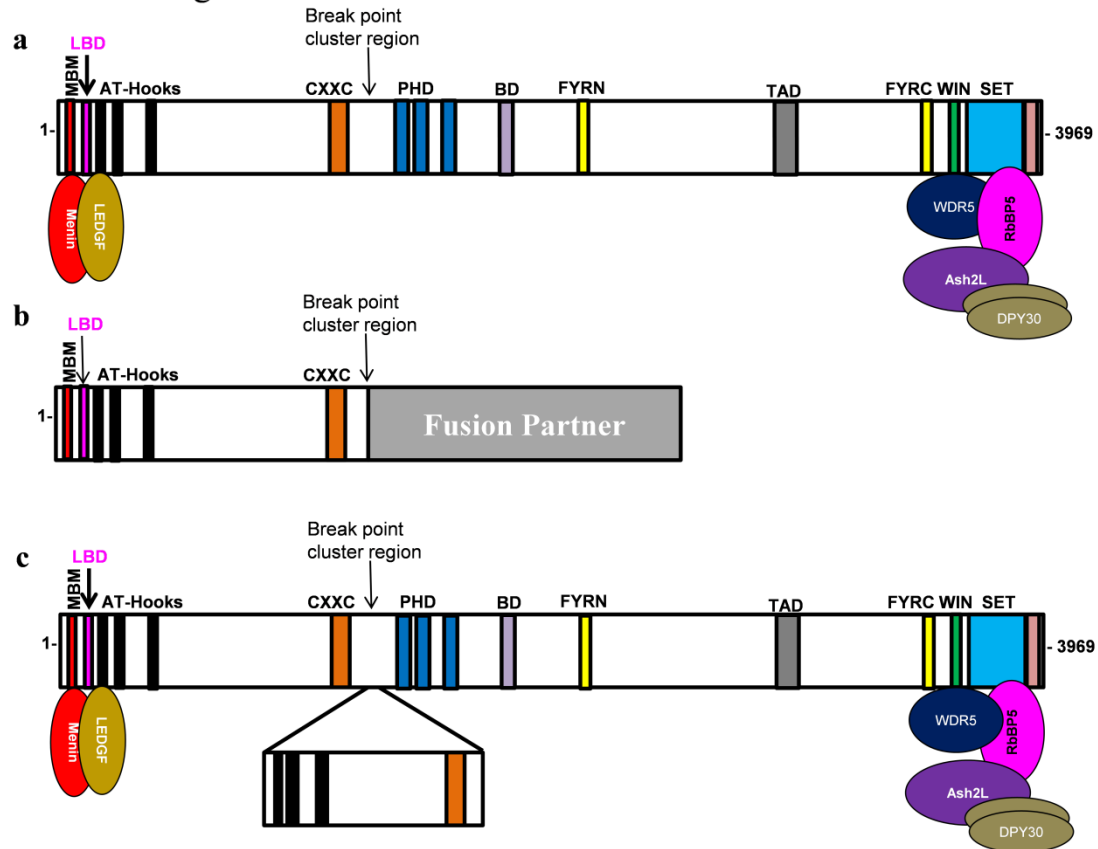
genetic studies have identified several direct and indirect interaction partners for MLL1 that include; the menin tumor suppressor protein (48, 53); cell cycle regulators such as E2Fs and HCF-1 (48, 161); polycomb group proteins, BMI-1 and HPC-2 (162); histone deacetylases (45, 162); nuclear cyclophilin , Cyp33 (162); acetyltransferases such as p300, CBP and MOF (40, 163); chromatin remodeling factors, INI1/SNF5 (164); and WDR5/RbBP5/Ash2L and DPY30, which are core components of SET1 family methyltransferases (51). In addition, a recent paper describes the identification of a gene internal promoter that transcribes the C-terminal half of MLL1 (107), the function of which is not known. Interestingly, the gene internal promoter coincides with an MLL1 breakpoint region, suggesting that the pathogenesis of MLL1 translocation induced leukemia could be due in part to the loss of this N-terminally truncated form of MLL1 when the breakpoint is 5' to the gene internal promoter. However, in cases where the MLL1 breakpoint is 3' to the gene internal promoter, it could result in aberrant expression of the fusion protein (107). Indeed, it has been noted that the common MLL1 translocation partners AF4, AF6, and ENL have potential AUG start codons in positions where they could be transcribed within the context of the translocated MLL1 gene (107). These studies suggest that the transcriptional activator properties of MLL1 are mediated through multiple functional domains through protein-protein and protein-DNA interactions. Many of these interactions are retained in leukemogenic MLL1s, some participating in gain-of function phenotypes, making them candidates for molecular targeted therapy. Recent technical advances that attempt to target these interactions as potential therapeutic targets to treat MLL1-associated leukemia are reviewed in sections 2. 5-2.8. In sections 2.2-2.4, I review the three most common genetic alterations associated with the MLL1 gene.

## ***2.2 Acute myeloid and lymphoblastic leukemia with 11q23 abnormalities***

### ***2.2.1 Incidence and clinical Significance of 11q23 Chromosomal translocations***

The MLL1 gene located at chromosome 11, band q23, is frequently involved in reciprocal translocations found in several cases of acute myeloid (AML) and acute lymphoblastic leukemia (ALL) (2, 3, 165) and identify a patient sub-population with a poor prognosis (1). Recurrent MLL1 translocations account for >70% of infant acute leukemias (both ALL and AML) and are also observed in approximately 10% of de novo AML in adults (8) and in therapy-related leukemias that develop in patients treated with topoisomerase II inhibitors (8-10). Chromosomal translocations fuse the N-terminal part (~1400 amino acids) of the MLL1 gene resulting in a protein product with an in-frame fusion to one of more than 60 partner proteins that range from nuclear factors to cytoplasmic proteins (1, 166, 167).

The five most common MLL1 translocations include: MLL1-AF4 or t(4;11)(q21;q23), MLL1-ENL or t(11;19)(q23;p13.3), MLL1-AF9 or t(9;11)(p23;q23), MLL1-AF10 or t(10;11)(p12;q23), and MLL1-AF6 or t(6;11)(q27;q23) and account for greater than 80% of MLL1-rearranged leukemias (6, 168-170). In addition, chimeric MLL1-fusions involving ELL, EEN, GAS7, AF1p, AFx, Septins, and histone acetyltransferases CBP/p300 have also been reported (4, 8, 14, 169, 171-177). The translocation partners identified to date are diverse and do not share any biochemical function or structural motifs. However, all known MLL1 fusion proteins share the property that the N-terminal portion containing the AT hooks and CxxC domains of MLL1 are retained, suggesting the preservation of DNA binding activity (178, 179). In contrast, the taspase cleavage site, gene internal promoter, TAD domain, PHD fingers, *Win* motif and the SET domain are lost (Figure 2.4b) (5).



**Figure 2.4: Schematic representation of the most common genetic alterations associated with the MLL1 gene**

The putative protein products generated by the two most common chromosomal alterations associated with the MLL1 gene are indicated along with wild-type MLL1: a) wild-type MLL1 b) reciprocal chromosomal translocations involving MLL1 c) partial tandem duplications (PTDs) in MLL1. The various functional domains are color coded as in figure 1.2. Chromosomal translocations fuse the N-terminal ~1400 amino acids of MLL1 in-frame to over one of the 60 different fusion partners (indicated in grey) (b). PTDs have a duplicated N-terminus (AT-hooks and CXXC motifs) in addition to all the functional domains present in wild-type MLL1 (c). Refer to text for a detailed description of MLL1 mutations.

### ***2.2.2 Pathogenesis of acute leukemias with MLL1-translocations***

While it is expected that the loss of the SET domain in MLL1 translocations would result in decreased H3K4 methylation and Hox gene expression, genetic studies have revealed that the individual fusion partners possess transcriptional activator properties and are indispensable for leukemogenesis (171, 176, 180, 181). Given the complexity of different translocation partners, MLL1-fusions may activate a common leukemia-associated gene expression program through multiple mechanisms. Indeed, AF4, AF5, AF9, ENL, ELL and AF10 proteins are all implicated in transcriptional elongation via association with the EAF complex, pTEFb kinase, and hDOT1 mediated methylation of H3K79 (86, 88, 182-185). Another mechanism could involve transcriptional activation via increased or aberrant histone acetylation (MLL-CBP/p300) (85, 186), protein arginine methyltransferase-1 (PRMT1) association (MLL-EEN) (187), SWI-SNF chromatin-remodeling complex recruitment (MLL-ENL, -AF9, -AF10)(89, 188, 189), and self-association or dimerization of the N-terminal part of MLL1 (MLL-GAS7, -AF1p, -beta-galactosidase, -gephyrin, -SEPT6) (84, 172, 176, 190).

Regardless of the mechanism, aberrant expression of MLL1 target genes are a common feature of MLL1-rearrangements examined to date. For example, HOXA7, HOXA9, and the HOX cofactor MEIS1 are consistently over expressed in human leukemias with MLL1-translocations (100, 101, 191-193) and act, at least partially, through the activation of the proto-oncogene c-Myb (194). Furthermore, retroviral co-transduction studies in mice have demonstrated that HOXA9 and MEIS1 expression immortalizes hematopoietic progenitors in vitro and rapidly accelerates leukemia development (195). In addition, MLL1-fusion proteins fail to transform bone marrow cells in which HOXA7 and HOXA9 expression are genetically ablated (101).

Similarly, it has been demonstrated that overexpression of HOXA9 and MEIS1 can replace the leukemogenic activity of MLL1-ENL (193). Collectively, these results suggest that HOXA9 dysregulation is an important factor in some MLL1-fusion induced leukemias (101). However, not all leukemogenic fusions result in HOXA9 dysregulation. For example, MLL1-GAS7 and MLL1-AF9 fusions were shown to transform bone marrow cells or mice that do not express HOXA9 (196, 197).

However, in addition to dysregulation of HOX genes, other signaling pathways are perturbed by MLL-translocations and may contribute to leukemogenesis. For example, transcriptional deregulation of FMS-like tyrosine kinase 3(FLT3), glycogen synthase kinase 3 (GSK3), heat shock protein-90 (HSP-90), myeloid cell leukemia sequence-1 (MCL-1), and components of the RAS pathway have been implicated in MLL1-induced leukemogenesis (198-206). MLL1's role as a master regulator of gene expression significantly complicates understanding its role in MLL1 associated leukemogenesis.

The master regulatory role of MLL1 in transcriptional control has implications that affect our normal understanding of malignancy. For example, it has been suggested that second-hit mutations are required to initiate the full leukemia phenotype (176). Indeed, recent studies have identified mutations in p53, ATM, Ras, and FLT3 genes in MLL1 leukemia patients (207-210). However, because of MLL1's role in epigenetic gene control, second hit mutations could also arise in the form of epigenetic mutations that result in silencing of tumor suppressors genes without changes in their DNA sequence. For example, it has been demonstrated that the FHIT tumor suppressor gene is epigenetically silenced in human primary tumor cells and tumor cell

lines with MLL1-translocations (211). These data suggest that epigenetic alterations may be just as important as genetic mutations in sources of so called “2<sup>nd</sup> hit” mutations that underlie the pathogenesis of leukemia. Recent advances in deep sequencing technologies such as RNA-SEQ and CHIP-SEQ will likely allow us to better distinguish genetic versus epigenetic mutations in future studies.

### ***2.3 Partial tandem Duplications: A cytogenetically normal rearrangement in MLL1***

#### ***2.3.1 Clinical significance of MLL1 partial tandem duplications (MLL1-PTDs)***

The second common MLL1-rearrangement, internal partial tandem duplication (MLL1-PTD) was first observed in de novo AML patients with a normal karyotype or trisomy 11 (11). MLL1-PTDs are found in 4-7% of the cases of AML and present a cytogenetically normal rearrangement that is associated with poor prognosis (12-15). MLL1-PTDs result from an in frame fusion of exons 11-5 or 12-5 upstream of exon 5, partially duplicating sequences in the 5' end of MLL1 (11, 13, 76). The protein product of MLL1-PTDs has a duplicated N-terminus that contains an additional AT-hook and CXXC domains while essentially retaining all the conserved domains in wild-type MLL1 (13, 76). In contrast to the variety of MLL1-fusions that delete the MLL1 C-terminus, MLL1-PTDs retain the 3'-portion of the gene that encodes the SET domain methyltransferase motif (Figure 2.4c). Recent data suggests that the enzymatic activity of the MLL-PTD SET domain participates in a gain-of function phenotype in AML.

#### ***2.3.2 Mechanism of leukemic transformation by MLL1-PTDs***

The molecular mechanisms that underlie MLL1-PTD transformation in AML are currently unknown. The crucial alteration is the duplication of AT hooks and CXXC DNA binding motifs,

which may alter target gene specificity. It has been suggested that duplication of these segments may also mimic dimerization observed in several MLL1 chimeric fusion proteins (190). Indeed, it has been demonstrated that a synthetic MLL1 N-terminal construct containing duplicated AT-hook and CXXC domains possesses potent transactivation activity in luciferase reporter assays (190). However, comparison of genome-wide gene expression data shows that MLL1-PTD primary cells have gene expression patterns that are distinct from that of cells bearing MLL1 chimeric fusions suggesting that the mechanism underlying transformation is distinct (212). Alternatively, it has been suggested that the additional amino acid sequences in MLL1-PTD might adopt a conformation that potentially interferes with the normal function of MLL1 by distancing the regulatory domains from its target site (7).

There is a growing body of evidence that suggests that epigenetic alterations underlie the pathogenesis of MLL1-PTDs. For example, it has been demonstrated that knock-in mice bearing  $MLL1^{PTD/WT}$  exhibit increased expression of HOXA7, HOXA9, and HOXA10 that is associated with increased H3K4 methylation and H3/H4 acetylation within these promoters (79). Similar gene expression and H3K4me2 methylation changes were seen in the presence and absence of the wild type MLL1 allele in primary MLL1-PTD mouse fetal liver cells, suggesting that MLL1-PTD behaves as a dominant gain-of-function mutation (78). This phenotype may also be due to other epigenetic alterations. For example, Whitman et al., (80) have demonstrated that the wild type MLL1 allele in  $MLL1^{PTD/WT}$  AML cells is silenced in a manner that is associated with hypoacetylation of histones H3 and H4. Treatment of MLL1-PTD cells with histone deacetylase inhibitors partially reactivates wild type MLL1 expression and reduces AML blast colony forming units (213). In addition, it has been demonstrated that the SLC5A8 tumor suppressor

gene is silenced in MLL-PTD cells in a manner that is associated with increased DNA methylation in its promoter, a phenotype that is partially reversed with DNA methyltransferase inhibitors (213). SLC5A8 encodes a membrane monocarboxylate transporter that regulates intracellular concentrations of histone deacetylase inhibitors, such as butyrate and pyruvate (214, 215). Together, these observations suggest that targeting MLL1-PTD may have therapeutic value in the treatment of AML (80). Indeed, it has been demonstrated that down regulation of MLL1-PTD using antisense oligodeoxynucleotides (aODNs) in primary human MLL-PTD AMLs results in reactivation of the wild type MLL1 gene, reduced AML blast-derived colony forming units, and increased sensitivity to cell death (213).

#### ***2.4 Acute myeloid leukemia with amplifications in MLL1***

Amplifications of the MLL1 gene, including trisomy 11 and intrachromosomal amplifications, are found with less frequency in AML and other myelodysplastic syndromes and are associated with a complex karyotype and poor prognosis (16-22). Amplifications involving MLL1 result in upregulation of several HOX genes including, HOXA7, HOXA9, and MEIS1 (16, 22). These studies suggest that MLL1-amplifications contribute to leukemogenesis through mechanisms that share some features with that of MLL1-fusions and MLL1-PTDs. In addition, gene expression analyses have identified other proteins that are upregulated in AML with 11q23 amplifications including cell surface receptors PROM1, ADAM10, and NKG2D, and the inosine triphosphatase (ITPA) (22). These data suggest that MLL1 amplifications are associated with a gain-of-function phenotype that may be responsive to targeted therapy (22).



In spite of these advances in our understanding of the pathogenesis of MLL1-linked leukemias, identification of inhibitors that specifically target MLL1 or MLL1-chimeric fusion proteins has so far proven elusive. Recent advances in our understanding of the protein-protein interactions involving MLL1 suggest several novel therapeutic strategies for targeted inhibition of MLL1 or MLL1 chimeric fusion activity in leukemic cells. These advances are reviewed in sections 2.5-2.8 (81) and summarized in Figure 2.4.

## ***2.5 Targeting MLL1-Menin interaction as a therapeutic strategy to treat Mixed Lineage Leukemia***

### ***2.5.1 Menin is a common component of wild-type and mutant MLL1 complexes***

Despite the growing evidence demonstrating a role for HOX genes in hematopoiesis, it is unclear how the array of mutations involving the MLL1 gene all contribute to altered HOX gene expression in acute leukemias. While there is little that is common among the different types of MLL1 aberrations, all types of MLL1 gene mutations including: chromosomal translocations, internal partial tandem duplications, and gene amplifications, retain the N-terminus of MLL1 (1, 87, 216). These ~1400 N-terminal amino acids contain the DNA binding AT hooks and CXXC domains as well as the binding site for nuclear proteins menin (called the Menin Binding Motif or MBM) and Lens epithelial derived growth factor (LEDGF binding domain or LBD) (Figure 2.4). Menin, which directly binds wild-type MLL1 and MLL1-oncogenic fusion proteins, is an essential co-factor for the maintenance of normal hematopoiesis and the leukemogenic activity of MLL1-associated translocations (48, 53, 217, 218). The importance of menin in the pathogenesis

of MLL1-related leukemia and progress on the development of inhibitors that target the MLL1-menin interaction is reviewed in this section.

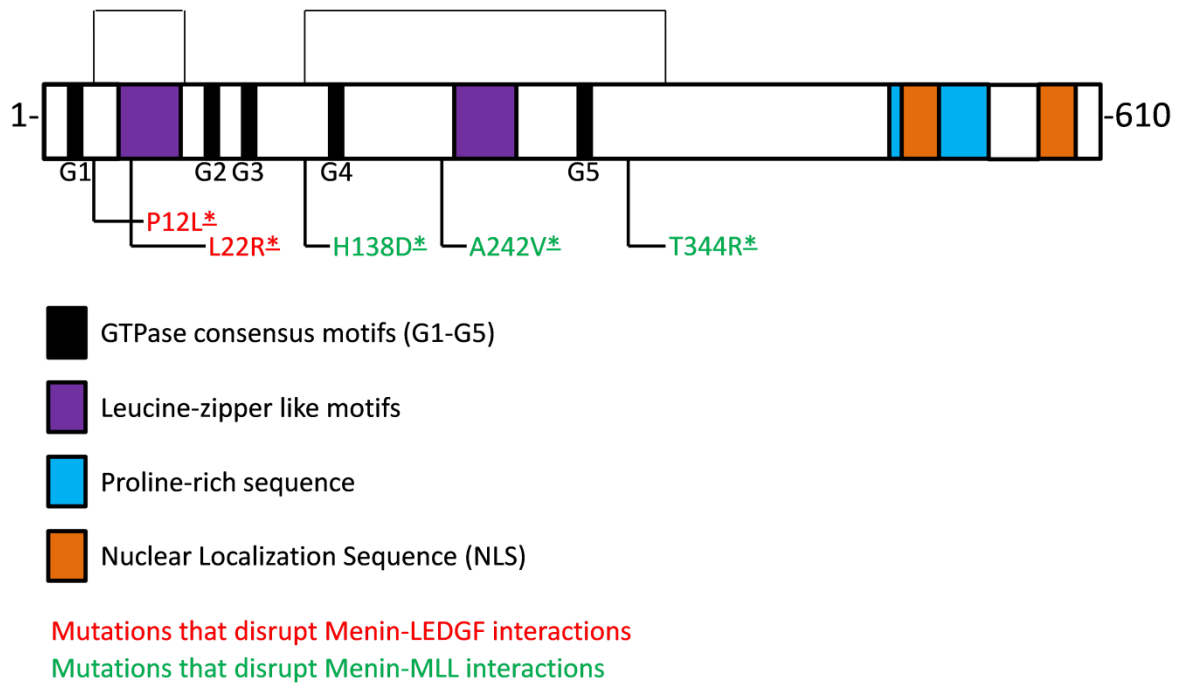
### **2.5.2 *MEN1 tumorigenesis***

Menin, the product of MEN1 gene located at chromosome band 11q13, is a tumor suppressor protein that is mutated in patients with an inherited syndrome, Multiple Endocrine Neoplasia 1 (MEN1) (219-221). To date, more than 400 nonsense and frame-shift mutations have been reported in MEN1 patients often developing parathyroid, pancreatic or pituitary tumors after the loss of the wild-type MEN1 allele (219, 222-224). Homozygous knockout of MEN1 (-/-) is embryonic lethal in mice, which die at the mid-gestation period with profound defects in liver, heart and the neural tube (225-227). Heterozygous knockout mice are viable until the adult stages, but develop tumors similar to human MEN1 syndrome in pancreatic islets, parathyroid, anterior pituitary, adrenal cortex and adrenal medulla (225, 228). Although the loss of menin results in tumors of the endocrine lineage, the MEN1 gene is ubiquitously expressed in most adult tissues and at all developmental stages (221, 229). In spite of the increasing evidence that highlights a role for menin in MEN1 tumors, the basic biology of menin dependent tumor suppression is unclear.

### **2.5.3 *Menin homology and conserved domain architecture***

Menin is highly conserved among vertebrates including, humans, mouse, rat, and zebrafish. However, menin orthologs have not been identified in budding yeast *Saccharomyces cerevisiae* or in the nematode *Caenorhabditis elegans* (226) (230-233). Strikingly, several of the disease-associated MEN1 mutations occur at conserved amino acid positions, highlighting a crucial role for menin in regulating cell proliferation in higher eukaryotes (229, 234). Menin is a novel

nuclear protein of ~610 amino acids and does not share significant sequence homology to any other known proteins (220, 234, 235). However, extensive analysis of sequence alignments among menin homologues has revealed several domains with putative roles in nuclear targeting and transcriptional regulation. Conserved domain search using the human menin amino acid sequence identified putative domains such as: consensus GTPase-like motifs (G1-G5), two leucine-zipper motifs, a proline-rich region and two nuclear localization signals (NLS) (229, 234, 236) (Figure 2.5). However, it is unclear as to how these conserved motifs contribute to the tumor suppressor function of menin.



**Figure 2.5: Domain architecture of human menin showing the putative conserved domains**

Schematic representation of the conserved regions in human menin with the various functional domains indicated in different colors: GTPase consensus-motifs G1-G5 (black), leucine-zipper-like motifs (purple), proline-rich sequence (blue), and nuclear localization signal or NLS (orange). Naturally occurring MEN1 mutations that disrupt its interaction with MLL1 (green) and LEDGF (red) are highlighted. Figure 3 was adapted from Poisson et al., (234) and modified based on the studies by Hughes et al., (53) and Yokoyama and Cleary (237).

#### ***2.5.4 Menin molecular interaction network***

The precise biochemical function of menin has so far proven elusive due to the lack of any known functional domains in menin. However, it has been suggested that the transcriptional regulatory properties of menin are regulated by protein-protein interactions (48, 234, 236-238). Recent biochemical efforts have been undertaken to elucidate the molecular pathways that underlie menin-dependent transcriptional regulation (236). Using yeast two-hybrid, GST pull-down and co-immunoprecipitation assays, menin was shown to interact with a cohort of proteins that are involved in cell cycle regulation, DNA replication and repair, genome stability, endocrine metabolism, bone morphogenesis and hematopoiesis (53, 229, 236, 239). These studies uncovered a wide variety of proteins that may or may not interact directly with menin. These proteins (menin interacting proteins or MIPs) can be grouped into four major functional classes: Class I, which includes transcription factors like JunD (240-242), NFkB (p50, p52 and p65) (240), Smad3 (243), BMP2 (bone morphogenic protein 2) (244), IGFBP-2 (Insulin-like growth factor binding protein 2) (245), FANCD2 (fanconi anemia complementation group D2 protein) (246), Pem (a homeobox containing transcription factor) (247), cMyb (238); Class II, which includes DNA damage and replication proteins such as RPA (replication protein A 1 and 2) (248); Class III, which includes cell cycle regulatory proteins such as CDK inhibitors (p18 and p27) (104), ASK (activator s-phase kinase) (249), type III intermediate filaments (glial fibrillary acidic protein or GFAP and Vimentin) (250); and Class IV, which includes transcriptional activators such as MLL1/2 (mixed lineage leukemia proteins) (48, 53), RNA polymerase II phosphorylated carboxy terminal domain (53), LEDGF (237), and CHD1 (chromo domain helicase I) (217). Based on these studies it has been suggested that MIPs modulate the transcriptional activator/repressor functions of menin (240, 241, 243, 251). While a direct role

for the interaction of menin with all MIPs has yet to be validated in vivo, it is possible that MIPs regulate transcription by binding directly or indirectly to menin (229, 236, 239). In addition, menin also binds to a putative tumor metastasis suppressor/nucleoside diphosphate kinase (Nm23), which stimulates the GTP hydrolyzing activity of menin (252). Moreover, menin also functions as a transcriptional co-activator of the nuclear receptor pathway by binding with estrogen receptor-alpha (ERalpha) in a hormone-dependent manner (253). A non-specific DNA binding activity through the C-terminal NLS has also been reported for menin (245). Of the multitude of interactions reported for menin, it is the association with the SET1 family methyltransferases MLL1/2 that has generated a lot of interest due to their roles in hematopoiesis and leukemia.

### ***2.5.5 Role of menin in Hematopoiesis***

Menin is an essential component of MLL1/2 family complexes with specific roles in the maintenance of HOX gene expression patterns during hematopoiesis (48, 53). Conditional knockouts of the MEN1 gene in mice decreases peripheral white blood cell counts as well as colony forming potential of bone marrow hematopoietic progenitors (217). Recent work by Maillard et al., (254), suggests that while conditional menin knockouts have modest effects on hematopoiesis under steady-state conditions, more severe defects are observed in competitive transplantation assays and during drug-mediated chemoablation (254). These studies suggest that menin functions as an essential regulator of hematopoietic stem cell (HSC) homeostasis specifically in situations of hematopoietic stress (255). These phenotypes may be due, at least in part, to menin's role in regulating HOX gene expression. For example, small-interfering RNA (siRNA) mediated knockdown of menin or conditional MEN1 (-/-) knockout embryos show significant decreases in the expression levels of several HOX genes including HOXA9, HOXC6,

and HOXc8. (48, 53, 217, 239). Interestingly, these hematopoietic defects are rescued by the ectopic expression of menin or its downstream targets HOXA9/MEIS1 (53, 217). Evidence indicates that menin's interaction with MLL1/2 complexes is required for its role in transcription. For example, chromatin immunoprecipitation studies using antibodies specific to menin, MLL1, trimethylated H3K4 or CHD1 (which binds trimethylated H3K4) (28, 29, 256) have further established that menin co-localizes with these components at HOXA9 promoters and is required for transcriptional activation (217, 239). Furthermore, it has been shown that a subset of naturally occurring mutations in menin (H139D, A242V, and T344R) disrupts its association with MLL1 and fails to associate with MLL1-dependent H3K4 methyltransferase activity (53) (refer to Figure 2.4 for MEN1 mutations). These findings underscore the importance of the menin-MLL1 interaction in regulating HOX gene expression in hematopoiesis.

### ***2.5.6 Role of menin in leukemogenesis***

The first evidence for the involvement of menin in MLL1-associated leukemogenesis came from the initial biochemical studies carried out by Yokoyama et al., (218) who identified a menin binding motif (MBM) located within the first 330 amino acids of MLL1, a region that is also retained in all types of MLL1 aberrations (Figure 2.4) (218). Using leukemia cells that carry specific MLL1-translocations (MLL1-AF6, MLL1-ENL, MLL1-AF9, MLL1-AF10, and MLL1-GAS7) it was demonstrated that menin specifically associates with these MLL1-fusion proteins at the promoter of HOX genes such as: HOXA7, HOXA9 and HOXA10, which are constitutively expressed in several cases of acute leukemia (48, 217, 218). Acute loss of menin reduces the aberrant HOX gene expression and abrogates the differentiation arrest associated with these MLL1-fusions (218). Furthermore, conditional knockouts of the MEN1 gene suggests that menin is essential for the initiation and maintenance of MLL1-associated, but not other oncogene

induced, myeloid transformations (217, 218, 257). Together, these results demonstrate that MLL1-fusion proteins are dependent on menin for their oncogenic properties and raise the possibility that molecular therapies that target the menin-MLL1 interaction might be an effective strategy to treat leukemias.

While it is unclear how menin modulates the activities of wild-type and MLL1-fusion proteins, recent studies suggest that menin functions to recruit other proteins that are required for targeting MLL1 to downstream genes. For example, a recent study by Yokoyama and Cleary (237) suggests that menin promotes LEDGF binding to MLL1. LEDGF contains a highly conserved PWWP motif that is required for MLL1's association with downstream target genes (237). In addition, menin amino acid substitutions that disrupt its interaction with LEDGF without affecting menin's interaction with MLL1 also display decreased Hoxa9 gene expression (237, 258). These data indicate that part of menin's function is to stabilize the interaction between MLL1 and LEDGF. More recently, it has been shown that menin recruits the transcription factor cMyb to the MLL1 complex, which is required for recruitment of MLL1 to the Hoxa9 promoter (238). It has been shown that depletion of cMyb decreases the transforming potential of the MLL1-ENL fusion protein (238), suggesting that molecules that inhibit the menin-cMyb interaction may also be useful therapeutic agents.

#### ***2.5.7 Menin interacts with the N-terminus of MLL1 through an evolutionarily conserved Menin Binding Motif (MBM)***

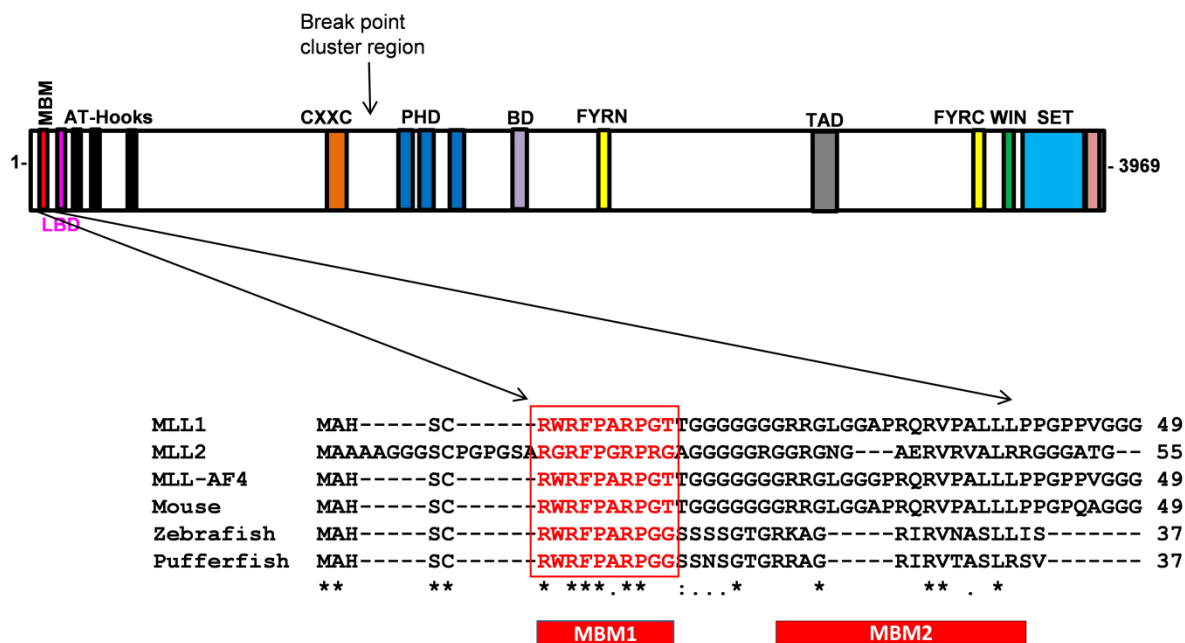
The interaction of menin with N-terminal sequences of wild-type MLL1 and MLL1-chimeric fusion proteins have been demonstrated by three independent reports (218, 257, 259). While the conclusions of these studies differ with respect to the exact length of the MLL1 fragment predicted to be involved in the interaction with menin, they have collectively identified a highly



conserved “menin binding motif” (MBM) localized within the first 46 residues of MLL1. For example, Yokoyama et al., (218) demonstrated that a consensus sequence (RXRFP), called the high-affinity menin binding motif (MBM), is present between MLL1 amino acids 6-10. Sequence analysis reveals that the MBM is highly conserved among MLL1 orthologs and in MLL2 (residues 17-21) (Figure 2.6). Deletions of the MBM in the MLL1-ENL and MLL1-GAS7 fusion proteins abrogates the interaction with menin in 293T cells (218). Furthermore, MLL1-ENL fusions that lack the MBM fails to induce acute myeloid leukemia in syngeneic recipient mice (218). This failure is associated with impaired expression of HOXA7 and HOXA9 genes in the MBM-deleted MLL1-ENL transduced murine myeloid progenitors (218). Moreover, MBM deletion mutants of MLL1-ENL lose their clonogenic potential and induce differentiation in leukemia blasts, a phenotype similar to the conditional knock out of menin or MLL1-ENL itself (101, 193, 218). These studies demonstrate the importance of menin-MLL1 interaction in the pathogenesis of MLL1-associated leukemias.

In an attempt to further characterize the physiological significance of MLL1-menin interaction in acute leukemias, Caslini et al., (257) employed a co-immunoprecipitation assay to demonstrated that MLL1 residues between 5 and 44 are required for high affinity binding with menin. This MBM region spans the RWRFP motif (residues 6-10) and also includes a second region between MLL1 residues 35 and 44 that is necessary, but not sufficient, for high-affinity interaction with menin (257). Internal deletions in the MLL1-AF9 fusion protein that lacked amino acids 5-15 (high affinity MBM) or 35-44 (low affinity MBM) failed to co-immunoprecipitate with menin from 293T cells (257). Furthermore, Caslini et al., (257) also demonstrated that MLL1-AF9 MBM sequences were essential for the transformation of hematopoietic progenitors by the MLL1-AF9 fusion protein. Interestingly, the MLL1 constructs

MLL1(2-167), MLL1(2-62), and MLL1(2-44) function as dominant negative inhibitors of the MLL1-menin interaction by titrating menin from the endogenous MLL1-AF9 protein resulting in reduced HOXA9 and MEIS1 expression and inhibition of the growth of transformed bone marrow progenitors (257). Together, these results suggest that expression of dominant negative MLL1 constructs or peptide inhibitors that mimic the MLL1-menin interaction can inhibit the transforming potential of MLL1-fusion proteins by specifically down regulating the expression of target HOX genes. Unexpectedly, dominant negative constructs of MLL1 also inhibited the colony-forming ability of wild-type hematopoietic progenitors since these constructs also mimic the interaction surface between wild-type MLL1 and menin (257). Collectively, these findings suggest that small molecule inhibitors that target menin-MLL1 interaction have a therapeutic potential to treat MLL1-associated leukemias, but with the caveat that normal hematopoiesis might also be impaired (257).



**Figure 2.6: Menin Binding Motifs (MBMs) are present in the wild-type, chromosomal translocations and partial tandem duplication mutations in MLL1**

Domain mapping experiments reveal that MBMs are present in the wild-type, chromosomal translocations, and PTD mutations in MLL1. The different functional domains in MLL1 are indicated and color coded as in Figure 1.2. The menin binding motif encompassing residues 5-44 (red) along with LEDGF binding domain (pink), AT-hooks (black), and the CXXC motif (orange) are retained by both MLL1-translocations and PTD mutations. The blow up shows the ClustalW-based multiple sequence alignment of the high affinity menin binding motifs, MBM1 and MBM2, present in human (Q03164), mouse (NP\_001074518), zebrafish (ACN88688), and pufferfish (AAC41377) MLL1, human MLL2 (O14686), and MLL1-AF4 (AAC37520) fusions (indicated in red). MBM1 and 2 were identified based on three independent studies (218, 257, 259).

### ***2.5.8 MBM based peptides as novel therapeutic agents for acute leukemias with MLL1-rearrangements***

Structural and biochemical studies that characterize the menin-MLL1 interaction in detail is an important step in the development of MBM-based small molecule inhibitors that can specifically help treat MLL1-mediated cancers. Grembecka et al., (259) carried out a detailed biophysical characterization of the interaction between menin and MLL1. Using a combination of NMR, Isothermal Titration Calorimetry (ITC) and Fluorescence Anisotropy (FP), it was demonstrated that MLL1 binds menin with high affinity ( $K_d \approx 10$  nM) utilizing two menin binding motifs (MBM1 and 2) located within the first 43 amino acids of MLL1 as previously suggested (218, 257). Furthermore, peptides derived from the MBM1 (amino acids 5-14) and MBM2 (23-40) in MLL1 bind menin with interaction affinities of 53 nM and 1400 nM, respectively (259). Using a series of peptide competition experiments, MBM1 and MBM2 peptides were shown to displace a construct of MLL1 (amino acids 2-43) from bound menin in vitro with  $IC_{50}$  values of 0.5  $\mu$ M and 37  $\mu$ M, respectively (259). Moreover, based on transfer-nuclear overhauser effects (Tr-NOEs) based NMR experiments, it was further suggested that MBM1 interacts with menin in an extended conformation and that the binding is facilitated by hydrophobic residues Phe9, Pro10 and Pro13 (259). Substitution to alanine of these amino acid residues significantly impaired the binding of MLL1 constructs to menin (259). Together, these studies have identified MBM1 (which encompasses the consensus RWRFP) as a potential drug target for leukemias with MLL1 translocations.

## ***2.6 Molecular targeting of MLL1-rearranged leukemias: Peptide inhibitors that target the activity of MLL1-AF4 and MLL1-AF9 fusion proteins***

### ***2.6.1 Clinical significance of t(4;11) and t(9;11) translocations***

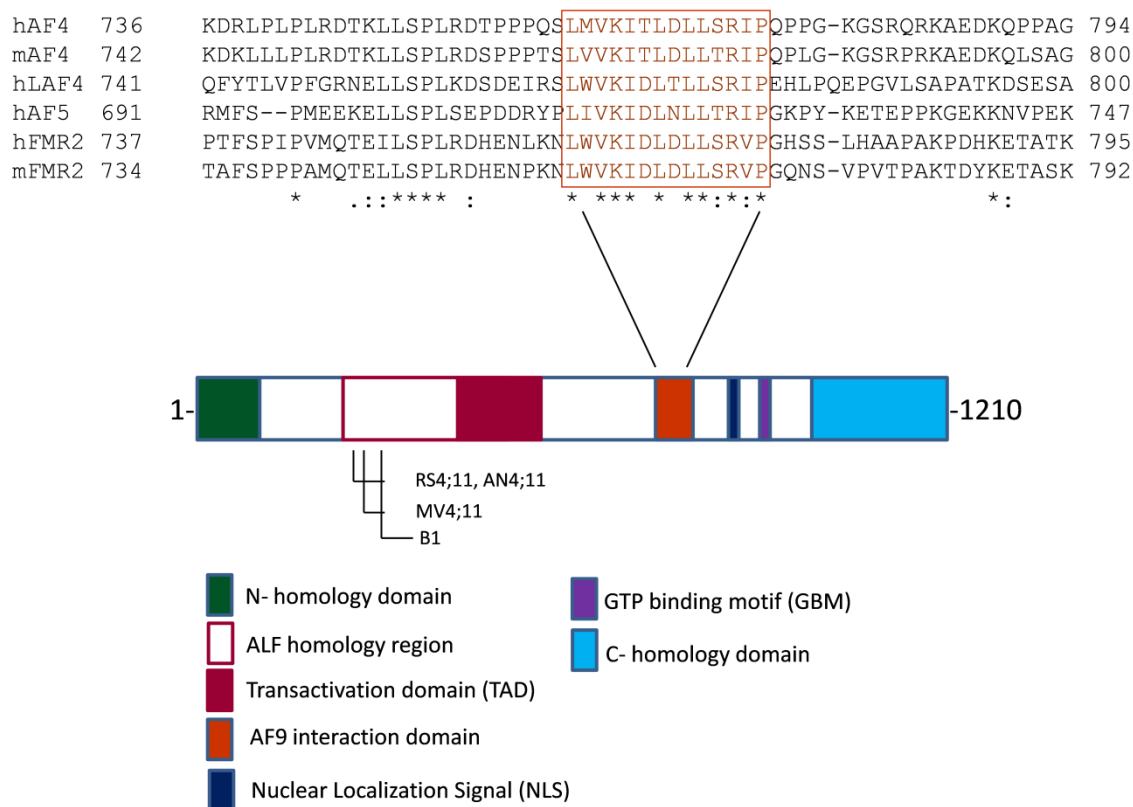
MLL1 translocations do not share a common structural motif or biochemical function. However, based on sequence similarities, the most commonly occurring MLL1 fusions can be grouped into three major gene families: AF10/AF17, ENL/AF9, and the largest family, AF4/LAF4/AF5q31/FMR2 (260). The most common translocation is t(4;11)(q21;q23) and is associated with more than 50% of acute leukemia cases in infants, and for 3-6% of cases in older children (261-263). The t(4;11) translocation results in leukemic blasts expressing phenotypic markers for ALL in 95% of the cases (264, 265). t4:11 translocations result in cancers that often spread beyond the hematopoietic lineage and have a poor prognosis (100, 265, 266). The t(4;11) translocation retains the 5' portion of MLL1 gene containing the menin binding motif (MBM), AT hooks and the CXXC DNA binding motifs, which are fused in-frame to the 3' portion of the gene at the 4q21 locus called AF4 (165). The high occurrence rates of t(4;11) translocation in infants along with the poor prognosis and absence of chemotherapeutics to treat these leukemias highlight an urgent need for the development of inhibitors that specifically target the gain-of-function phenotypes associated with the MLL1-AF4 fusion. In this regard, inhibitors that target the interaction between AF4 and its partner protein AF9 have been developed recently and show promising results in inhibiting the transforming potential of leukemia cell lines bearing MLL1-AF4 or MLL1-AF9 translocations (267-269). The biochemical studies that form the basis for these conclusions are summarized in this section.

### **2.6.2 Domain architecture and the function roles of AF4 family**

AF4, also known as AFF1/FEL, is a serine/proline-rich nuclear protein with crucial roles in B and T lymphocyte development. AF4 has several putative functional domains including the ALF (AF4/LAF4/FMR2 homology) domain, which mediates the interaction with a family of ubiquitin ligases called SIAH (seven in absentia homolog), a serine/proline-rich transcriptional activation domain (TAD), nuclear localization signals (NLS), a guanosine triphosphate (GTP) binding motif (GBM), and a C-terminal homology domain involved in intra-nuclear localization and binding to pre-mRNA splicing factors (265, 270-274) (Figure 2.7). AF4 is located at a fragile break-point region on chromosome 4 and is associated with a wide variety of chromosomal translocations. AF4 is a member of AF4/LAF4/AF5q31/FMR2 family of nuclear transcription factors (275-277) and also shows significant homology to the *Drosophila melanogaster* pair-rule gene *Lilliputian* (278). Surprisingly, three of these family members (AF4/LAF4/AF5q31) are associated with infant leukemias involving reciprocal translocations with the MLL1 gene (279-281). The second family member LAF4, isolated from Burkitt's lymphoma, is a lymphoid-specific transcription factor and has transcriptional activation domains and nuclear localization signals that are highly similar to AF4 (281). Indeed, the MLL1-LAF4 fusion proteins also retain the TAD in LAF4, which can functionally substitute for the activation domain in MLL1, thereby contributing to the leukemogenic potential of these chimeric fusion proteins (281). The third AF4 homologue AF5q31 was originally identified from infant acute leukemias with a (5;11)(q31;q13q23) translocation (279). Strikingly, each of these MLL1-AF4 gene family fusions manifest very similar clinical characteristics- that is, early onset, poor prognosis, and a mixed immunophenotype. However, the fourth AF4 family member, FMR2, is associated with mental retardation and is located in the folate-sensitive break-point region at

chromosome X band q28 (275, 282). FMR2 has not been reported to be a part of any MLL1 fusions.

Despite the growing amount of clinical evidence suggesting the involvement of MLL1-AF4 family translocations in human acute leukemias, the molecular mechanisms by which the MLL1-AF4 fusion proteins induce neoplasia are not fully understood. One possibility is that the resulting MLL1-AF4 fusion protein has a gain-of function phenotype wherein the transcriptional activation domain of AF4 is fused to the DNA binding domains of MLL1 and alters the regulation of MLL1 dependent genes (82). However, genetic studies using mouse knock-in models suggest that the two fusion proteins MLL1-AF4 and AF4-MLL1 resulting from the balanced reciprocal translocation between MLL1 and AF4 could independently contribute to the pathogenesis of t(4;11) mediated acute leukemias (181, 271, 283). Importantly, the AF4-MLL1 fusion protein retains the C-terminal portion of MLL1 including its SET domain, and is sufficient to induce leukemia in the absence of MLL1-AF4 (283).



**Figure 2.7: Domain representation of human AF4 indicating the putative functional domains including the AF9 interaction region**

Overview of the protein domain structure of human AF4 adapted from (267, 270) and modified accordingly. The different domains are highlighted: N-homology region with an unknown function (green), ALF (AF4/LAF4/FMR2 homology) region which includes the transcription activation domain (dark red), AF9 interaction region (orange), nuclear localization signals (dark blue), a putative consensus guanosine triphosphate (GTP) binding motif (purple), and C-terminal homology domain (blue). The domains are not drawn to scale. The ClustalW-based multiple protein sequence alignment of the AF9 binding regions of AF4 family members: human AF4 or hAF4 (P51825), mouse AF4 or mAF4 (AAU93698), human LAF4 or hLAF4 (NP002276), human AF5 or hAF5 (Q9UHB7), human FMR2 or hFMR2 (BAC81113) and mouse FMR2 or mFMR2 (CAA04821) are shown with the AF9 interaction region highlighted in brown. Chromosomal break points are also indicated for cell lines carrying t(4;11) translocations.



### ***2.6.3 Molecular interaction network for AF4 fusion proteins***

These studies raise the possibility that MLL1-AF4 and AF4-MLL1 participate in different sets of interaction networks as compared to the native full-length proteins (MLL1 and AF4) and result in gene expression signatures that are a representative of the physiological function of the fusion partners. This hypothesis is supported by the studies of Benedikt et al., (284) who used affinity purified AF4 and AF4-MLL1 complexes from 293T cells to elucidate the subunit composition of the two complexes. Wild-type AF4 was purified in a complex containing the CDK9/Cyclin T heterodimer, which resembles the positive transcription elongation factor b (p-TEFb) (284, 285) and also co-purifies with the wild type versions of two other known MLL1 fusion partners, AF9 and ENL, both of which bind to the C-terminus of AF4 family members (88, 284, 286). Furthermore, ENL binding to AF4 creates a binding site for AF10, DOT1 methyltransferase and histone H3 (185, 284). Both the DOT1 mediated H3K79 methylation and the p-TEFb mediated phosphorylation of RNA Pol II C-terminal domain could lead to transcriptional elongation and is facilitated by the interaction of wild-type AF4 with these proteins (184, 284). In contrast, the subunit composition of the AF4-MLL1 complex is different from the wild-type AF4 protein alone due to additional proteins that interact with the fused MLL1 C-terminal fragment and modulate the transcriptional activating properties of AF4-MLL1 fusion protein (284). MLL1-AF4 chimeric fusions on the other hand have an intact C-terminus of AF4 and therefore retain the transcriptional activation domains and its ability to interact with AF9 and ENL.

### ***2.6.4 Functional significance of AF9/ENL family***

Reciprocal translocations involving the AF9 (t(9;11)(p22;q23)) and ENL (t(11;19)(q23;p13.)) genes are also associated with several cases of ALL and AML (287). Similar to AF4, AF9 also

belongs to a family of serine/proline-rich transcription factors (82, 288, 289) and shares significant similarity to ENL and the yeast protein ANC1 (290, 291). Interestingly, ANC1 has been demonstrated to be a part of the yeast RNA polymerase II complex, as well as the SWI/SNF nucleosome-remodeling complex, a macromolecular complex which functions as ATP-dependent chromatin remodeler (292-294) The high degree of sequence similarity between human AF9/ENL and the yeast ANC1 protein has led to the hypothesis that AF9 and ENL may also interact with a human SWI/SNF remodeling complex similar to yeast ANC1, and the MLL1-AF9/ENL fusion proteins may retain these features. The biological functions of AF4, AF9 and ENL are not clearly understood, however, gene deletion studies in mice have demonstrated important roles for these proteins during development (271, 295, 296) Furthermore, the endogenous ENL protein was purified as a part of a macromolecular complex (ENL associated protein complex or EAP) that also contains p-TEFb, DOT1 and AF4 and plays a putative role in transcriptional elongation (184, 185). Despite the fact that MLL1-AF4 translocations and MLL1-AF9/ENL fusions account for more than 50% of MLL1 11q23 associated leukemias (168-170), there exists no functional similarity between these MLL1 fusion partners. It is tempting to hypothesize that MLL1 fusion proteins might hijack the p-TEFb/DOT1 mediated transcriptional elongation activity or the SWI/SNF dependent nucleosome remodeling activity through the fusion partner and result in constitutive target gene expression leading to leukemia. These studies suggest that pTEFb, DOT1 and the MLL1 fusion partners AF4/AF9/ENL are all molecular targets in the development of therapeutics that target MLL1-fusion mediated leukemias.

### ***2.6.5 Domain mapping of the interaction region between AF4 and AF9***

Co-localization studies carried out by (286) have established that the two most common MLL1 fusion partners AF4 and AF9 form a stable complex within the nucleus and are restricted to discrete nuclear foci called “AF4 bodies”. AF4 bodies are nuclear speckle-like in appearance and are distinct from the nucleolus, cajal bodies, PML body or regions associated with DNA replication and repair (286). Using yeast two-hybrid screens it was demonstrated that the minimum motif in human AF4 required for binding AF9 encompasses 14 residues (761-774) that are proximal to the bipartite nuclear localization signal (NLS) in AF4 (refer to Figure 2.7 for domain representation). Deletion of these residues completely abolishes the interaction between AF4 and AF9 and also results in diffuse AF4 bodies (267, 286). Site-directed mutagenesis studies reveals that bulky hydrophobic residues within the conserved AF9 binding region in mouse AF4 ( mAF4 ) dictate its binding affinity to AF9 (267). Using fluorescent tagged AF4 and AF9 constructs it was further demonstrated that both the 14 residue AF9 binding motif and the bipartite NLS was required for the punctate nuclear speckle distribution of AF4-AF9 complexes (286). Likewise, the terminal 93 residues in AF9 (and the terminal 84 in the case of ENL) were identified as the minimal region required for binding AF4 (286). Interestingly, the mutual interaction domains between AF4 and AF9 are highly conserved in the AF4 and AF9 homologues (refer to figure 5 for sequence alignment) (267, 286) and the AF4-AF9 interaction region is retained by the MLL1 fusion proteins (MLL1-AF4 and MLL1-AF9) (180, 280, 286) suggesting that AF4-AF9 interaction might be an important step in the pathogenesis of these MLL1 fusions. The functional significance of the interaction between AF4 and AF9 is unclear; however, the co-localization of these two proteins to specific sub-nuclear foci suggests that AF4-AF9 interaction might be required for normal cellular functions as well as in the pathogenesis of

MLL1-AF4 or MLL1-AF9 associated leukemias (286). Furthermore, the presence of this interaction region in MLL1-AF4 and MLL1-AF9 fusion proteins, and the ability of MLL1-AF4 fusions to alter the localization of endogenous AF9 also suggest that AF4-AF9 protein complex is a pharmacological target for leukemia therapy (267, 286).

#### ***2.6.6 A synthetic peptide PFWT disrupts the interaction between AF4 and AF9***

Based on the initial mapping studies of the AF9 binding region in AF4, a synthetic peptide, designated “PFWT” that mimics the interaction region was developed and tested for its ability to disrupt the AF4-AF9 interaction both in vitro and in vivo (267, 286). The initial PFWT peptide developed in this study was based on the highly conserved AF4-AF9 interaction sequence in the mouse FMR2 protein and encompasses residues (759-771) (see Figure 2.7 for the sequence alignment of the AF9 interaction region in AF4 family members). To enable nuclear uptake, the PFWT peptide was conjugated to a penetratin transporter sequence at its N-terminus (267). Using pull-down assays with GST-tagged AF4 and biotinylated AF9, it was shown that this PFWT peptide disrupts the interaction between human AF4 and AF9 in a concentration dependent manner (267). However, a control peptide (containing amino acid substitutions at V763E and I765S) did not interfere with the binding of AF4-AF9 complex (267). The specificity of PFWT peptide against AF4-AF9 complexes was further demonstrated by its inability to disrupt the interaction between AF9 and two other proteins, the Polycomb protein (MPC3) and the mouse homolog of BCL-6 co-repressor (mBCoR), both of which interact through the C-terminus of AF9 (267, 288, 297). Furthermore, the PFWT peptide was shown to be readily taken up by NIH3T3 cells and disrupts the co-localization of AF4-AF9 in vivo (267). These results corroborated the in vitro findings that the PFWT peptide has the ability to specifically disrupt the interaction between AF4 and AF9 protein complexes.

### ***2.6.7 PFWT peptide inhibits the cell proliferation of leukemia cell lines with t(4;11) and t(9;11) translocations***

Based on these observations it was further predicted that the PFWT peptide will have the ability to inhibit the proliferation of leukemia cell lines that carry a t(4;11)(q21;q23) or t(9;11)(p22;q23) translocations (267-269). As expected, PFWT peptide specifically inhibits the proliferation of leukemia cell lines B1, MV4-11 and RS4;11 (298-300) that harbor the MLL1-AF4 translocation (267-269). Interestingly, the PFWT peptide also inhibits the survival capacity of KP-L-RY cell lines (300) that are characterized by a t(5;11) translocation associated with MLL1-AF5q31 fusions (267). In contrast to cell lines that carry either the MLL1-AF4 or MLL1-AF5q31 translocations, the PFWT peptide shows mixed effects in the inhibition of the proliferative capacity of leukemia cell lines that carry MLL1-AF9 translocations (267, 268). For instance, based on the study by (267), the PFWT peptide failed to inhibit the growth of THP-1 cells (301) that are associated with MLL1-AF9 translocations. However, a more recent study by (268) demonstrated that Molm13 leukemia cells (302) that are also associated with MLL1-AF9 translocations are sensitive to treatments with PFWT peptide in a dose dependent manner. While these findings suggest that there are differences in the inhibitory properties of the PFWT peptide towards cell lines that carry a similar translocation, it highlights the fact that these leukemias might operate through additional pathways that lead to misregulated gene expression profiles (267, 268). Furthermore, the specificity of PFWT peptide in inhibiting the leukemia cells with MLL1-AF4 and MLL1-AF9 mutations is corroborated by the findings that the MOLT-4 (T-ALL) cell line (303) that does not contain a MLL1-AF4 translocation is not affected by the PFWT peptide even at higher doses (267). However, the Reh (B-precursor) cell lines (304) that do not carry a MLL1-AF4 chimeric fusion are susceptible to the PFWT peptide (267). These results suggest that some cell lines might require the function of normal AF4-AF9 complexes for

its carcinogenesis and these native complexes are also targeted by the PFWT peptides (267). Nevertheless, these findings demonstrate the wide range of specificity for the PFWT peptide and suggest that PFWT-like peptides are promising lead compounds in the development of treatment regimens for the most frequent 11q23 translocations. The pharmacological importance of PFWT based peptides is further enhanced by the fact that these peptides have little effect on the number and colony forming potential of hematopoietic progenitor cells (267), suggesting that PFWT-like peptides may have fewer side effects .

### ***2.6.8 PFWT peptide as a lead compound in the treatment of MLL1-rearranged leukemias***

While initial studies suggest that the growth inhibitory properties of the PFWT peptide is mediated through apoptosis (267), recent work by (268) demonstrates that PFWT peptide induces cell death by necrosis in MV4-11 and Molm13 cell lines. Necrotic cell death was further demonstrated by the changes in plasma membrane integrity in the absence of traditional apoptotic markers: nuclear disintegration, caspase activation, DNA fragmentation or mitochondrial membrane depolarization (268). Given the resistance of t(4;11) translocations to conventional pro-apoptotic chemotherapeutic drugs, PFWT mediated necrotic cell death is an attractive alternative strategy to treat these acute leukemias. More recently, Bennett et al., (269) demonstrated that the PFWT peptide in combination with standard chemotherapeutic agents such as etoposide, 17AAG (17-(allylamino)-17-demthoxygeldamycin), cytarabine, and Flt-3 kinase inhibitor results in synergistic cytotoxicity in MV4-11 cells that harbor t(4;11) translocations. Furthermore, cell death was mediated through both apoptotic and necrotic pathways suggesting that PFWT peptides could be useful in combinatorial therapy to treat leukemia cell lines that are relatively resistant to current apoptotic drugs (269, 305-307). Together, these findings emphasize that PFWT peptides inhibit the proliferation of t(4;11), t(5;11) and t(9;11) leukemia cell lines by

specifically disrupting the association of AF4 or AF5q31 with AF9 and serves as a starting point for the development of more effective therapeutic strategies. In the absence of a crystal structure that clearly demonstrates the interaction mode between AF4 and AF9 proteins, systematic site-directed mutagenesis experiments can help identify amino acid positions that are more important for the overall binding affinity. Such an approach has the potential to help identify additional PFWT-based peptido-mimetic compounds that have better inhibitory properties than the initial PFWT peptide and prevent the potential toxic side effects that could occur at high doses of the PFWT peptide.

## ***2.7 Win motif-based inhibitors that target the assembly and enzymatic activity of the MLL1 core complex***

### ***2.7.1 Novel strategies that target the histone methyltransferase activity of MLL1 amplifications and MLL1-PTDs***

Growing evidence suggests that increased HOXA gene expression associated with MLL1 amplifications and MLL1-PTDs underlie the pathogenesis of these leukemias (79, 308). HOXA gene expression is dependent on the histone methyltransferase activity of MLL1 ((54), and given the increased H3K4 methylation observed in MLL1-PTDs (78, 79), inhibitors that down-regulate the histone methyltransferase activity of MLL1-PTD have the potential to reverse this aberrant epigenetic program. Until recently, inhibitors that specifically target MLL1's H3K4 methyltransferase activity have been difficult to develop due to lack of a clear understanding of MLL1's structure and function. However, recent findings have established that H3K4 methylation levels are precisely regulated by the function of two independent methyltransferases: the MLL1 SET domain and a novel multi-subunit enzyme, WDR5-RbBP5-Ash2L-DPY30 (WRAD), that lacks sequence homology to known methyltransferases (32, 33). Because multiple

methylation on H3K4 is catalyzed by two different enzymes, it is interesting to speculate that inhibitors that prevent the association of two methyltransferases will have clinical significance in the treatment of acute leukemias that are characterized by aberrant H3K4 methylation (78, 79).

### ***2.7.2 Peptides derived from the MLL1 Win motif inhibit the H3K4 dimethylation activity of MLL1 core complex***

Based on mapping studies of the WDR5 binding region in MLL1, Patel et al (60) developed a peptide, designated the MLL1 *Win* motif peptide (3762-3773) that can bind WDR5 with high affinity ( $K_d=1700$  nM) and disrupt the assembly of the MLL1 core complex in vitro (60, 61). Consequently, this 12-residue MLL1 *Win* motif peptide was further shown to inhibit the in vitro H3K4 dimethylation activity of the MLL1 core complex in a dose dependent manner by competing with wild-type MLL1 for the arginine binding pocket in WDR5 (61). However, a control p53 peptide that has an arginine in a different sequence context failed to inhibit the H3K4 dimethylation activity even with a 60-fold excess, suggesting that *Win* motif peptides are highly potent inhibitors of MLL1 family complexes (61). Furthermore, data presented in this dissertation in Chapter 3 demonstrate that peptides derived from other human SET1 family members bind WDR5 with dissociation constants ranging from 50 nM-2800 nM (Figure 3.3) (309). In accordance with the binding studies, MALDI-TOF mass spectrometry based methylation assays further demonstrate that other human SET1 family *Win* motif peptides are 14 to 72-fold more potent inhibitors of the H3K4 dimethylation activity of the MLL1 core complex as compared to the MLL1 *Win* motif peptide (Figure 3.6) (309). These studies suggest that the human SET1 family *Win* motif based peptides represent an excellent starting point for the design of lead compounds that would specifically disrupt the interaction between MLL1 and WDR5 and inhibit the H3K4 dimethylation activity of MLL1 core complex in vivo.



### ***2.7.3 Analysis of the binding of MLL1 Win motif and histone H3 peptides to WDR5: identification of key structural elements required for binding WDR5***

In order to characterize the protein structural features within the MLL1 Win motif peptide that is required for its high affinity binding to WDR5, Karatas et al (310) performed a systematic deletion analysis on the original 12-residue MLL1 *Win* motif peptide (3762-3773) and showed that a three residue sequence composed of  $\text{Acetyl-ARA-NH}_2$  is the minimal motif that is required for its interaction with WDR5 (310). The  $\text{Acetyl-ARA-NH}_2$  peptide binds WDR5 in peptide competition experiments with similar affinity to that of the original 12 residue MLL1 *Win* motif peptide (~120 nM) (61, 310). In addition to the  $\text{Acetyl-ARA-NH}_2$  peptide, Karatas et al (310) also identified another peptide designated the Acetyl-10mer ( $\text{acetyl-ARAEVHLRKS-NH}_2$ ) peptide encompassing residues 3764-3773, which binds WDR5 with 10-fold better affinity ( $K_i = 3$  nM (310)) compared to that of the  $\text{Acetyl-ARA-NH}_2$  or MLL1 *Win* motif peptide. The results of these studies are described in more detail in Chapter 5 along with the binding and structural studies for the  $\text{Acetyl-ARA-NH}_2$  and the Acetyl-10mer peptides that were carried out as a part of this dissertation.

### ***2.7.4 Win motif based inhibitors: a novel class of compounds with potential to treat Acute Myeloid Leukemia***

The extensive biochemical characterization of the peptide derived from the MLL1 *Win* motif have demonstrated that these peptidomimetic compounds have the potential to inhibit the H3K4 methyltransferase activity of the MLL1 core complex by disrupting the association of MLL1 with the WRAD sub-complex (61). Furthermore, the crystal structures of WDR5 bound to human SET1 family *Win* motif peptides, described in Chapter 4 of this dissertation, provides a

rationale for the design of new peptides and nonpeptide mimetics with better inhibitory properties. The structures also reveal that conserved residues from the human SET1 Win motifs form a cyclical shaped  $3_{10}$ -helix that fits snugly into the outer opening in WDR5 and raises the possibility that other cyclic peptidomimetic compounds might also bind WDR5 with similar or better inhibitory properties. The efficacy of *Win* motif based inhibitors in down regulating the H3K4 methylation activity of MLL1 core complex *in vitro* has yet to be demonstrated *in vivo*. However, based on the *in vitro* studies it is expected that the *Win* motif based peptides will have the ability to down regulate the increased H3K4 dimethylation, and aberrant HOX gene expression associated with MLL1-PTDs and gene amplification mutations in MLL1. Hence, the *Win* motif peptide is a novel “first in class” inhibitor that are expected to have the ability to downregulate H3K4 dimethylation levels in the cell without perturbing K3K4 monomethylation, or the methylation activities of other H3K4 methyltransferases.

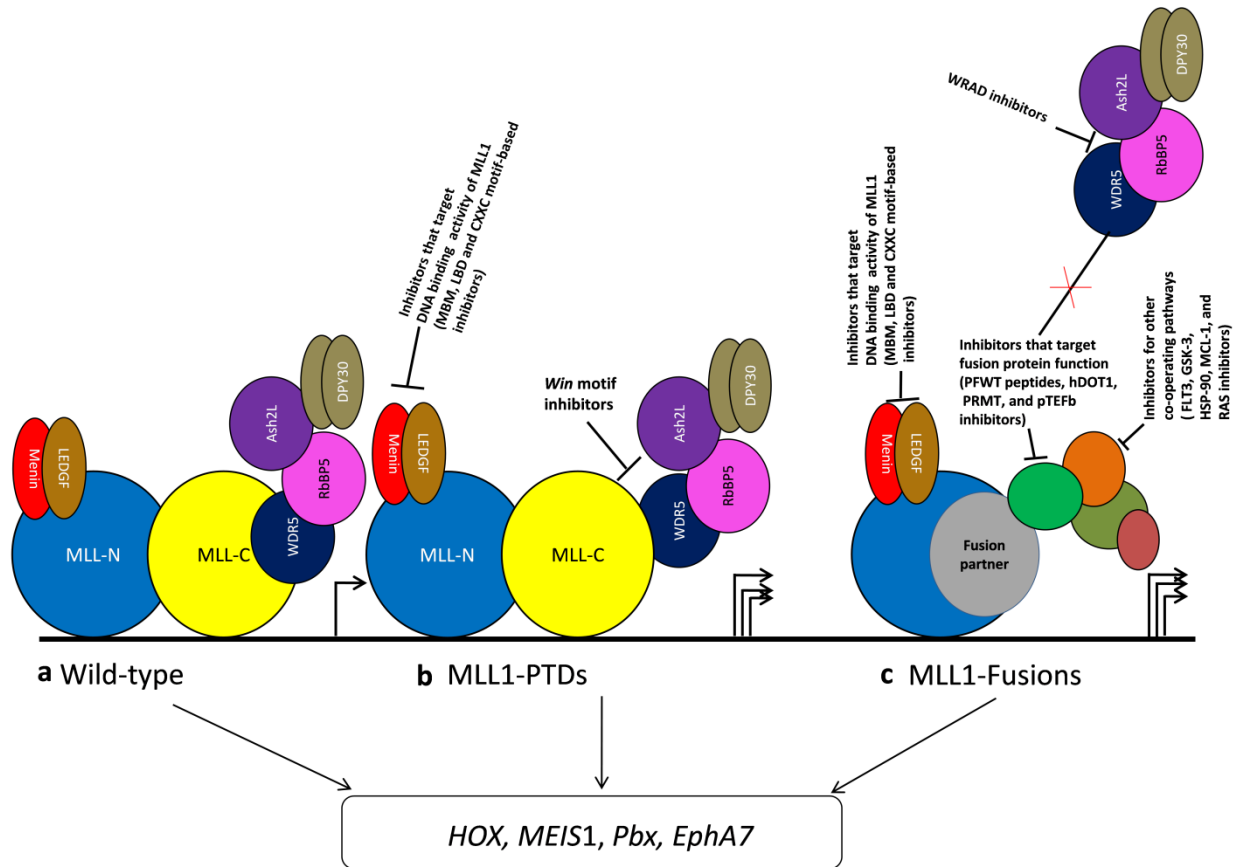
### ***2.8 Conclusions and future perspectives-are there other molecular targets for leukemia therapy?***

Over the past decade, we have witnessed remarkable strides towards understanding the fundamental mechanisms of MLL1-mediated transcription and leukemogenesis, and many targets that are likely to be biomedically important are beginning to be characterized (7, 23, 311). Genetic, biochemical and structural studies have demonstrated that several MLL1 partner proteins (AF4, AF9, AF10, ENL, and EEN) are either directly or indirectly associated with macromolecular complexes involved in transcriptional control and elongation (88, 182, 184, 185, 187, 311, 312). Based on these studies, it was also suggested that leukemogenic properties of these MLL1-fusions are mediated in part through their association with the histone methyltransferases: hDOT1 (182, 184, 312) and protein arginine methyltransferase-1 or PRMT1

(187). These studies have led to the proposal of the “MLL1 web hypothesis”, which states that MLL1 fusion partners are components of larger macromolecular complexes that are involved in transcriptional activation and/or elongation (286). Molecular therapies directed at the interactions of these proteins within the “MLL1 web hypothesis” or inhibiting the activity of associated histone modifying enzymes are emerging as promising targets (7, 23, 311). In addition to targeting the function of MLL1-fusion proteins, inhibitors that target the association of MLL1-fusion proteins to target DNA might also have potential therapeutic implications. In this regard, inhibitors that disrupt the interaction between MLL1’s CXXC motif and unmethylated CpG containing target DNA are being investigated as molecular drug targets in MLL1-related leukemias (313, 314). In addition to these molecular targets, recent studies have also identified other cooperating pathways such as positive transcription elongation factor-b (pTEFb) recruitment (88, 185, 311, 315), activation of FMS-like tyrosine kinase 3 (FLT3) (199, 200, 202, 205), glycogen synthase kinase 3 (GSK3) (206), heat shock protein-90 (HSP-90) (198, 202), myeloid cell leukemia sequence-1 (MCL-1) expression (316), and RAS pathways (204) that are implicated in MLL1-induced leukemogenesis. The therapeutic significance of these molecular targets in MLL1-therapy were reviewed recently (7, 23, 311) and since then, additional potential pharmacological targets such as the multi-subunit complex WRAD, which regulates the degree of H3K4 methylation activity by MLL1 core complex, have emerged (32, 33) (Figure 2.8).

In conclusion, this chapter summarizes the recent biochemical studies that have contributed significantly to our understanding of how MLL1 works and have led to the identification of promising therapeutic targets for MLL1-related leukemias. In particular, inhibitors that target molecular interactions between menin and MLL1, AF4 and AF9 complex,

and MLL1 and WRAD sub-complex have emerged as novel candidate pharmacological targets. These MLL1-targeted therapies have enhanced pharmacological potential as compared to the existing broadly-based chemotherapeutics for MLL1-associated leukemias and will hopefully result in better treatment outcomes. In the future, similar biochemical, structural and genetic studies will be instrumental in identifying additional molecular targets that will form the basis for novel treatment strategies.



**Figure 2.8: Potential protein-protein interaction targets for the development of novel therapies for MLL1-associated leukemias**

Cartoon representation of the different forms of MLL1 (wild-type, PTDs, and chromosomal translocations) and a summary of emerging potential molecular targeted therapies for mixed lineage leukemia. Key molecular targets include: i) inhibitors that disrupt the association of MLL1 to its target genes (MBM-based, LBD-based or CXXC motif-based inhibitors), ii) inhibitors that target the assembly of MLL1 core complex (*Win* motif-based inhibitors), iii) inhibitors that target the functions of fusion proteins (PFWT-like peptides, hDOT1L inhibitors, pTEFb kinase inhibitors), iv) WRAD sub-complex inhibitors and v) Inhibitors that target other co-operating pathways in leukemia (GSK-3, FLT3, MCL-1, and Ras). MLL1-N (blue) and MLL1-C (yellow) fragments are shown as localized to a MLL1 target gene. In a) menin (red), LEDGF (brown) and WRAD sub-complex associate with wild-type MLL1 to regulate the normal expression (indicated by a single arrow) of *HOX*, *MEIS1*, *Pbx*, and *EphA7*. Target gene expression is dysregulated in MLL1-PTDs (b) and MLL1-fusions (c) as indicated by multiple arrows. b) MLL1-PTDs retain all the functional interactions as seen in wild-type MLL1 and c) MLL1-fusions do not retain the MLL1-C fragment and many interacting proteins. Refer to text for a detailed description of novel molecular targets that are emerging as potential candidates in the design of treatment strategies for mixed lineage leukemia. Figure 2.8 was adapted from Liedtke and Cleary (23) and modified based on the studies by Patel et al., (32, 33, 60, 61).

### **CHAPTER 3: Thermodynamic characterization of the interaction between WDR5 and Human SET1 family *Win* motifs.**

This chapter is a reprint of reference ((309)) and modified accordingly to include additional data for this dissertation. Binding and inhibition studies for the 12-residue MLL1 *Win* motif peptide was described previously by Patel et al (32, 60) and have been included in this dissertation for completeness. Co-Immunoprecipitation studies described in this chapter were carried out by Dr. Jeong-Hong Lee in Dr. David Skalnik's group at Department of Pediatrics, Indiana University.

## INTRODUCTION

We have previously shown that the conserved R3765 of the MLL1 *Win* motif is required for the assembly and H3K4 dimethylation activity of the MLL1 core complex. We have also shown that a 12 residue peptide derived from the *Win* motif of MLL1 (amino acid residues 3762-3773) can bind WDR5 with high affinity ( $K_d=1700$  nM) and disrupt the assembly and H3K4 dimethylation activity of the MLL1 core complex in vitro (60, 61). These initial studies suggest WDR5-MLL1 *Win* motif interaction is conserved in vivo and raises the possibility that peptides that mimic this interaction surface will be useful inhibitors of MLL1 core complex dependent H3K4 dimethylation activity. These initial studies are extended further in Chapter 3 where we (in collaboration with David Skalnik's lab) demonstrate that the conserved arginine 3765 of the MLL1 *Win* motif is required for co-immunoprecipitation of WDR5, RbBP5 and Ash2L from mammalian cells, confirming its critical role in the assembly the MLL1 core complex (309). Furthermore, the *Win* motif sequence is highly conserved among metazoan SET1 family members (refer to Figure 1.2 and appendix Figures 1a-1b for alignment); however, little is known about how WDR5 recognizes the different SET1 family members. To begin to understand the molecular basis for the interaction between WDR5 and human SET1 family *Win* motifs, this chapter describes a detailed thermodynamic and structural analysis of the binding of several different *Win* motif peptides to WDR5. While the six residue *Win* motif is highly conserved among the human SET1 family members, we were also interested in understanding the role of the non-conserved residues flanking the six-residue *Win* motif in recognition by WDR5. To delineate the potential roles of these additional amino acids in binding WDR5, in Chapters 3 and 4, we have performed a systemic structural and functional analysis of the interaction between WDR5 and six different human SET1 family *Win* motif peptides containing the six-residue *Win*

motif sequence flanked on both N- and C-termini by four additional naturally occurring amino acid residues.

In Chapter 3, we compare the thermodynamic binding parameters and inhibition constants of six-different human SET1 family *Win* motif based peptides bound to WDR5. We show that *Win* motif peptides derived from human SET1 family members (MLL2, MLL3, MLL4, SETd1a and SETd1b) interact with WDR5 with dissociation constants ( $K_d$ ) ranging from 54-541 nM. These binding affinities are **5 to 51-fold stronger** than the interaction affinity between MLL1 *Win* motif and WDR5 (2762 nM) (309). We further demonstrate that the other human SET1 family *Win* motif peptides are also **14 to 72-fold** more potent inhibitors of the H3K4 dimethylation activity of the MLL1 core complex as compared to that of MLL1 *Win* motif peptide (309). In Chapter 4, we provide evidence based on WDR5-*Win* peptide co-crystal structures that these binding energy differences are governed by the differences in the amino acid sequences between the SET1 family *Win* motif peptides both within and outside the conserved *Win* motif sequence.

Collectively, these studies demonstrate that WDR5-*Win* motif interaction is conserved among human SET1 family members and also suggest a framework for the design of better *Win* motif based peptide inhibitors for the MLL1 core complex.

## METHODS

***Co-immunoprecipitation and Immunoblotting:*** Human embryonic kidney cells (HEK293) were transiently transfected with pCMV-Myc tagged MLL-C180 constructs expressing either the



wild-type or mutant (R3765A) as previously described (41). After 48 hours of transfection, nuclear extracts were prepared as previously described (41) and incubated with anti-Myc agarose beads (Sigma) for 3 hours. Bound proteins were eluted with SDS sample buffer after extensive washing, and were analyzed by western blotting. Antisera utilized are as follows. Anti-Myc antibody was obtained from Santa Cruz Biotechnology, Inc. Antisera directed against Ash2L and Rbbp5 were obtained from Bethyl Laboratories. Anti-Wdr5 antiserum was previously described (42).

***Protein expression and purification:*** 6X-His tagged versions of full-length WDR5 (1-334) and an N-terminal truncated form of WDR5 (residues 23-334,  $\Delta$ N-WDR5) were expressed and purified as described previously (60, 61). For the first step of purification, the crude lysate was passed through a HisTrap column (GE healthcare) containing Nickel beads. The bound 6x-His-WDR5 was eluted from the column using an elution buffer that contains 500 mM Imidazole. The peak fractions that contain WDR5 was then combined and dialyzed with three changes against the column buffer with no imidazole (at 4 °C) to remove the imidazole and to cleave the 6X-His tag in the presence of TEV (Tobacco Etch Virus) protease. For the second step of purification the dialyzed protein was passed through the HisTrap column and the flow through fractions that contain the untagged version of WDR5 are collected. As a final step of purification, the protein was passed through a gel filtration column (Superdex 200<sup>TM</sup> GE Healthcare) pre-equilibrated with the sample buffer containing 20 mM Tris (pH 7.5), 300 mM sodium chloride, 1 mM Tris (2-carboxyethyl) phosphine, and 1  $\mu$ M zinc chloride. All other proteins used in the inhibition assays were purified as previously described (60, 61).

**Peptide synthesis:** All six human SET1 family *Win* motif peptides used in this dissertation were synthesized by Genscript (refer to Tables 3.1-3.2 for peptide sequences). All peptides were synthesized with an acetyl- and amide-capping group at the N- and C-terminus, respectively, to eliminate the contributions of unnatural N- and C-terminal charges on binding. The isoelectric points (pI) for the peptides are listed in Table 3.3. The pI values were calculated using the EXPASY protparam tool. The net charge on each *Win* motif peptide at pH 7.5 is reported.

**Isothermal Titration Calorimetry:** Isothermal Titration Calorimetry (ITC) experiments were carried out using VP-ITC calorimeter (MicroCal). All ITC experiments were performed at 20°C in a sample buffer containing 20 mM Tris (pH 7.5), 300 mM sodium chloride, 1 mM Tris (2-carboxyethyl) phosphine and 1 μM zinc chloride. Prior to ITC, all peptides and proteins were dialyzed against sample buffer to minimize variations in sample preparations. Individual ITC experiments for each of the six human SET1 family peptides were carried out by titrating a known concentration of *Win* motif peptide (determined by amino acid analysis at the KECK proteomics facility at Yale University) diluted in the sample buffer into a sample cell containing a known concentration of full-length WDR5 (0.037-0.050 mM) in the same buffer. For each ITC experiment, a 180-s delay at the start of the experiment was followed by 30 injections of 10 μl of the titrant solution, spaced 300-s apart and the sample was stirred at 300rpm throughout the experiment. Binding stoichiometry (N), dissociation constant ( $K_d$ ), standard enthalpy ( $\Delta H$ ), entropy ( $\Delta S$ ), and free energy ( $\Delta G$ ) changes associated with *Win* peptide binding to WDR5 were derived by fitting the binding isotherm to an one-site binding model (Origin 7.0).

***MLL1 core complex inhibition assays:*** MALDI-TOF mass spectrometry based

methyltransferase assays were performed as described previously (61). Methylation assays were carried out in the absence and presence of increasing concentrations of the different *Win* motif peptide inhibitors. 7.3  $\mu\text{M}$  of the MLL1 core complex [MLL1<sup>3745</sup> (amino acid residues 3745-3969), WDR5, RbBP5 and Ash2L] was incubated with 250 $\mu\text{M}$  S-adenosyl-methionine (Sigma Aldrich), 10 $\mu\text{M}$  histone H3 peptide (residues 1-20) (Global Peptides) and *Win* peptides (0-1 mM) at 15°C in an assay buffer containing 50 mM Tris-Cl (pH 9.0), 200 mM sodium chloride, 3 mM dithiothreitol, 1 $\mu\text{M}$  Zinc chloride, and 5% glycerol. The reactions were quenched after 12 hours by the addition of trifluoroacetic acid to a final concentration of 0.5%. The quenched samples were diluted 1:5 in deionized water and mixed 1:5 with  $\alpha$ -cyano-4-hydroxycinnamic acid. MALDI-TOF mass spectrometry analysis was carried out on Bruker AutoFlex mass spectrometer (State University of New York, College of Environment and Forestry, Syracuse, NY) operated in reflectron mode. Final spectra were averaged from 200 shots per position at 10 different positions. For the pre-incubation methylation assays, 7.3  $\mu\text{M}$  of the WRA-subcomplex was pre-incubated with increasing concentrations of the different *Win* motif peptides (0-250 $\mu\text{M}$ ) for 1 hour and then mixed with 7.3  $\mu\text{M}$  MLL1<sup>3745</sup>. The methylation reactions and MALDI-TOF analysis were performed as described above. MLL1 core complex inhibition assays with WDR5 were carried out similarly with 0, 5, 10 or 30-fold excess WDR5 as compared to WRAD (7.3  $\mu\text{M}$ ) and the relative amounts of unmodified, mono-, di- and trimethylated histone H3 peptides were quantified after 12 hours.

***Curve fitting and IC<sub>50</sub> analysis:*** MALDI-TOF mass spectrometry was used to determine the relative distribution of unmodified, mono-, di-, and trimethylated species in each reaction as

described previously (61). IC<sub>50</sub> value is defined as the concentration of the *Win* motif peptide that is required to inhibit the H3K4 dimethylation activity of the MLL1 core complex by 50% of the initial value. IC<sub>50</sub> values were determined by plotting the percent dimethylation as a function of *Win* peptide concentration. The data was fitted using Sigma plot 11.0 to a four parameter logistic model (Equation 1):

**Equation 1:**  $y = c + ((a-c)/(1 + ((x/IC_{50})^b)))$ ,

where a = % inhibition at zero *Win* peptide concentration, c = percent inhibition at infinitely high peptide concentration, x is the concentration of inhibitor peptide, y is percent dimethylation, and b= Hill slope. IC<sub>50</sub> values derived from the fits were used to calculate the inhibition constants (K<sub>i</sub>) utilizing the Cheng-Prusoff relation (317) (Equation 2):

**Equation 2:**  $K_i = IC_{50} / (1 + [S]/K_d)$ ,

where S=concentration of the MLL1 (7.3 μM) used in the methylation assays and K<sub>d</sub>= dissociation constant for the interaction between MLL1<sup>3745</sup> and WDR5 proteins, previously determined to be 120 nM (61).

***Nucleosome methylation inhibition assays:*** Radiometry based methyltransferase assays were performed as described previously (33). Nucleosome methylation assays were carried out in the absence and presence of increasing concentrations of the MLL3 *Win* motif peptide inhibitor. Histones were expressed, purified, refolded, and reconstituted into nucleosomes using the small

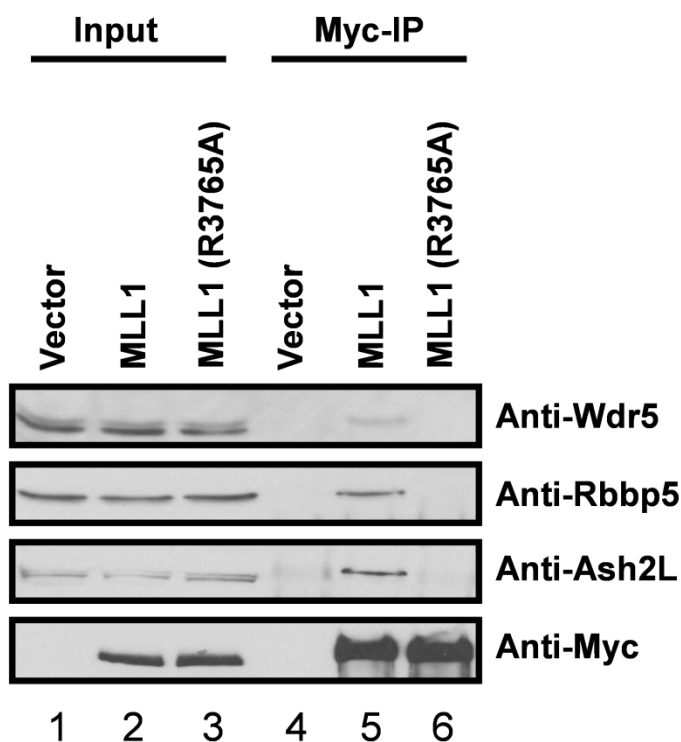
scale reconstitution procedure as described by Dyer *et al.* (318) using the 255-base pair 54A54 double-stranded DNA fragment from the murine mammary tumor virus 3-LTR promoter and flanking sequences as described by Flause *et al.* (319). 7.3  $\mu\text{M}$  of the MLL1 core complex was incubated with 1  $\mu\text{Ci}$  of [ $^3\text{H}$ ]methyl *S*-adenosylmethionine ([ $^3\text{H}$ ]-SAM; MP Biomedicals), 1  $\mu\text{l}$  (0.325 mg/ml) of fully assembled nucleosome and MLL3 *Win* motif peptide (0-4 mM) in an assay buffer containing 50 mM Tris-Cl (pH 8.5), 200 mM sodium chloride, 3 mM dithiothreitol, 5 mM magnesium chloride, and 5% glycerol. The reactions were carried out at 15°C for 2 hours and quenched by the addition of 5X SDS loading buffer. The reactions were separated on a 18% Tris-glycine PAGE and the destained gels were soaked for 30 min in autoradiography enhancer solution (ENLIGHTNING; PerkinElmer Life Science), vacuum dried, and exposed to film at -80°C for 10-days.

## RESULTS

***R3765 of the MLL1 Win motif is required for the co-immunoprecipitation of WDR5, RbBP5 and Ash2L from mammalian cells (the co-immunoprecipitation experiments were done by J. H. Lee in David Skalnik's laboratory).***

We have previously established that R3765 of the MLL1 *Win* motif is required for the assembly and H3K4 dimethylation activity of MLL1 core complex *in vitro* (61). To determine if R3765 of MLL1 is required for the assembly of the MLL1 core complex in mammalian cells, we compared wild-type and R3765A MLL1 proteins for their ability to co-immunoprecipitate WDR5, RbBP5 and Ash2L (WRA) from HEK293 cells. pCMV-Myc vectors encoding the 180 kDa C-terminal fragment of wild-type or R3765A MLL1 proteins were transfected into HEK293 cells, immunoprecipitated with antibodies against c-Myc, and probed by western blotting for the presence of WDR5, RbBP5 and Ash2L. As shown in Figure 3.1 (lane 5), while the wild-type

MLL1 co-immunoprecipitates the WRA sub-complex, co-immunoprecipitation is abolished when R3765 of MLL1 is replaced with alanine (Figure 3.1, lane 6). These results are consistent with the suggestion from our in vitro studies that R3765 of the MLL1 *Win* motif is crucial for the assembly of the MLL1 core complex in cells.



**Figure 3.1 R3765 of MLL1 *Win* motif is required for the co-immunoprecipitation of MLL1 core complex components**

In lanes (1-3), western blots using antibodies against WDR5, RbBP5, Ash2L and Myc show that equal amounts of the MLL1 wild-type (lane 2) and MLL1<sup>R3765A</sup> mutant (lane 3) proteins are present in the input. Lane 1 shows the vector only control. Lanes 4-6 show western blots from samples immunoprecipitated with the anti-Myc antibody. WDR5, RbBP5, and Ash2L are co-immunoprecipitated along with wild-type MLL1 (lane 5), but are absent in the MLL1 R3765A mutant (lane 6) or vector only control (lane 4).

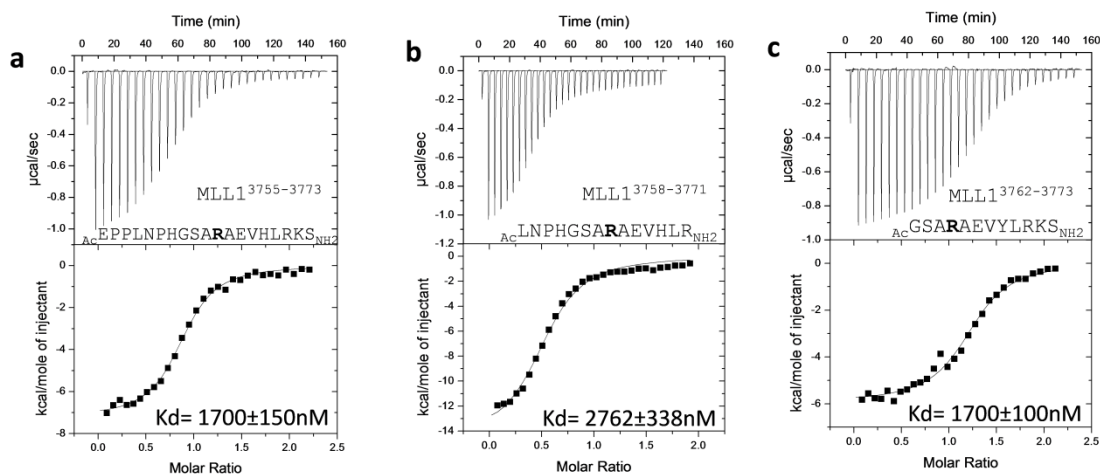
### ***Thermodynamic binding analysis of human SET1 family Win motif peptides***

In our previous crystal structure of WDR5 bound to the 12-residue MLL1 *Win* motif peptide (**GSARAEVHLRKS**), we observed that the N-terminus was well ordered while the last three residues of the C-terminus were disordered (60). The high quality of density at the N-terminus suggested that additional interactions may be observed using peptides with additional N-terminal amino acids. However, Isothermal Titration Calorimetry (ITC) binding experiments with a longer 19-residue MLL1 *Win* motif peptide (EPPLNPH**GSARAEVHLRKS**) shows identical affinity, suggesting that our original *Win* motif peptide captured the majority of the salient interactions (60). To better understand SET1 family *Win* motif recognition by WDR5, we obtained synthetic *Win* motif peptides derived from the six human SET1 family members: MLL1-4, SETd1a and SETd1b. On the basis of our previous structural results, the length of each peptide was 14-residues and contained the conserved 6-residue *Win* motif sequence with an additional 4 amino acid residues flanking both the N- and C-termini (Tables 3.1-3.2). All peptides were synthesized with acetyl- and amide capping groups to prevent unnatural charge-charge interactions from influencing the results.

Using these peptides, we first compared the 14-residue MLL1 *Win* motif peptide (LNPH**GSARAEVHLR**, conserved *Win* motif sequence highlighted in bold) with that of the original 12- and 19-residue MLL1 *Win* motif peptides (tested in (60)) for their ability to bind WDR5 by ITC (Figure 3.2). While the 12- and 19-residue MLL1 peptides bind WDR5 with identical affinity (1.7  $\mu$ M), the 14-residue MLL1 peptide binds ~2-fold weaker with a dissociation constant of 2.8  $\mu$ M (Figure 3.2 and Table 3.1). It is likely that this difference is due to the absence of lysine and serine residues on the C-terminus of the 14-residue peptide, as it is



present in both 12- and 19-residue peptides that bind WDR5 with identical affinity. A similar conclusion was reached upon deletion of the RKS residues of a *Win* motif peptide in a related analysis (310). However, it is not clear how the RKS residues contribute to the binding affinity for WDR5 as they are disordered in all X-ray structures that contain these sequences (60, 68), and therefore were not included in the peptides used for further analysis. Because of the similarity in length to all other *Win* motif peptides in this study and the relatively small differences in binding affinities, we used the 14-residue MLL1 *Win* motif peptide as the basis for comparison for all other human SET1 family *Win* motif peptides. The thermodynamic binding parameters such as dissociation constants ( $K_d$ ), number of binding sites (N), enthalpy change ( $\Delta H$ ), entropy change ( $\Delta S$ ), and free energy change ( $\Delta G$ ) derived from the ITC experiments for different MLL1 *Win* motif peptides are summarized in Table 3.1.



**Figure 3.2 Different length MLL1 *Win* motif peptides bind WDR5 with similar affinities**  
 Isothermal Titration Calorimetry (ITC) data show that the different MLL1 *Win* motif peptides have similar affinities for WDR5. (a-c). ITC data for WDR5 binding to MLL1<sup>3755-3773</sup> (a), MLL1<sup>3758-3771</sup> (b), and MLL1<sup>3762-3773</sup> (c) *Win* motif peptides. Upper panels show heat of binding plotted as a function of time. Lower panels show the binding isotherms fit to a one-site binding model. MLL1 *Win* peptide sequence and the dissociation constants ( $K_d \pm \text{S.E.M.}$ ) derived from the fit are indicated.

Table 3.1: Thermodynamic binding parameters for the different MLL1 *Win* motif peptides

MLL1 motif	Win	Peptide Sequence	Dissociation Constant (Kd) $\pm$ S.E.M, nM	Enthalpy Change ( $\Delta H$ ) $\pm$ S.E.M, kCal/mol	Entropy Change ( $\Delta S$ ) $\pm$ S.E.M, Cal/mol/K	TAS $\pm$ S.E.M, kCal/mol	Number of binding sites (N) $\pm$ S.E.M.	Free energy Change ( $\Delta G$ ) $\pm$ S.E.M, kCal/mol
<sup>a</sup> MLL1 <sup>3755-3773</sup>		Ac <sup>1</sup> EPPLNPHGSARAEVHLRKS <sub>NH2</sub>	1700 $\pm$ 150	-7.80 $\pm$ 0.42	-0.45 $\pm$ 1.56	-0.13 $\pm$ 0.45	1.20 $\pm$ 0.32	-7.71 $\pm$ 0.52
MLL1 <sup>3758-3771</sup>		Ac <sup>1</sup> LNPHGSARAEVHLR <sub>NH2</sub>	2762 $\pm$ 338	-5.80 $\pm$ 0.92	4.47 $\pm$ 3.23	1.31 $\pm$ 0.95	1.48 $\pm$ 0.34	-7.09 $\pm$ 0.04
<sup>a</sup> MLL1 <sup>3762-3773</sup>		Ac <sup>1</sup> GSARAEVHLRKS <sub>NH2</sub>	1700 $\pm$ 100	-5.77 $\pm$ 0.10	7.04 $\pm$ 0.07	2.06 $\pm$ 0.20	1.40 $\pm$ 0.17	-7.83 $\pm$ 0.10

We next compared the binding of WDR5 to the different 14-residue SET1 family *Win* motif peptides using ITC (Figure 3.3 and Table 3.2). Despite the high sequence conservation of the six-residue *Win* motif within each peptide, we observed a wide range of affinities for WDR5 among different peptides. For example, while the 14-residue MLL1 *Win* motif peptide binds WDR5 with a dissociation constant ( $K_d$ ) of  $\sim 2,800$  nM, the other human SET1 family *Win* motif peptides bind with 5- to 51-fold greater affinity (Table 3.2). These results suggest that the non-conserved residues flanking the *Win* motif contributes to the specificity of WDR5 for different SET1 family *Win* motifs. The MLL3 *Win* motif peptide binds with the greatest affinity at 54 nM, followed by MLL2 (75 nM), MLL4 (88 nM) SETd1b (103 nM), SETd1a (541 nM), and MLL1 (2,762 nM). The thermodynamic binding parameters for different SET1 family *Win* motif peptides are summarized in Table 3.2. The binding data shows that all six SET1 family *Win* motif peptides bind WDR5 with a one-to-one stoichiometry (Table 3.2, column 7) suggesting that there is a single *Win* motif binding pocket in WDR5.

### Figure 3.3 Human SET1 family *Win* motif peptides bind WDR5 with a wide range of affinities

ITC data for the binding of human SET1 family *Win* motif peptides to full-length WDR5. Isothermal Titration Calorimetry (ITC) data showing that *Win* motif peptides derived from human SET1 family members have differential affinities for WDR5. (a-f). ITC data for WDR5 binding to MLL1 (a), MLL2 (b), MLL3 (c), MLL4 (d), SETd1a (e), and SETd1b (f) *Win* motif peptides. Upper panels show heat of binding plotted as a function of time. Lower panels show the binding isotherms fit to a one-site binding model. Peptide sequence and dissociation constants ( $K_d \pm S.E.M.$ ) derived from the fits are indicated.

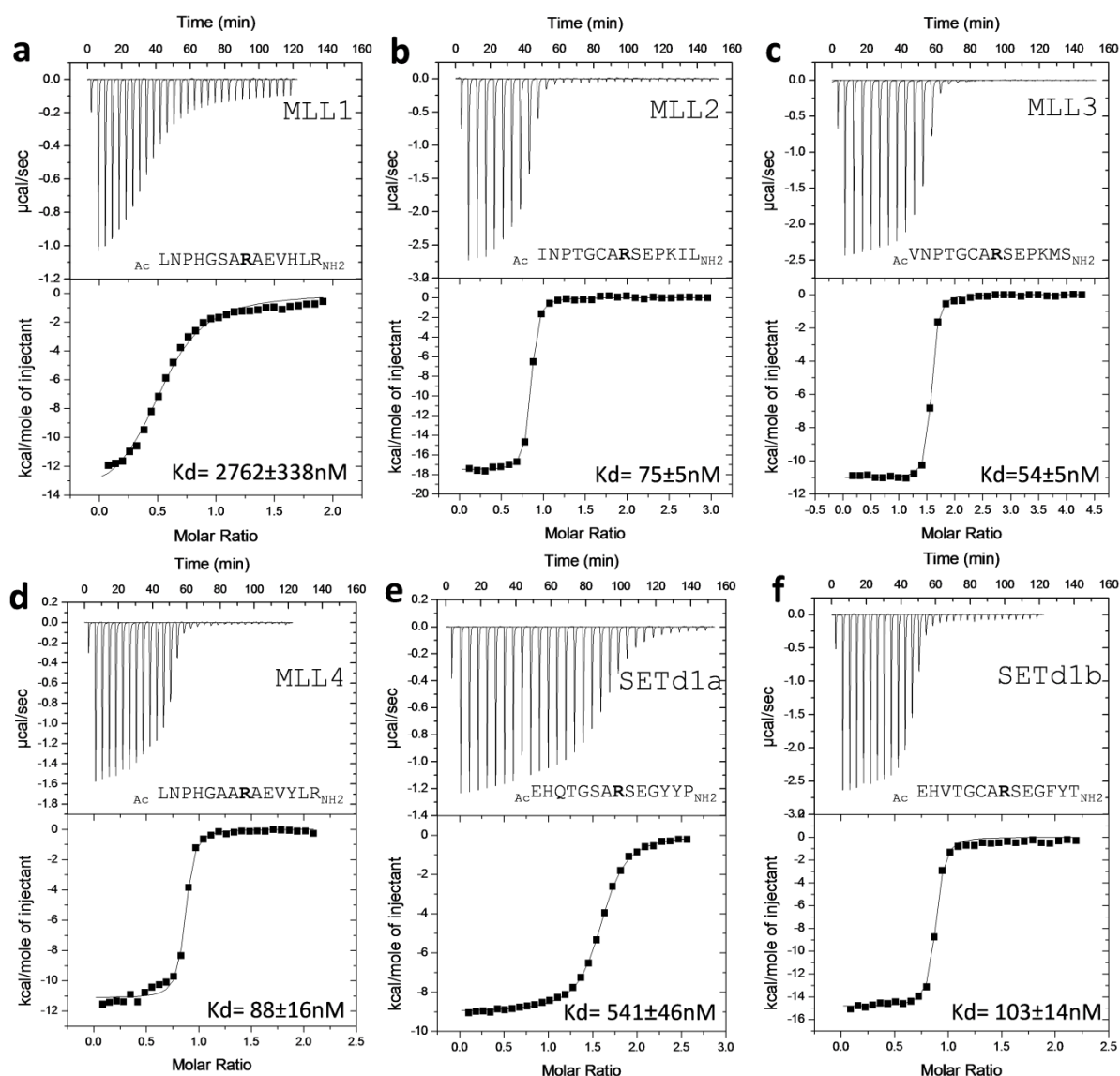


Table 3.2: Summary of thermodynamic binding parameters for the binding of human SET1 family *Win* motif peptides to WDR5

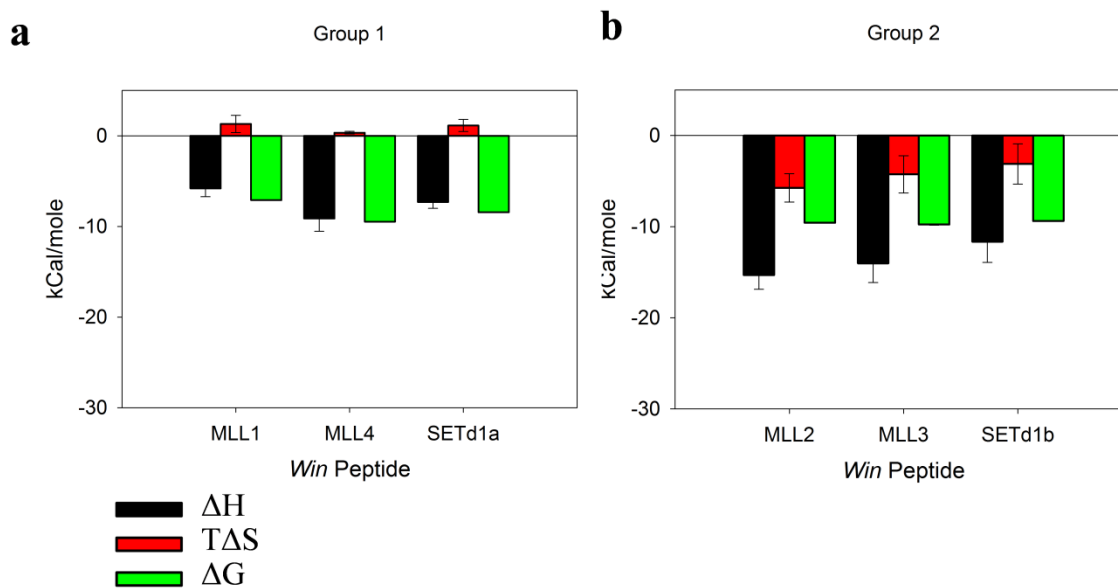
SET1 family <i>Win</i> motif	<i>Win</i> motif sequence	Dissociation Constant (Kd) $\pm$ S.E.M, nM	Enthalpy Change( $\Delta H$ ) $\pm$ S.E.M, kCal/mol	Entropy Change( $\Delta S$ ) $\pm$ S.E.M, Cal/mol/K	TAS $\pm$ S.E.M, kCal/mol	Number of binding sites (N) $\pm$ S.E.M	Free energy Change ( $\Delta G$ ) $\pm$ S.E.M, kCal/mol
MLL1	$\lambda_C$ -LNPHGSARAEVHLR $_{NH2}$	2762 $\pm$ 338	-5.80 $\pm$ 0.92	4.47 $\pm$ 3.23	1.31 $\pm$ 0.95	1.48 $\pm$ 0.34	-7.09 $\pm$ 0.04
MLL2	$\lambda_C$ -INPTGCARSEPKIL $_{NH2}$	75 $\pm$ 5	-15.32 $\pm$ 1.56	-19.60 $\pm$ 5.30	-5.74 $\pm$ 1.56	0.87 $\pm$ 0.04	-9.58 $\pm$ 0.00
MLL3	$\lambda_C$ -VNPTGCARSEPKMS $_{NH2}$	54 $\pm$ 5	-14.03 $\pm$ 2.11	-14.54 $\pm$ 6.97	-4.26 $\pm$ 2.04	1.21 $\pm$ 0.23	-9.77 $\pm$ 0.07
MLL4	$\lambda_C$ -LNPHGAARAEVYLR $_{NH2}$	88 $\pm$ 16	-9.13 $\pm$ 1.41	-7.12 $\pm$ 0.82	0.34 $\pm$ 0.15	1.13 $\pm$ 0.14	-9.47 $\pm$ 0.03
SET1a	$\lambda_C$ -EHQTGCARSEGYYP $_{NH2}$	541 $\pm$ 46	-7.30 $\pm$ 0.71	6.54 $\pm$ 4.10	1.13 $\pm$ 0.67	1.53 $\pm$ 0.03	-8.43 $\pm$ 0.04
SET1b	$\lambda_C$ -EHVTGCARSEGFYT $_{NH2}$	103 $\pm$ 14	-11.65 $\pm$ 2.27	-10.67 $\pm$ 7.54	-3.13 $\pm$ 2.21	1.15 $\pm$ 0.21	-9.39 $\pm$ 0.06

### ***Human SET1 family Win motif peptides have different enthalpic and entropic binding signatures***

Comparison of the enthalpy ( $\Delta H$ ) and entropy ( $\Delta S$ ) changes associated with the binding of human SET1 family *Win* motif peptides to WDR5 can provide valuable insights into the driving forces involved in the differential binding affinity. Enthalpic interactions are typically driven by hydrogen bonding and van der Waals interactions whereas favorable entropic binding affinity is derived from hydrophobic interactions (320-323). Thermodynamic binding data (Table 3.2 and Figure 3.3) suggest that the hydrogen bonding and van der Waals interactions are the primary forces that dictate the binding of all *Win* motif peptides to WDR5 (as indicated by large negative values of  $\Delta H$  in Table 3.2, column 4). The entropic contribution (hydrophobic interactions and de-solvation effects) to the overall binding affinity is relatively smaller (indicated by smaller values of  $T\Delta S$  in Table 3.2, column 6). Interestingly, the *Win* motif peptides MLL1, MLL4, and SETd1a have positive  $T\Delta S$  values suggesting that these peptides might utilize somewhat different binding forces (additional hydrophobic interactions and/or changes in solvent entropy) to bind WDR5 as compared to MLL2, MLL3, and SETd1b (negative values of  $T\Delta S$ ) (Table 3.2 and Figure 3.4). Based on the thermodynamic analysis, the human SET1 family *Win* motif peptides are grouped into two different thermodynamic binding signatures: Group1 (includes MLL1, MLL4, and SETd1a) that has a more favorable  $\Delta H$  and less favorable  $T\Delta S$  (characteristics of hydrogen bond contribution, van der Waals, and fewer hydrophobic interactions) and Group 2 (includes MLL2, MLL3, and SETd1b) that has a favorable  $\Delta H$  and a more favorable  $T\Delta S$  (characteristics of hydrogen bond, van der Waals, and significant contributions from hydrophobic interactions/de-solvation entropy) (Figure 3.4).

A similar thermodynamic analysis has been reported for HIV-protease inhibitors: Indinavir, nelfinavir, saquinavir, ritonavir, amprenavir, lopinavir, KNI-272 and KNI-764 that were used to deduce guidelines for the design of HIV-1 protease inhibitors with greater specificity (favorable  $\Delta H$ ) and membrane solubility (favorable  $T\Delta S$ ) (320-323). *Win* motif peptides could be subjected to a similar thermodynamic analysis to determine their initial binding signatures that could then be optimized to identify new lead compounds with better inhibitory properties. Based on this analysis, we predict that the MLL2 *Win* motif peptide that has the largest  $\Delta H$  value (highest specificity derived from hydrogen bonds) would represent a good starting point for the design of additional *Win* motif-like inhibitors by incorporating new hydrophobic features required for enhanced membrane solubility.





**Figure 3.4 Thermodynamic binding signatures for Human SET1 family *Win* motif peptides**

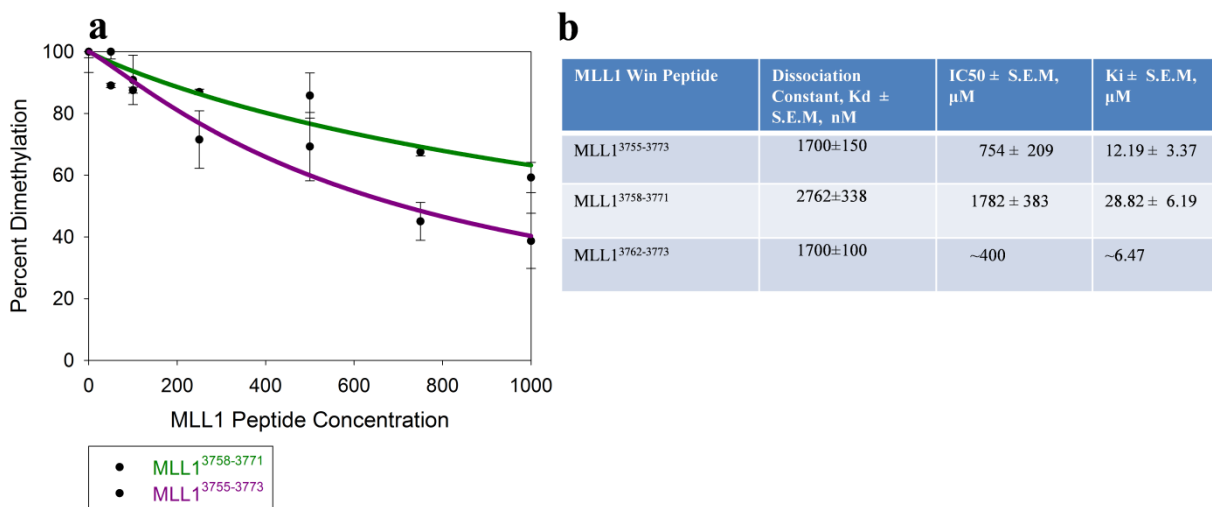
Thermodynamic binding signatures for the binding of six different human SET1 family *Win* motif peptides to WDR5. *Win* motif peptides are classified based on the enthalpic ( $\Delta H$ ), entropic ( $T\Delta S$ ) and free energy ( $\Delta G$ ) changes associated with binding to WDR5. Group 1: MLL1, MLL4 and SETd1a (a). Group 2: MLL2, MLL3, and SET1b (b). See text for further description.

***SET1 family Win motif peptides are potent inhibitors of the H3K4 dimethylation activity of the MLL1 core complex***

We have previously demonstrated that the 12-residue MLL1 *Win* motif peptide (3762-3773) inhibits the H3K4 dimethylation activity of the MLL1 core complex in a concentration dependent manner by competing with MLL1 for the arginine-binding pocket of WDR5 (61). Since all six human SET1 family *Win* motif peptides described here bind WDR5 with similar or better interaction affinities as compared to that of the 12-residue MLL1 *Win* motif peptide, we predicted that the other SET1 family *Win* motif peptides would also inhibit the H3K4 dimethylation activity of the MLL1 core complex. Moreover, since the different human SET1 family *Win* motif peptides differ in their affinities to bind WDR5, we further hypothesized that these *Win* motif peptides will inhibit the H3K4 dimethylation activity with varying efficiencies. To determine if other human SET1 family *Win* motif peptides also inhibit the enzymatic activity of the MLL1 core complex, we compared inhibition constants among different *Win* motif peptides in enzymatic assays with the fully assembled MLL1 core complex using MALDI-TOF mass spectrometry. As a negative control, we performed the same assays with a peptide derived from the P53 tumor suppressor protein containing amino acids HSSHLKSKKGQSTSRHKK (P53<sup>365-382</sup>).

Using this strategy, we first compared the inhibition constants among the different MLL1 *Win* motif peptides. While we previously showed that the 12-residue MLL1 *Win* motif peptide has an IC<sub>50</sub> of ~400 μM (61), the 14-residue MLL1 *Win* motif peptide inhibits with an IC<sub>50</sub> of 1.8 mM (Figure 3.5b). However, the 19-residue *Win* motif peptide, which binds to WDR5 with similar affinity as compared to the 12-residue peptide, inhibits the H3K4 dimethylation activity

with an  $IC_{50}$  of 750  $\mu$ M (Figure 3.5a-b). The ~2-fold difference in  $IC_{50}$  values between the 19- and 14-residue MLL1 *Win* motif peptides is consistent with the differences observed in the binding affinities of the different MLL1 *Win* motif peptides (Figure 3.5).



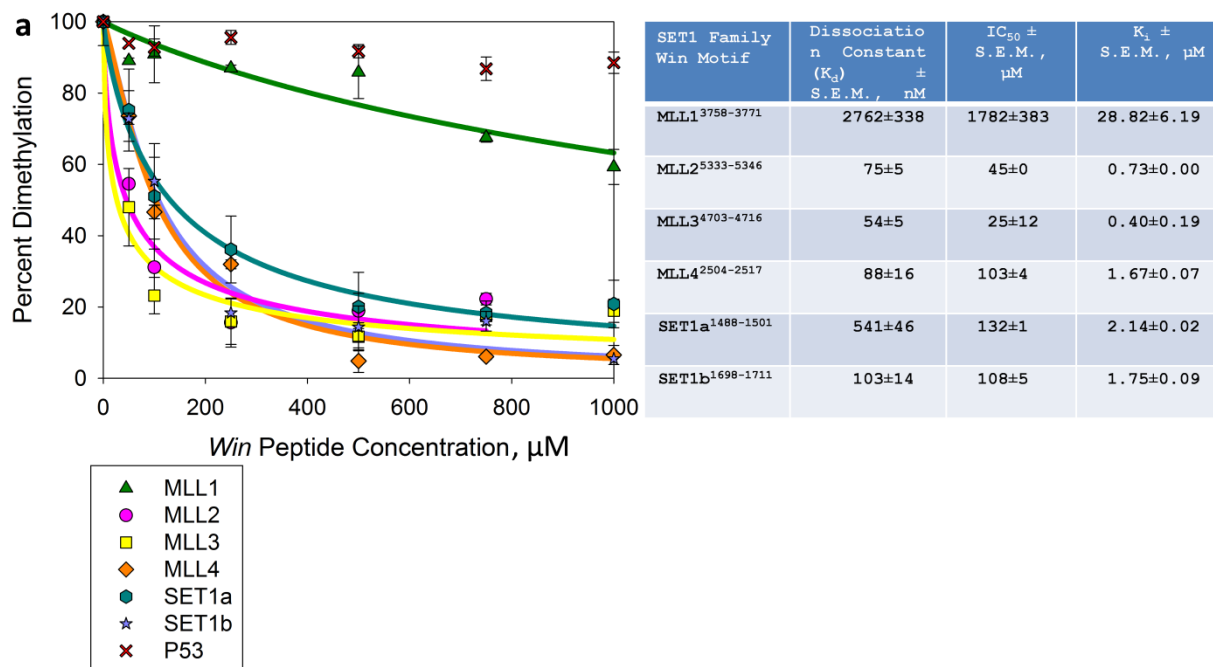
**Figure 3.5 Comparison of inhibition constants for different MLL1 *Win* motif peptides**

MLL1 *Win* motif peptides inhibit the H3K4 dimethylation activity of MLL1 core complex with varying efficiencies (a-b). MALDI-TOF assays for the H3K4 methylation activity catalyzed by the MLL1 core complex (MLL<sup>3745</sup>, WDR5, RbBP5 and Ash2L). MALDI-TOF assays were carried out with 7.3  $\mu$ M MLL1 core complex for 12 hours in the absence or presence of increasing concentrations of MLL1<sup>3755-3773</sup>, MLL1<sup>3758-3771</sup>, and MLL1<sup>3762-3773</sup> *Win* motif peptides. In a, relative H3K4 dimethylation levels in the presence of increasing concentrations of different MLL1 *Win* motif peptides are shown. Error bars represent the variation observed in two or three independent experiments. Percent dimethylation at zero *Win* peptide concentration is normalized to 100%. In b, IC<sup>50</sup> values derived from fitting the data to a four parameter logistic model described in equation 1. Inhibition constants (K<sub>i</sub>) were derived from cheng-prusoff equation (see equation 2) using a substrate concentration (S) of 7.3  $\mu$ M and a dissociation constant K<sub>d</sub> of 120 nM (MLL1-WDR5 complex interaction, (61)).

We next compared the inhibition constants among the six different 14-residue human SET1 family *Win* motif peptides. The results show that all SET1 family *Win* motif peptides inhibit the H3K4 dimethylation activity of the MLL1 core complex in a concentration dependent manner (Figure 3.6). For example, at 6-, 30- and 60-fold molar excesses of all six human SET1 family *Win* motif peptides, the relative amounts of dimethylation catalyzed by the MLL1 core complex decreases dramatically while the monomethylation activity is correspondingly increased (Appendix Figures 3a-f). In contrast, even with a 120-fold excess of a control arginine-containing p53 peptide, the H3K4 dimethylation activity is not significantly altered (Appendix Figure 3g). The  $IC_{50}$  values for the inhibition of H3K4 dimethylation activity of MLL1 core complex derived from these experiments are summarized in Figure 3.6b.

Interestingly, all other 14-residue human SET1 family *Win* motif peptides inhibit with  $IC_{50}$  values that are between 14- to 72-fold lower than that of the 14-residue MLL1 *Win* motif peptide (Figure 3.6 and Appendix 3a-g). Little inhibition was observed with the P53 peptide, which could not be fit to the same model described in equation 1 (Figure 3.6a). The MLL3 *Win* motif peptide is the best inhibitor of the H3K4 dimethylation activity of the MLL1 core complex (Figure 3.6b) with an  $IC_{50}$  value of 25  $\mu$ M, followed by MLL2, MLL4, SETd1b, SETd1a, and MLL1. The relative inhibitory efficiencies of the human SET1 family *Win* motif peptides correlate well with the differences in their binding affinities for WDR5 as measured by ITC (Figure 3.6b). The original MALDI-TOF spectra for the relative distribution of different H3K4 methylation (unmodified, mono-, di-, and trimethylation) levels in the absence or presence of different *Win* motif peptides is presented in the Appendix 3a-g. The MALDI-TOF inhibition assays described in Figure 3.6 were classified as *no pre-incubation* experiments due to the fact

that the *Win* motif peptides were added to a pre-assembled MLL1 core complex (MLL1 and WRA). The next set of inhibition assays are classified as pre-incubation experiments due to the fact that *Win* motif peptides were pre-incubated with the WRA-subcomplex before mixing with MLL1<sup>3745</sup>.



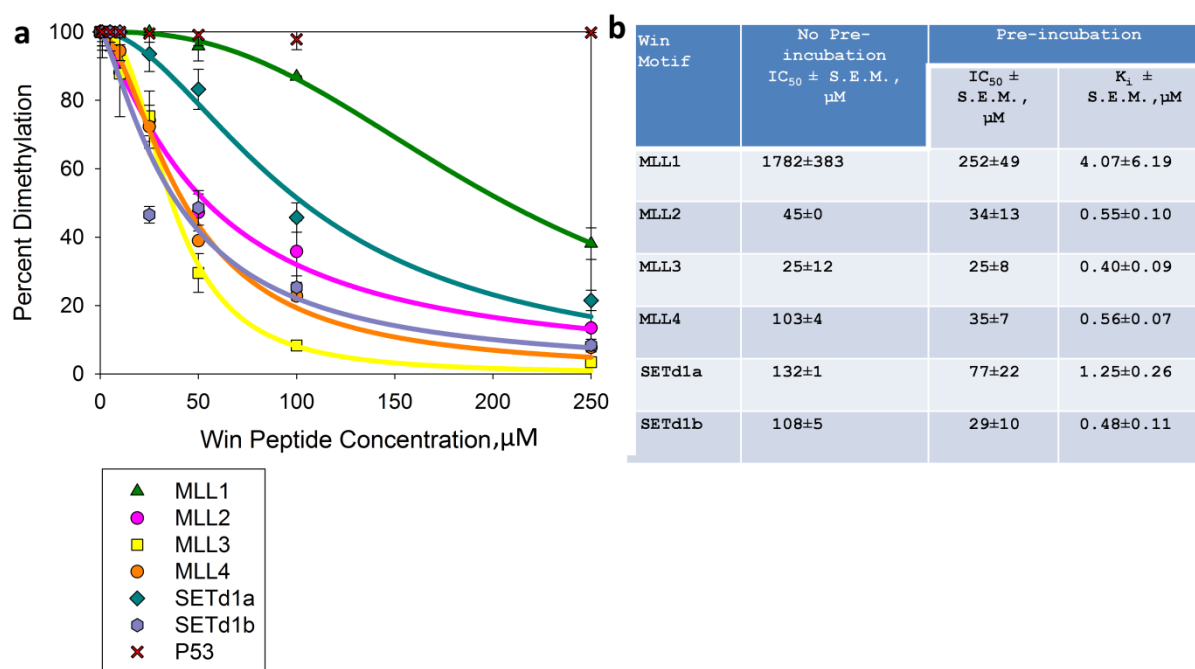
**Figure 3.6 Human SET1 family *Win* motif peptides inhibit the H3K4 dimethylation activity of the MLL1 core complex with varying efficiencies**

SET1 family *Win* motif peptides inhibit the H3K4 dimethylation activity of the MLL1 core complex. a-b. MALDI-TOF assays for the H3K4 methylation activity catalyzed by the MLL1 core complex (MLL<sup>3745</sup>, WDR5, RbBP5 and Ash2L). MALDI-TOF assays were carried out with 7.3  $\mu\text{M}$  MLL1 core complex for 12 hours in the absence or presence of increasing concentrations of MLL1 (green), MLL2 (pink), MLL3 (yellow), MLL4 (orange), SETd1a (cyan), SETd1b (dark blue) *Win* motif peptides or the control P53 peptide (red). a. Relative H3K4 dimethylation levels in the presence of increasing concentrations of *Win* motif peptides or control P53 peptide. Error bars represent the variation observed in two or three independent experiments. b.  $IC_{50}$  values and inhibition constants ( $K_i$ ) were derived as described in Figure 3.5. MALDI-TOF spectra for non pre-incubation experiments are shown in Appendix 3.

We next asked if pre-incubation of *Win* motif peptides with the WRA sub-complex before mixing with MLL1<sup>3745</sup> alters the ability of *Win* motif peptides to compete with MLL1<sup>3745</sup> in our MALDI-TOF methylation assays. To test this hypothesis, we pre-incubated the WRA-sub complex with a 1-, 6-, 12- or 30-fold excess of different SET1 family *Win* motif peptides and examined the H3K4 dimethylation activity catalyzed by the MLL1 core complex (Figure 3.7 and Appendix Figure 4a-g). It should be noted that the MLL1 *Win* motif peptide used in the pre-incubation assays is the 19-residue peptide while the no pre-incubation assays were carried out with the 14-residue MLL1 peptide. Comparison of IC<sub>50</sub> values for the different *Win* motif peptides shows that MLL3 and SET1b *Win* motif peptides have similar IC<sub>50</sub> values of 25 μM and 29 μM, respectively, when pre-incubated with the WRA sub-complex (Figure 3.7). These IC<sub>50</sub> values are 10-fold better than that of the MLL1 *Win* motif peptide that has an IC<sub>50</sub> value of 251 μM (Figure 3.7 and appendix figures IVa-g). Similar experiments with MLL2 and MLL4 *Win* motif peptides show that these peptides inhibit the H3K4 dimethylation activity with similar IC<sub>50</sub> values of 34 μM, which is 7-fold better than the MLL1 *Win* motif peptide (Figure 3.7b). Finally, when pre-incubated with the WRA sub-complex, the SET1a *Win* motif peptide inhibits the H3K4 dimethylation activity of the MLL1 core complex with an IC<sub>50</sub> value of 77 μM (Figure 3.7b) that is in correlation with its intermediate interaction affinity to WDR5. However, when the same experiments were conducted with the control p53 peptide, no change in H3K4 dimethylation activity was observed, even with a 30-fold excess of peptide (Figure 3.7 and appendix Figure 4g). Overall, the inhibition studies show that the human SET1 family *Win* motif peptides are 2 to 3-fold better inhibitors of H3K4 dimethylation activity when pre-incubated with the WRA sub-complex than when the *Win* peptides are not pre-incubated. These results suggest



that human SET1 family *Win* motif peptides are better at inhibiting the assembly of MLL1 core complex as compared to inhibiting pre-formed MLL1 core complexes. The SET1 family *Win* motif peptides ranked in the order of strongest to weakest inhibitor efficiencies when pre-incubated with WRA-sub complex are as follows: MLL3> SET1b> MLL2> MLL4> SET1a> MLL1.



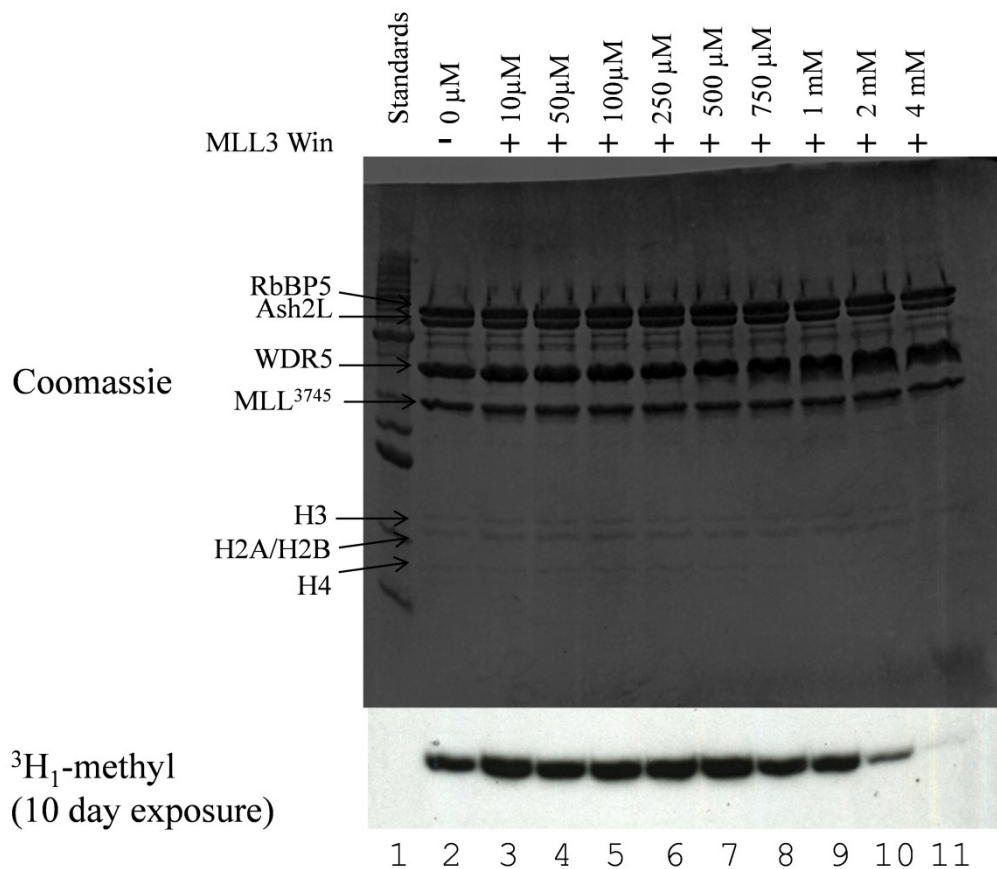
**Figure 3.7 Pre-incubation with WRAD lowers the inhibition constants for SET1 family *Win* motif peptides**

SET1 family *Win* motif peptides are better inhibitors of the H3K4 dimethylation activity of the MLL1 core complex when pre-incubated (a-b). MALDI-TOF assays were carried out as described in 3.6. For pre-incubation, increasing concentrations of human SET1 family *Win* motif peptides were incubated with 7.3  $\mu\text{M}$  of WRAD sub-complex for 1 hour and then mixed with 7.3  $\mu\text{M}$  of MLL1<sup>3745</sup> to initiate the methylation reaction. a. Relative H3K4 dimethylation levels in the presence of increasing concentrations of *Win* motif peptides or control P53 peptide. Error bars represent the variation observed in two or three independent experiments. b.  $\text{IC}_{50}$  values and inhibition constants ( $K_i$ ) were derived as described in Figure 3.5. MALDI-TOF spectra for pre-incubation experiments are shown in Appendix 4.

***Nucleosomal H3 methylation by MLL1 core complex is inhibited by the MLL3 Win motif peptide***

Based on these studies, MLL3 *Win* motif peptide was identified as the best inhibitor of the H3K4 dimethylation activity of MLL1 core complex. However, the MALDI-TOF assays described earlier were carried out using a histone H3 peptide (N terminal 1-21 residues of histone H3) substrate and it is possible that the histone peptide may not fully recapitulate the additional features present in the physiological substrate for MLL1 core complex, which is a nucleosome. Indeed, it was previously shown that while the individual methyltransferases, MLL1 and WRAD, are both capable of methylating histone H3 peptides in vitro, a fully assembled MLL1 core complex is required for the methylation of a nucleosomal histone H3 (33). Since the *Win* motif peptides target an interaction that is required for the assembly of the MLL1 core complex, we predicted that *Win* motif peptides will prevent MLL1 core complex mediated nucleosomal H3K4 methylation. To test this hypothesis, we carried out radiometric histone methyltransferase assays using a nucleosome substrate in the absence or presence of increasing concentrations of MLL3 *Win* motif peptide. The results shows that while the fully assembled MLL1 core complex catalyzes efficient nucleosomal H3 methylation (Figure 3.8, lane 2), nucleosomal methylation is inhibited in the presence of MLL3 *Win* motif peptide (Figure 3.8, lanes 10-11). The fold excess MLL3 *Win* motif peptide required to inhibit nucleosomal methylation is 250-500 fold (as compared to the MLL1 concentration) and is 80-fold higher than the IC<sub>50</sub> value (25 μM) determined for the MLL3 *Win* motif peptide using MALDI-TOF assays (Figure 3.6). No inhibition was seen up to 125-fold excess MLL3 *Win* peptide concentration (Figure 3.8, lanes 3-9) when nucleosomes were used as a substrate. These results suggest that additional features present in the nucleosome substrate stabilize the interaction between MLL1 core complex and a fully assembled nucleosome H3 and hence a large excess of MLL3 *Win* motif peptide is required

to completely abolish this interaction. Nevertheless, these results demonstrate that MLL3 *Win* motif acts as a highly specific inhibitor of nucleosomal H3 methylation at higher concentrations and possibly through the disruption of the assembly of MLL1 core complex. It is interesting to speculate that complete inhibition of nucleosomal H3K4 dimethylation could be obtained by pre-incubating the MLL3 *Win* motif peptide with the WRA sub-complex before the addition of MLL1<sup>3745</sup> in the methylation reactions. Intriguingly, the methylation assays also demonstrate that at lower concentrations of MLL3 *Win* motif peptide there is a slight increase in MLL1 core complex mediated nucleosomal methylation (Figure 3.8, lane 3). One possible explanation for this slight increase in nucleosomal methylation is that one of the components of the enzymatic assays (present in excess) acts as an intrinsic inhibitor of MLL1 core complex activity and the addition of small amount of MLL3 *Win* motif peptide relieves that inhibition. Since WDR5 is the only component that is expected to bind the MLL3 *Win* motif peptide based on its three-dimensional structure, we predicted that free WDR5 can act as an intrinsic inhibitor of MLL1 core complex activity when present in excess by competing with WRAD sub-complex for binding the *Win* motif of MLL1. Methylation assays carried out using a MLL1 core complex that is purified over gel filtration (and hence should have stoichiometric amounts of individual core complex components) can help address this problem.

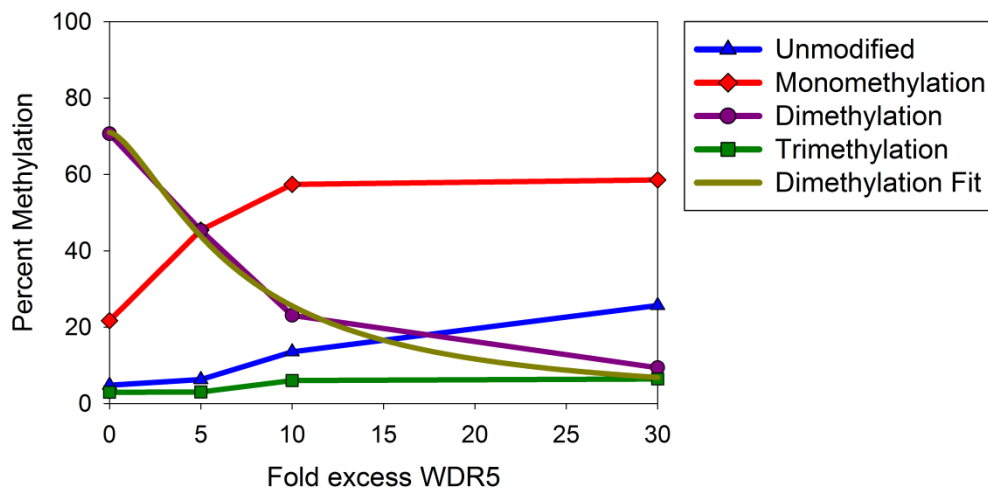


**Figure 3.8 MLL3 *Win* motif peptide inhibits MLL1 core complex mediated nucleosomal methylation**

Nucleosomal H3 methylation assays for the H3K4 methylation activity catalyzed by the MLL1 core complex (MLL<sup>3745</sup>, WDR5, RbBP5 and Ash2L). Methylation assays were carried out with 7.3  $\mu$ M MLL1 core complex for 8 hours in the absence or presence of increasing concentrations of MLL3 *Win* motif peptide. Protein bands were separated using a 4-12% SDS gel and stained with coomassie to view the protein bands and the methylated nucleosomal H3 bands were visualized using <sup>3</sup>H<sub>1</sub> fluorography.

***Free WDR5 acts as inhibitor of MLL1 core complex activity***

To test if free WDR5 inhibits MLL1 core complex mediated H3K4 methylation, I carried out MALDI-TOF based methyltransferase assays in the absence or presence of increasing concentrations of WDR5 as compared to that of WRAD sub-complex. The data demonstrate that MLL1 core complex mediated H3K4 dimethylation activity decreases with increasing concentrations of free WDR5 (Figure 3.9). The  $IC_{50}$  value derived from the fit for inhibition of H3K4 dimethylation is 50  $\mu$ M and suggests that WDR5 acts as a strong inhibitor of the MLL1 core complex. However, it should be noted that WDR5 has been previously shown to bind histone H3 peptides with a weak affinity (10-60  $\mu$ M) (64-67) and adding excess WDR5 could titrate away histone H3 peptide from the MLL1 core complex active site instead of acting as an inhibitor of MLL1 core complex. The 3-fold increase in MLL1 core complex mediated monomethylation even at 30-fold excess WDR5 (Figure 3.9) suggests that the H3 peptide can still bind to the first active site present in the MLL1 core complex (which catalyzes the monomethylation event). Therefore it is likely that the histone H3 peptide binding to the MLL1 core complex active site is not perturbed in the presence of excess WDR5 and the decrease in H3K4 dimethylation is due to WDR5's role as an inhibitor of MLL1 core complex activity.



**Figure 3.9 Free WDR5 acts as an inhibitor of MLL1 core complex mediated H3K4 dimethylation activity**

Methylation assays were carried out with 7.3  $\mu$ M MLL1 core complex for 12 hours in the absence or presence of increasing concentrations (0 to 30-fold excess) of WDR5. Relative amounts of unmodified, mono-, di-, and trimethylation on H3 peptide is plotted as a function of fold excess WDR5 concentration.

<i>Win</i> motif	Sequence	pI	Net charge @pH7.5
MLL1	$_{\text{Ace}}\text{LNPHGSARAEVHLR}_{\text{NH}_2}$	9.61	+
MLL2	$_{\text{Ace}}\text{INPTGCARSEPKIL}_{\text{NH}_2}$	8.22	+
MLL3	$_{\text{Ace}}\text{VNPTGCARSEPKMS}_{\text{NH}_2}$	8.19	+
MLL4	$_{\text{Ace}}\text{LNPHGAARAEVYLR}_{\text{NH}_2}$	8.75	+
SET1A	$_{\text{Ace}}\text{EHQTGSARSEGYYP}_{\text{NH}_2}$	5.40	-
SET1B	$_{\text{Ace}}\text{EHVTGCARSEGFYT}_{\text{NH}_2}$	5.40	-

**Table 3.3 Isoelectric points (pI) for human SET1 family *Win* motif peptides at pH 7.0**



## DISCUSSION

WDR5 is an essential component of SET1 family complexes required for regulating the degree of H3K4 methylation in cells (63). WDR5 functions to bridge interactions between MLL1 and WRAD by binding to a highly conserved WDR5 interaction or *Win* motif within the N-SET region of MLL1 (61, 68). It was previously shown that arginine 3765 of the MLL1 *Win* motif is essential for the interaction between MLL1 and WDR5 (61). Consistent with the in vitro results, we demonstrate here that R3765 of MLL1 is required for co-immunoprecipitation of WDR5, RbBP5 and Ash2L from mammalian cells, suggesting that the *Win* motif is important for the assembly of the MLL1 core complex in cells. We also showed previously that a peptide that mimics the MLL1 *Win* motif disrupts the assembly of the MLL1 core complex in vitro and inhibits its H3K4 dimethylation activity (61). These results suggest that disrupting the interaction between MLL1 and WDR5 using *Win* motif peptides or related compounds may be an effective strategy for the inhibition of H3K4 methylation in malignant cells. It is expected that such a strategy would be far more specific compared to efforts to design s-adenosyl methionine-related mimics that may bind to a wider variety of methyltransferases. A challenge remains in the identification of *Win* motif inhibitors that target preferentially MLL1 over other human SET1 family members. However, little is known about the relative stability of SET1 family complexes. To begin to address this question and to identify more potent *Win* motif inhibitors of the MLL1 core complex, we compared thermodynamic binding parameters and inhibition constants of several *Win* motif peptides derived from the other members of the human SET1 family (MLL2, MLL3, MLL4, SETd1a, and SETd1b).

Although the interaction affinity between the full-length SET family proteins and WDR5 is currently unknown, the differences in the binding affinities shown here suggest that the SET1 family complexes may have differing overall stabilities. If this hypothesis is correct, then it is predicted that the weakest complex would be disrupted at the lowest concentration of *Win* motif inhibitors before other SET1 family complexes are disrupted. The binding studies presented here suggest that MLL1 has the weakest affinity for WDR5 and therefore would likely be most sensitive to inhibition by peptides or other compounds that mimic the MLL1-WDR5 interaction interface. The increased sensitivity of the MLL1 core complex to *Win* motif peptides could provide a mechanism to selectively inhibit MLL1 over other human SET1 family members in MLL1 gain-of-function leukemias. The best peptide inhibitor we have identified is derived from the MLL3 *Win* motif sequence, which binds with 51-fold greater affinity to WDR5 than the equivalent MLL1 sequence. We have also shown that *pre-incubation* of *Win* motif peptides with WRA sub-complex lowers the inhibition constants 2 to 3-fold as compared to the inhibition constants derived from *no pre-incubation* experiments. These results suggest that *Win* motif peptides are better inhibitors in preventing the assembly of MLL1 core complex as compared to disrupting complexes that are already formed. This result could be physiologically relevant in that new MLL1-like complexes are being actively formed in rapidly dividing cells such as cancer cells and *Win* motif peptides could be useful in targeting these cells over normal cells.

We have also shown that MLL3 *Win* motif peptide is a highly potent inhibitor of MLL1 core complex mediated nucleosomal methylation. This result suggests that *Win* motif peptides will have the potential to be useful inhibitors of MLL1 core complex activity *in vivo*. We have also demonstrated that free WDR5 acts as an inhibitor of MLL1 core complex activity. This

finding is of high significance and suggests that the intracellular concentration of free WDR5 should be taken into account when designing inhibitors that target *Win* motif-WDR5 interaction.

In summary, data presented in this chapter suggests that the MLL1 *Win* motif is essential for the assembly of the MLL1 core complex in mammalian cells. We have also shown that *Win* motif peptides bind to WDR5 with a wide range of affinities, suggesting that human SET1 family complexes may be differentially sensitive to *Win* motif inhibitors-with the MLL1 core complex being the most sensitive. These results provide the basis for the rational design of improved *Win* motif inhibitors. We expect that the MLL3 *Win* motif peptide will be a good template for future rational drug design efforts.

## **CHAPTER 4: Structural analysis of the interaction between WDR5 and Human SET1 family *Win* motifs.**

This chapter is a reprint of reference (309) and modified accordingly to include additional data for this dissertation.

## INTRODUCTION

Design of *Win* motif based inhibitors with better efficiency could be greatly enhanced by determining the protein structural features within the *Win* motif that are required for the high affinity binding to WDR5. To facilitate the process of structure-based drug design and to understand the molecular basis for *Win* motif-WDR5 interaction, we previously determined the three-dimensional structure of WDR5 bound to the 12-residue (3762-3773) MLL1 *Win* motif peptide (60, 68). The structure revealed that the 12-residue MLL1 *Win* motif peptide binds WDR5 by adopting a partial  $3_{10}$ -helical conformation and by inserting the side chain of conserved *Win* motif R3765 into the central tunnel in WDR5 (60, 68). While the MLL1 *Win* peptide binding mode is similar to the previously determined structures of WDR5 bound to histone peptides, which bind by inserting R2 of histone H3 into the central tunnel in WDR5 (64-66), the MLL1 *Win* motif peptides participate in more favorable interactions with WDR5 (60, 68). These structural studies further corroborate the role of MLL1 R3765 in mediating the interaction with WDR5 and explain the high sequence conservation of arginine within SET1 family *Win* motifs.

Since the six residue *Win* motif sequence is highly conserved within SET1 family, we predicted that peptides derived from other human SET1 family members will also bind WDR5 using the same arginine binding pocket utilized by MLL1 *Win* motif peptide. Moreover, binding and inhibition studies described in Chapter 3 demonstrate that *Win* motif peptides derived from other human SET1 family members bind WDR5 with affinities that are **5 to 51-fold** better than that of the MLL1 *Win* motif peptide and suggests that there are subtle differences in the binding

mode for different *Win* motif peptides. To understand how WDR5 recognizes the different human SET1 family *Win* motifs with different binding affinities, we determined the three-dimensional crystal structures of WDR5 bound to 6-different naturally occurring human SET1 family *Win* motif sequences at resolutions ranging from 1.9-1.2 Angstroms.

The structures demonstrate that all SET1 family *Win* motif peptides bind WDR5 using the same arginine binding pocket previously shown to be utilized by the MLL1 *Win* motif or histone peptides (60, 64-68). The conserved residues of the *Win* motif adopt a similar  $3_{10}$  helical conformation. The structures also reveal that non-conserved residues C-terminal to the *Win* motif participate in different sets of interactions with WDR5. In particular, the amino acid four residues C-terminal to the conserved arginine (+4 position, refer to Figure 4.1 for residue numbering) is highly variable among the different *Win* motif peptides and may contribute to the binding energy differences through the presence or absence of an additional hydrogen bond with WDR5 residues. Moreover, the structural analysis also reveals subtle variation within the conserved *Win* motif sequence (at the -2 and +1 position), which may also contribute to binding energy differences-possibly through stabilization of the bound conformation of the peptides when free in solution. The structures also demonstrate that the residues N-terminal to the *Win* motif adopt a conformation that may further stabilize the bound conformation of the *Win* motif. To further characterize the contribution of each of these structural differences to the overall differences in binding affinity, we synthesized additional *Win* motif peptides bearing individual amino acid substitutions and quantified their contribution to WDR5 binding energy differences. The results of this structure-function analysis are described in Chapter 4.

Taken together, the structural studies reported in this chapter demonstrate how WDR5 specifically recognizes the different human SET1 family *Win* motifs. Based on the crystal structures, we have deduced some rules that could be useful for the design of additional *Win* motif peptides with better inhibitory properties.

## METHODS

***Protein expression and purification:*** A N-terminal truncated form of WDR5 (residues 23-334,  $\Delta$ N-WDR5) was expressed and purified using the conditions described previously in Chapter 2 for the full-length WDR5 (60, 61). As a final step of purification, the protein was passed through a gel filtration column (Superdex 200<sup>TM</sup> GE Healthcare) pre-equilibrated with the sample buffer containing 20 mM Tris (pH 7.5), 300 mM sodium chloride, 1 mM Tris (2-carboxyethyl) phosphine, and 1  $\mu$ M zinc chloride. The purified protein was concentrated to 13 mg/ml and used in the crystallization studies described below.

***Crystallization and Structure Determination:*** Crystals of the  $\Delta$ N-WDR5:*Win* peptide binary complex were obtained using the hanging drop vapor diffusion method. Immediately before crystallization, a 13 mg/ml stock solution of  $\Delta$ N-WDR5 in the sample buffer was mixed with a stock solution of each of the six *Win* motif peptides dissolved in the same buffer. The final concentration of  $\Delta$ N-WDR5 and the *Win* peptides were 11.7 mg/ml and 1 mM, respectively. The mother liquor used for crystallization contained 20-30 mM ammonium sulfate, 25-27% polyethylene glycol (PEG)-3350, and 100 mM HEPES (pH 7.3-7.5). The crystals were flash frozen in the mother liquor containing 40% PEG3350. Diffraction data for  $\Delta$ N-WDR5:SETd1a*Win* peptide complex was collected at the National Synchrotron Light Source

(NSLS) on the beamline X25 using quantum 210 CCD detector. The diffraction data for the other binary complexes were collected at the Cornell High Energy Synchrotron Source (CHESS) at the F1 beamline using ADSC quantum 270 CCD detector. All dataset were indexed, reduced and scaled with HKL-2000 (324) and CCP4i suite (325). Data collection statistics are summarized in Table 4.1. Each of the six *Win* motif peptide-WDR5 binary complex structures were solved by molecular replacement with MOLREP (326) using as a search model the coordinates of the previously determined structure of WDR5 (PDB code:2H68) (67). After an initial rigid body refinement, the structures were further refined with rounds of simulated annealing, energy minimization, and individual B-factor refinement with a maximum likelihood target using CNS (327). CNS was used to calculate the initial difference Fourier maps ( $F_o - F_c$ ) and locate the electron density corresponding to each of the *Win* motif peptides, and the structures were built using O (328). Final refinements were carried out using the Phenix.refine module within the Phenix program (147) and the refinement statistics are reported in Table 4.2. Molprobity was used to analyze the protein geometry and steric clashes in final refined structures and the statistics are reported as molprobity score in Table 4.2 (329, 330). All structure figures were generated using PyMOL (331). All structure refinement programs used in this study were a part of SBGrid Consortium.

**Peptide synthesis:** All human SET1 family *Win* motif peptides used in this chapter were synthesized by Genscript (refer to Figure 4.1 for peptide sequences) except the MLL1<sup>H3769Y</sup> peptide, which was obtained from Pi-Proteomics. All peptides were synthesized with an acetyl- and amide-capping group at the N- and C-termini, respectively, to eliminate the contributions of unnatural N- and C-terminal charges on binding. It should be noted that the MLL1<sup>H3769Y</sup> peptide



is insoluble in ITC sample buffer (refer to Chapter 3) and therefore no ITC data could be collected with this peptide. However, the MLL1<sup>H3769Y</sup> peptide is soluble in methyltransferase assay buffer (refer to Chapter 3) and was therefore used to determine inhibition constants described in this chapter.

***MLL1 core complex inhibition assays:***

MALDI-TOF mass spectrometry based methyltransferase assays for the MLL1<sup>H3769Y</sup> and MLL1<sup>S3763A</sup> peptides were performed as described previously (61). The methylation reactions, MALDI-TOF analysis and curve-fitting were performed as described in Chapter 3.

## **RESULTS**

***Structure determination of WDR5 bound to human SET1 family Win motif peptides.***

To begin to understand the structural basis for the observed differences in affinity for WDR5, we determined the three-dimensional X-ray structures of WDR5 bound to the 14-residue MLL2, MLL3, MLL4, SETd1a, and SETd1b *Win* motif peptides with resolutions ranging from 1.2 to 1.9 Å (Table 4.1 and 4.2). In addition, we also determined the structure of WDR5 bound to the 19-residue MLL1 *Win* motif peptide at 1.7 Å resolution for structural comparisons. Crystals grown with the 14-residue MLL1 peptide did not show density for peptide and hence was not included in the comparative analysis. The data collection and refinement statistics for the different *Win* peptide structures are summarized in Tables 4.1 and 4.2.

	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6	+7	+8
MLL1 <sup>3755-3773</sup>	<u>E</u> <sup>3755</sup>	<u>P</u>	<u>P</u>	<u>L</u>	<u>N</u>	<u>P</u>	<u>H</u>	<u>G</u>	<u>S</u>	<u>A</u>	<u>R</u> <sup>3765</sup>	<u>A</u>	<u>E</u>	<u>V</u>	<u>H</u>	<u>L</u> <sup>(a)</sup>	<u>R</u>	<u>K</u>	<u>S</u> <sup>3773</sup>
MLL2 <sup>5333-5346</sup>				<u>I</u> <sup>5333</sup>	<u>N</u>	<u>P</u>	<u>T</u>	<u>G</u>	<u>C</u>	<u>A</u>	<u>R</u> <sup>5340</sup>	<u>S</u>	<u>E</u>	<u>P</u>	<u>K</u>	<u>I</u>	<u>L</u> <sup>5346</sup>		
MLL3 <sup>4703-4716</sup>				<u>V</u> <sup>4703</sup>	<u>N</u>	<u>P</u>	<u>T</u>	<u>G</u>	<u>C</u>	<u>A</u>	<u>R</u> <sup>4710</sup>	<u>S</u>	<u>E</u>	<u>P</u>	<u>K</u>	<u>M</u>	<u>S</u> <sup>4716</sup>		
MLL4 <sup>2504-2517</sup>				<u>L</u> <sup>2504 (a)</sup>	<u>N</u>	<u>P</u>	<u>H</u>	<u>G</u>	<u>A</u>	<u>A</u>	<u>R</u> <sup>2511</sup>	<u>A</u>	<u>E</u>	<u>V</u>	<u>Y</u>	<u>L</u>	<u>R</u> <sup>2517</sup>		
SET1a <sup>1488-1501</sup>				<u>E</u> <sup>1488</sup>	<u>H</u>	<u>Q</u> <sup>(a)</sup>	<u>T</u>	<u>G</u>	<u>S</u>	<u>A</u>	<u>R</u> <sup>1495</sup>	<u>S</u>	<u>E</u>	<u>G</u>	<u>Y</u>	<u>Y</u>	<u>P</u> <sup>1501</sup>		
SET1b <sup>1698-1711</sup>				<u>E</u> <sup>1698</sup>	<u>H</u>	<u>Y</u>	<u>T</u>	<u>G</u>	<u>C</u>	<u>A</u>	<u>R</u> <sup>1705</sup>	<u>S</u>	<u>E</u>	<u>G</u>	<u>F</u>	<u>Y</u>	<u>T</u> <sup>1711</sup>		

\* Amino acid residues underlined are not modeled in the crystal structure. Residues with the superscript (a) are modeled as alanine. Conserved Win motif residues are indicated in red.

**Figure 4.1: Amino acid sequence of the human SET1 family *Win* motif peptides used in crystallization**

	<b>MLL1</b>	<b>MLL2</b>	<b>MLL3</b>	<b>MLL4</b>	<b>SETd1a</b>	<b>SETd1b</b>
Sequence	Acc <sup>L</sup> EPPLPNPHGSARAEVHL RK <sub>NH2</sub>	Acc <sup>L</sup> INPTGCARSEPKI L <sub>NH2</sub>	Acc <sup>S</sup> VNPTGCARSEPKM S <sub>NH2</sub>	Acc <sup>L</sup> LNPHGAARAEVY LR <sub>NH2</sub>	Acc <sup>L</sup> EHQTGSARSEGY YP <sub>NH2</sub>	Acc <sup>L</sup> EHVTGCARSEGFY T <sub>NH2</sub>
Space group	P2 <sub>1</sub>	C121	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C121	C222 <sub>1</sub>	C222 <sub>1</sub>
Cell dimensions a, b, c (Å)	38.606, 80.154, 83.517	139.78, 80.79, 88.12	46.53, 68.25, 87.88	139.03, 80.29, 87.63	78.49, 98.695, 80.280	78.254, 98.272, 79.879
Angles $\alpha, \beta, \gamma$ (°)	90, 89.96, 90	90, 90.2, 90	90, 90, 90	90, 90.04, 90	90, 90, 90	90, 90, 90
Resolution (Å) <sup>a</sup>	50-1.7 (1.76-1.7)	50-1.90 (1.93-1.90)	50-1.22 (1.24-1.22)	50-1.75 (1.78-1.75)	50-1.50 (1.55-1.50)	50-1.81 (1.87-1.81)
Redundancy <sup>a</sup>	6.7 (3.4)	7.5 (6.3)	12.3 (6.3)	7.4 (7.5)	13.6 (9.5)	13.2 (11.2)
Completeness (%) <sup>a</sup>	99.4 (93.8)	100 (99.4)	99.6 (96.9)	99.8 (100)	99.7 (98.9)	98.3 (89.8)
I/ $\sigma$ <sup>a</sup>	28.5 (2.61)	34.76 (4.42)	58.9 (3.17)	44.87 (6.07)	37.1 (3.25)	60.9 (12.06)
R <sub>merge</sub> <sup>a</sup>	8.2 (47.8)	8.5 (54.3)	13.2 (58.9)	7.9 (54.6)	6.8 (52.1)	12.1 (35.7)

<sup>a</sup> Values in parentheses are for the highest resolution shell.

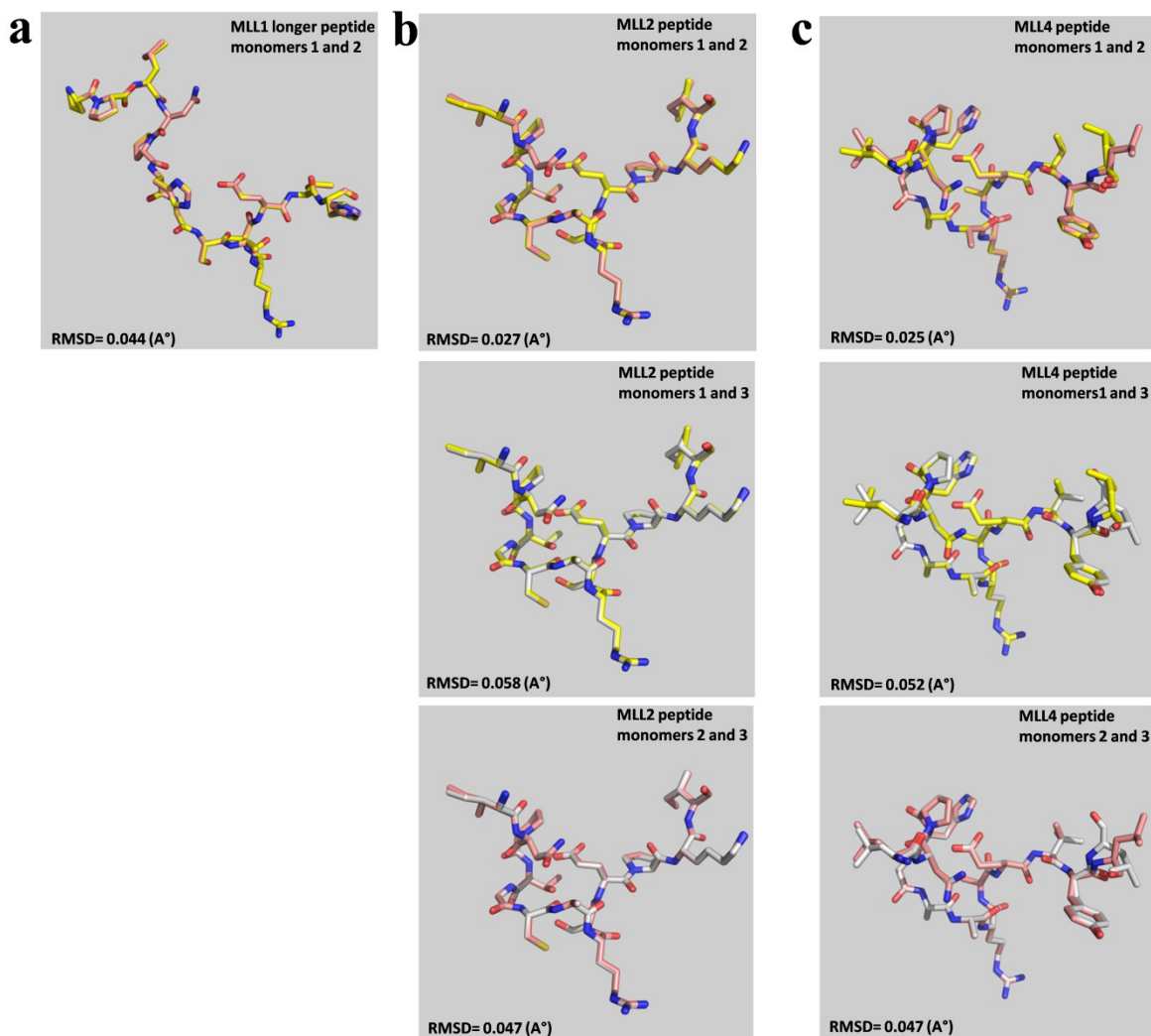
Table 4.1: Summary of the X-ray data collection statistics for the different human SET1 family *Win* motif peptides

	MLL1	MLL2	MLL3	MLL4	SETd1a	SETd1b
Number of molecules in the asymmetric unit	2	3	1	3	1	1
Resolution (Å)	1.75	1.91	1.3	1.75	1.75	1.82
No.reflections	51223	76235	69347	96631	31743	27254
R <sub>work</sub> / R <sub>free</sub>	15.89/20.06	16.53/19.97	18.66/20.91	18.59/21.61	15.68/18.95	21.44/24.21
No.atoms						
Protein	4713	7070	2356	7070	2356	2356
Peptide	236	285	100	264	92	45
Water	679	705	560	559	338	116
B-factors						
Protein	15.01	23.35	13.07	21.08	13.01	24.16
Peptide	20.09	26.16	14.92	25.58	21.37	27.26
Water	26.77	32.45	25.88	27.27	24.67	28.63
root mean square (r.m.s) deviations						
Bond lengths (Å)	0.007	0.008	0.006	0.007	0.007	0.007
Bond angles (°)	1.148	1.137	1.178	1.174	1.165	1.154
Ramachandran plot						
% most favored	88.90	87.20	88.30	87.70	90.30	93.40
% allowed	11.10	12.60	11.70	12.30	9.70	6.60
% outliers	0	0.20	0	0.53	0.96	0.33
Molprobability score	1.36	1.28	1.39	1.17	1.34	1.21

Table 4.2: Summary of the refinement statistics for the different human SET1 family *Win* motif peptides

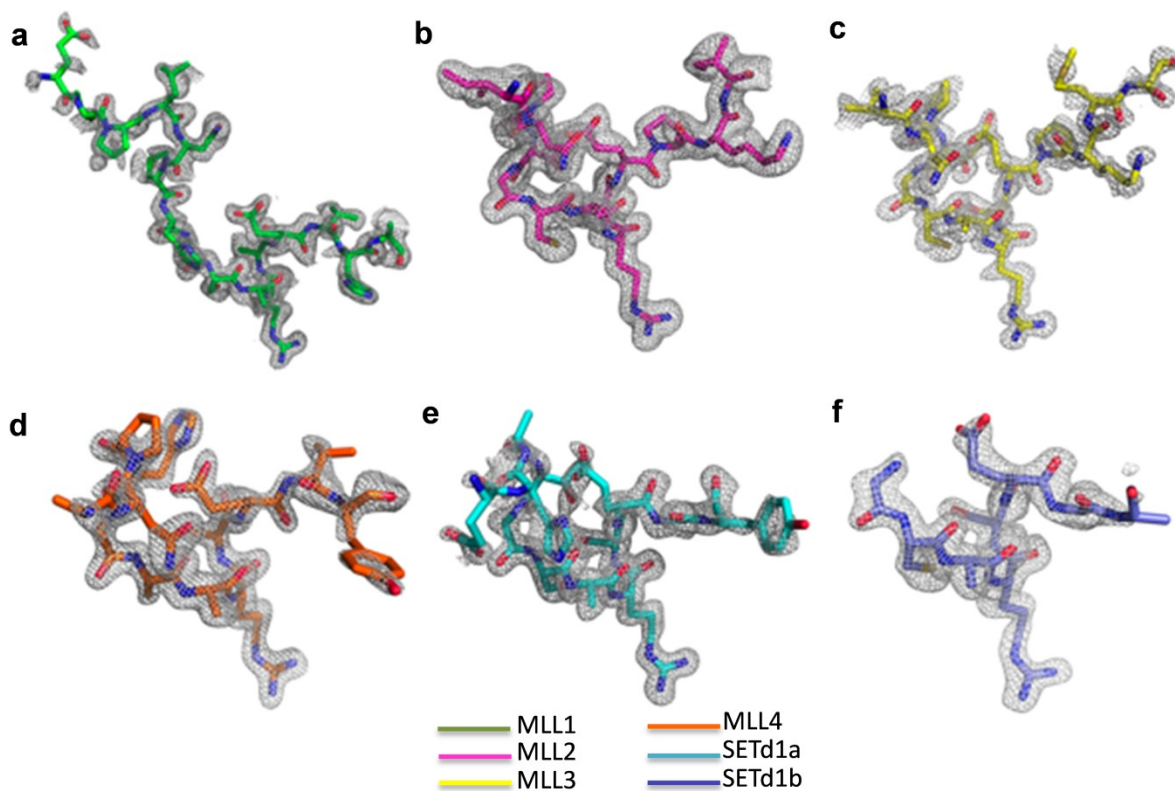
### ***Crystal structure of MLL1 Win motif peptide bound to WDR5***

We determined the crystal structure of WDR5 bound to a MLL1 *Win* motif peptide encompassing residues 3755-3773. This 19-residue MLL1 *Win* motif peptide co-crystallized with WDR5 in the  $P2_1$  space group and the structure was determined at 1.70 Å resolution. The molecular replacement solution revealed that there are two molecules in the asymmetric unit. The structural superposition of the two individual monomers in the asymmetric unit is reported in Figure 4.2a and the two peptide molecules superpose with a root mean square deviation (RMSD) of 0.044 Å. The structure was refined using the Phenix program and the final refinement statistics include a  $R_{\text{work}}$  of 15.89% and  $R_{\text{free}}$  of 20.06 % (Table 4.2). Fifteen residues could be modeled into the electron density with the last four residues on the C-terminus being disordered (Figure 4.3a). The structure was analyzed by Procheck (332) for stereochemical quality and the Ramachandran plot did not show any deviations from the allowed conformations for Phi ( $\phi$ ), Psi ( $\psi$ ) angles (refer to appendix Figure 5a). The structure coordinates have been deposited in the protein data bank (PDB code: 4ESG).



**Figure 4.2** Pair-wise superposition of peptide monomers in the asymmetric unit for MLL1, MLL2 and MLL4 *Win* motif structures

**a-c.** Structural superposition of individual peptide monomers in the asymmetric unit of MLL1, MLL2 and MLL4 *Win* motif structures. In a, superposition of the two peptide chains (1 and 2) from the MLL1 *Win* motif structure is shown. In b, individual pair-wise superposition of the three peptide chains (1, 2 and 3) of MLL2 *Win* motif peptide structure is shown. In c, individual pair-wise superposition of the three peptide chains (1, 2 and 3) of MLL4 *Win* motif peptide structure is shown. Root mean square deviation (RMSD Å) values for the superposition of each monomer pairs are indicated. Structural differences are present in the ends of each peptide monomer chain. Structural figures were generated using PyMOL (331).



**Figure 4.3 Simulated Annealing omit maps showing the different human SET1 family *Win* motif peptides bound to WDR5**

**a-f.** Simulated annealing Fo-Fc Omit maps showing the different *Win* motif peptides bound to WDR5. MLL1 (c) and MLL3 (e) *Win* motif peptides are contoured at  $2\sigma$  and MLL2 (d), MLL4 (f), SETd1a (g), and SETd1b (h) *Win* peptides are contoured at  $3\sigma$ . The peptides are shown as sticks and color-coded as MLL1 (green), MLL2 (pink), MLL3 (yellow), MLL4 (orange), SETd1a (cyan), and SETd1b (dark blue).

### ***Crystal structure of MLL2 Win motif peptide bound to WDR5***

Next, we determined the crystal structure of MLL2 *Win* motif peptide (encompassing residues 5333-5346) bound to WDR5. The 14-residue MLL2 *Win* motif peptide co-crystallized with WDR5 in the  $C_{121}$  space group and the structure was determined at 1.90 Å resolution. The molecular replacement solution revealed three molecules in the asymmetric unit. The structural superposition of the three individual monomers in the asymmetric unit is reported in Figure 4.2b and the three peptide molecules superpose with RMSD values that range from 0.027-0.058 Å. The structure was refined using the Phenix program and the final refinement statistics include a  $R_{\text{work}}$  of 16.53% and  $R_{\text{free}}$  of 19.97 % (Table 4.2). Thirteen residues could be modeled into the electron density with the last leucine residue on the C-terminus being disordered (Figure 4.3b). The structure was analyzed by Procheck (332) for stereochemical quality and the Ramachandran plot did not show any deviations from the allowed conformations for Phi ( $\phi$ ), Psi ( $\psi$ ) angles except for the WDR5 residue D212 from chain A, which is present in a disallowed region (refer to appendix Figure 5b). The structure coordinates have been deposited in the protein data bank (PDB code: 4ERQ).

### ***Crystal structure of MLL3 Win motif peptide bound to WDR5***

Next, we determined the crystal structure of WDR5 in complex with a MLL3 *Win* motif peptide encompassing residues 4703-4716. The 14-residue MLL3 *Win* motif peptide co-crystallized with WDR5 in the  $P2_12_12_1$  space group and the structure was determined at 1.22- Å resolution. The molecular replacement solution revealed only one molecule in the asymmetric unit. The structure was refined using the Phenix program and the final refinement statistics include a  $R_{\text{work}}$  of 18.66% and  $R_{\text{free}}$  of 20.91 % (Table 4.2). All fourteen residues could be modeled into the



electron density owing to the high quality diffraction data (Figure 4.3c). The structure was analyzed by Procheck (332) for stereochemical quality and the Ramachandran plot did not show any deviations from the allowed conformations for Phi ( $\phi$ ), Psi ( $\psi$ ) angles (refer to appendix Figure 5c). The structure coordinates have been deposited in the protein data bank (PDB code: 4ERY).

#### ***Crystal structure of MLL4 Win motif peptide bound to WDR5***

The crystal structure of WDR5-MLL4 *Win* motif peptide (residues 2504-2517) binary complex was determined at 1.75 Å resolution. The 14-residue MLL4 *Win* motif peptide co-crystallized with WDR5 in the  $C_{121}$  space group and the molecular replacement solution revealed three molecules in the asymmetric unit. The structural superposition of the three individual monomers in the asymmetric unit is reported in Figure 4.2c and the three peptide molecules superpose with RMSD values that range from 0.025-0.052 Å. The structure was refined using the Phenix program and the final refinement statistics include a  $R_{\text{work}}$  of 18.59% and  $R_{\text{free}}$  of 21.61% (Table 4.2). Eleven residues could be unambiguously modeled into the electron density with the last leucine residue on the N-terminus and the last two residues on the C-terminus being disordered (Figure 4.3d). The structure was analyzed by Procheck (332) for stereochemical quality and the Ramachandran plot did not show any deviations from the allowed conformations for Phi ( $\phi$ ), Psi ( $\psi$ ) angles (refer to appendix Figure 5d). The structure coordinates have been deposited in the protein data bank (PDB code: 4ERZ).

***Crystal structure of SETd1a Win motif peptide bound to WDR5***

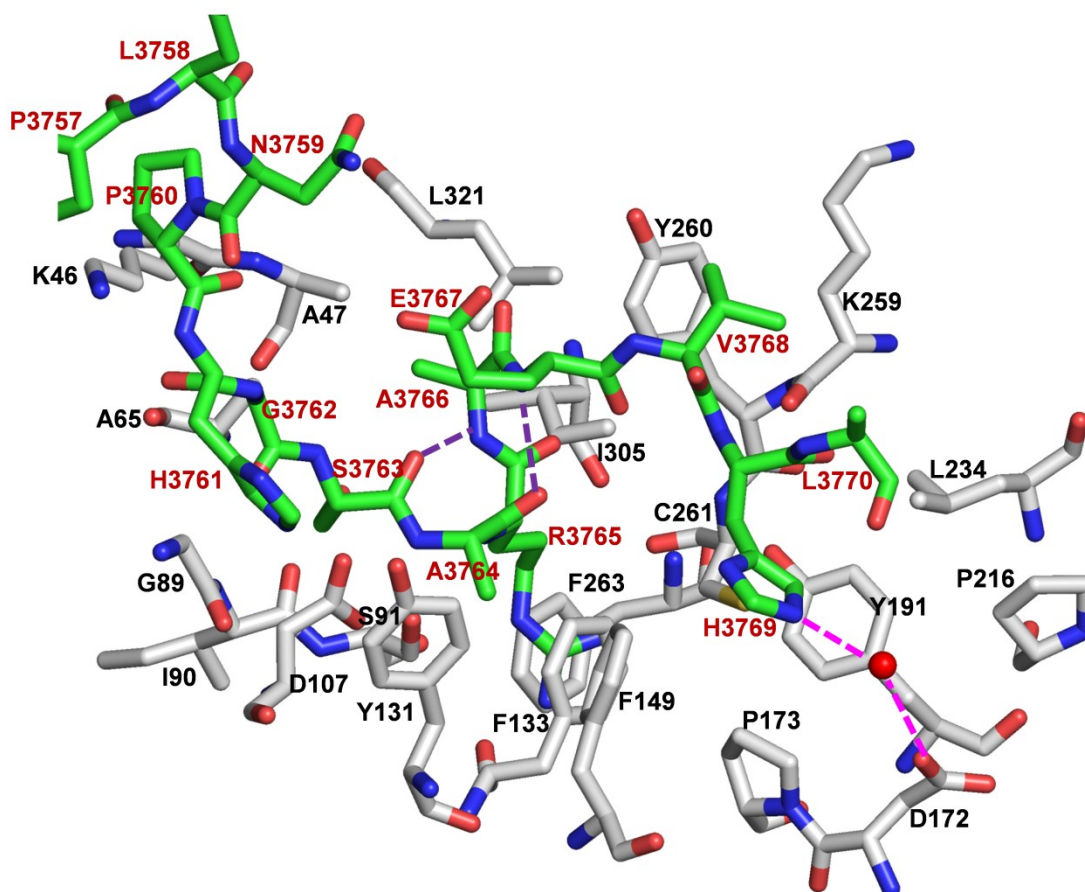
The crystal structure of WDR5-SETd1a *Win* motif peptide (residues 1488-1501) binary complex was determined at 1.50 Å resolution. The 14-residue SETd1a *Win* motif peptide co-crystallized with WDR5 in the C222<sub>1</sub> space group and the molecular replacement solution revealed only one molecule in the asymmetric unit. The structure was refined using the Phenix program and the final refinement statistics include a R<sub>work</sub> of 15.68% and R<sub>free</sub> of 18.95% (Table 4.2). Nine residues could be unambiguously modeled into the electron density with the last three residues on the N-terminus and the last two residues on the C-terminus being disordered (Figure 4.3e). The structure was analyzed by Procheck (332) for stereochemical quality and the Ramachandran plot did not show any deviations from the allowed conformations for Phi (φ), Psi (ψ) angles (refer to appendix Figure 5e). The structure coordinates have been deposited in the protein data bank (PDB code: 4EWR).

***Crystal structure of SETd1b Win motif peptide bound to WDR5***

The crystal structure of WDR5-SETd1a *Win* motif peptide (residues 1488-1501) binary complex was determined at 1.81 Å resolution. The 14-residue SETd1b *Win* motif peptide co-crystallized with WDR5 in the C222<sub>1</sub> space group and the molecular replacement solution revealed only one molecule in the asymmetric unit. The structure was refined using the Phenix program and the final refinement statistics include a R<sub>work</sub> of 21.44% and R<sub>free</sub> of 24.21% (Table 4.2). Seven residues could be unambiguously modeled into the electron density with the last four residues on the N-terminus and the last three residues on the C-terminus being disordered (Figure 4.3f). The structure was analyzed by Procheck (332) for stereochemical quality and the Ramachandran plot did not show any deviations from the allowed conformations for Phi (φ), Psi (ψ) angles (refer to

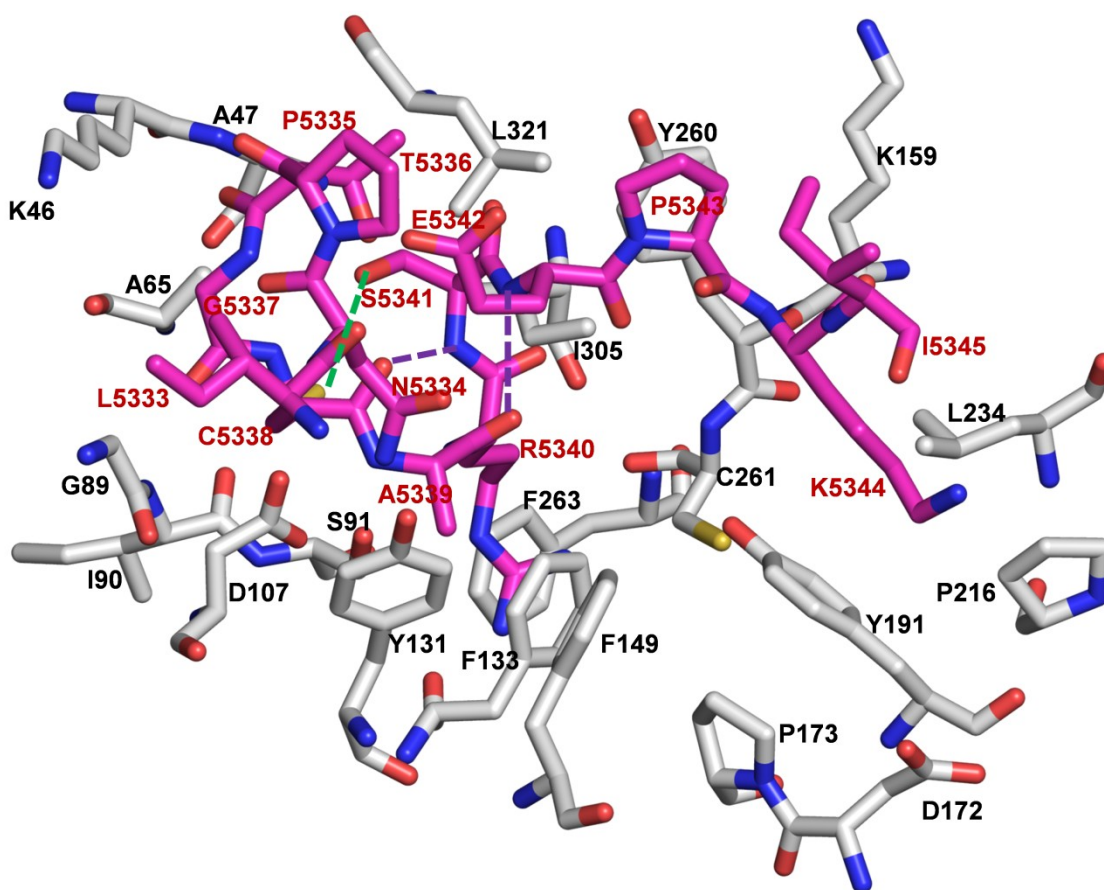
appendix Figure 5f). The structure coordinates have been deposited in the protein data bank (PDB code: 4ESO).

An overview of the structure of each of the *Win* motif peptides bound to WDR5 is presented in Figures 4.4a-f. The individual structures were analyzed for the presence of intra- and intermolecular hydrogen bond, van der Waals contacts between the *Win* motif peptide and WDR5, which are summarized in Figures 4.5-4.7. The contribution of solvent molecules is ignored in this structural analysis. The structures of other human SET1 family *Win* motif peptides are then compared and contrasted with the three-dimensional structure of the 19-residue MLL1 *Win* motif peptide and the similarities and differences observed in these structures are described in the following sections (Figures 4.8-4.12).



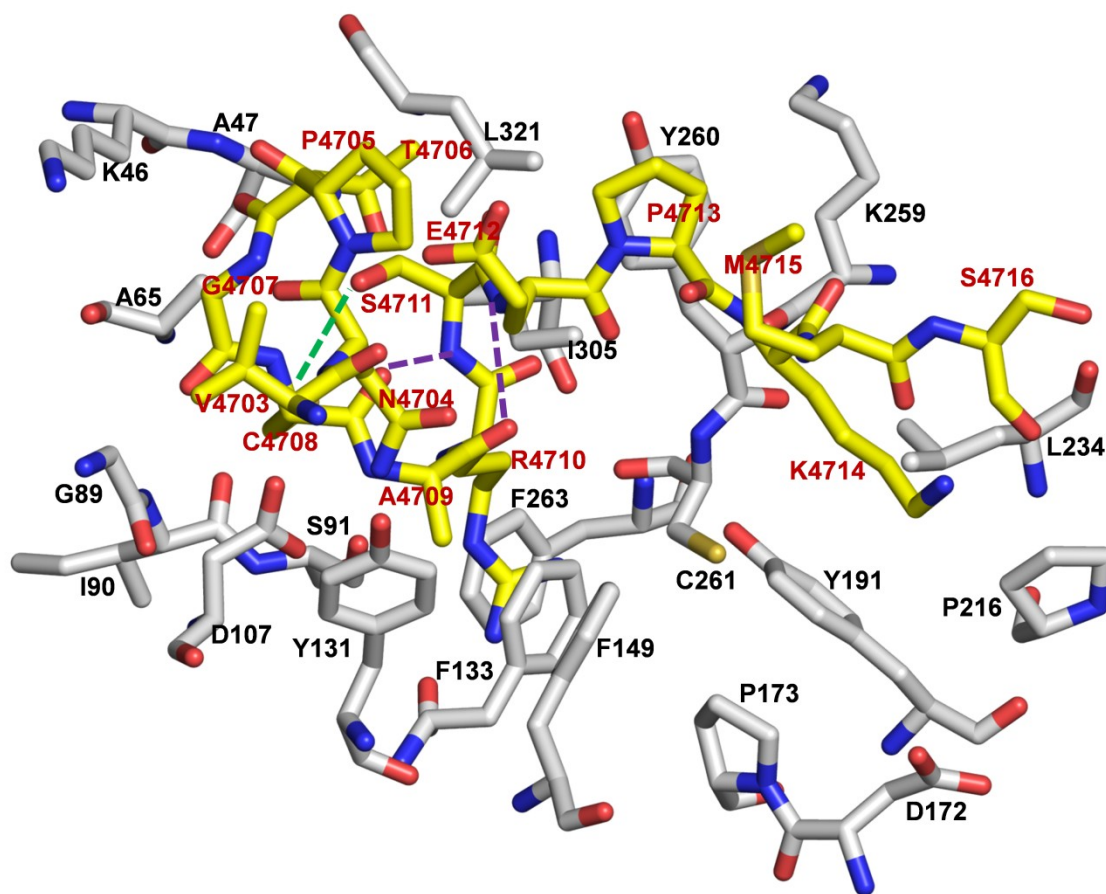
**Figure 4.4a: Structure of MLL1 *Win* motif peptide bound to WDR5**

The amino acid residues that form the interaction surface between WDR5 and MLL1 *Win* motif peptide are shown. WDR5 residues are shown as white sticks and the MLL1 *Win* motif peptide residues are shown as sticks in green. Intramolecular  $i$  to  $i+3$  (purple dotted lines) and intermolecular (pink dotted lines) hydrogen bonds present within the MLL1 *Win* peptide structure are shown. Water molecule is shown as red sphere. WDR5 residues are labeled in black and MLL1 *Win* motif peptide residues are labeled in red.



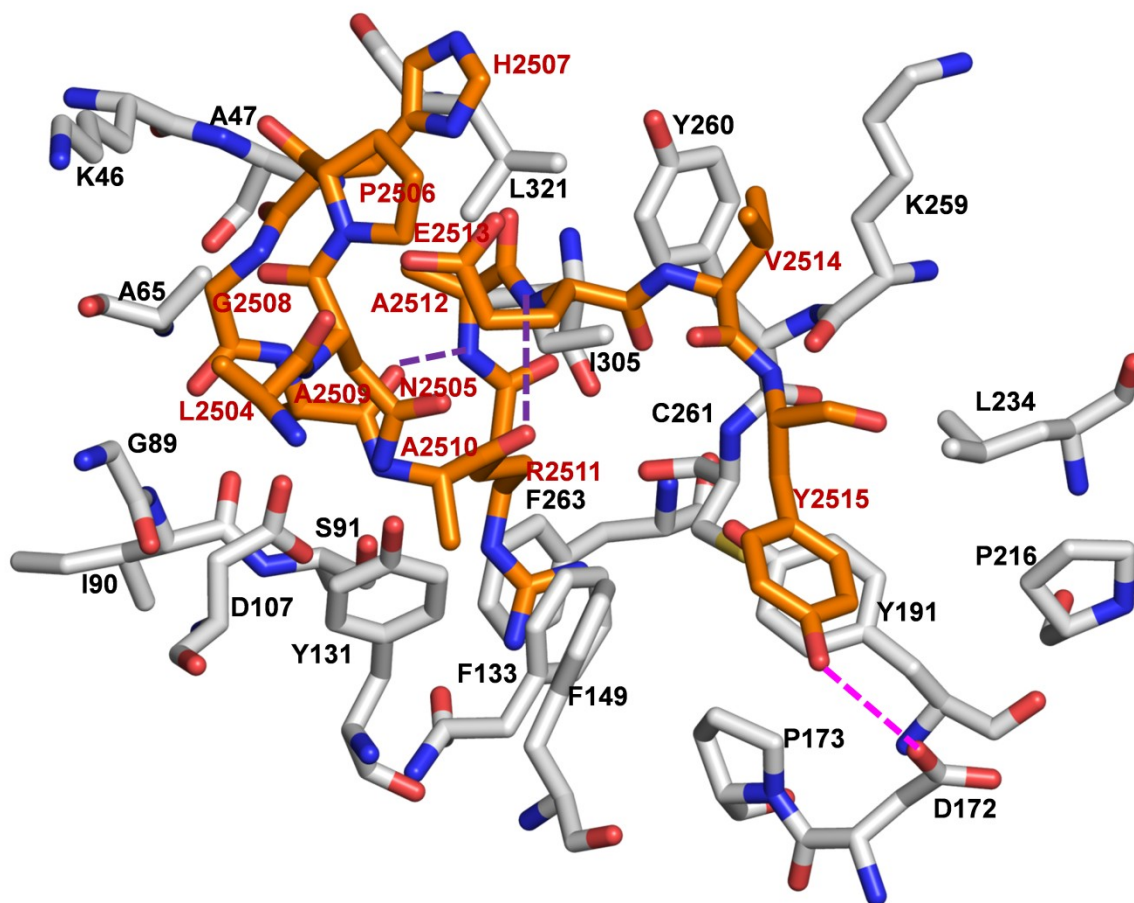
**Figure 4.4b: Structure of MLL2 *Win* motif peptide bound to WDR5**

The amino acid residues that form the interaction surface between WDR5 and MLL2 *Win* motif peptide are shown. WDR5 residues are shown as white sticks and the MLL2 *Win* motif peptide residues are shown as sticks in pink. Intramolecular hydrogen bonds are shown as purple dotted lines (i to i+3) and green dotted lines, respectively. WDR5 residues are labeled in black and MLL2 *Win* motif peptide residues are labeled in red.



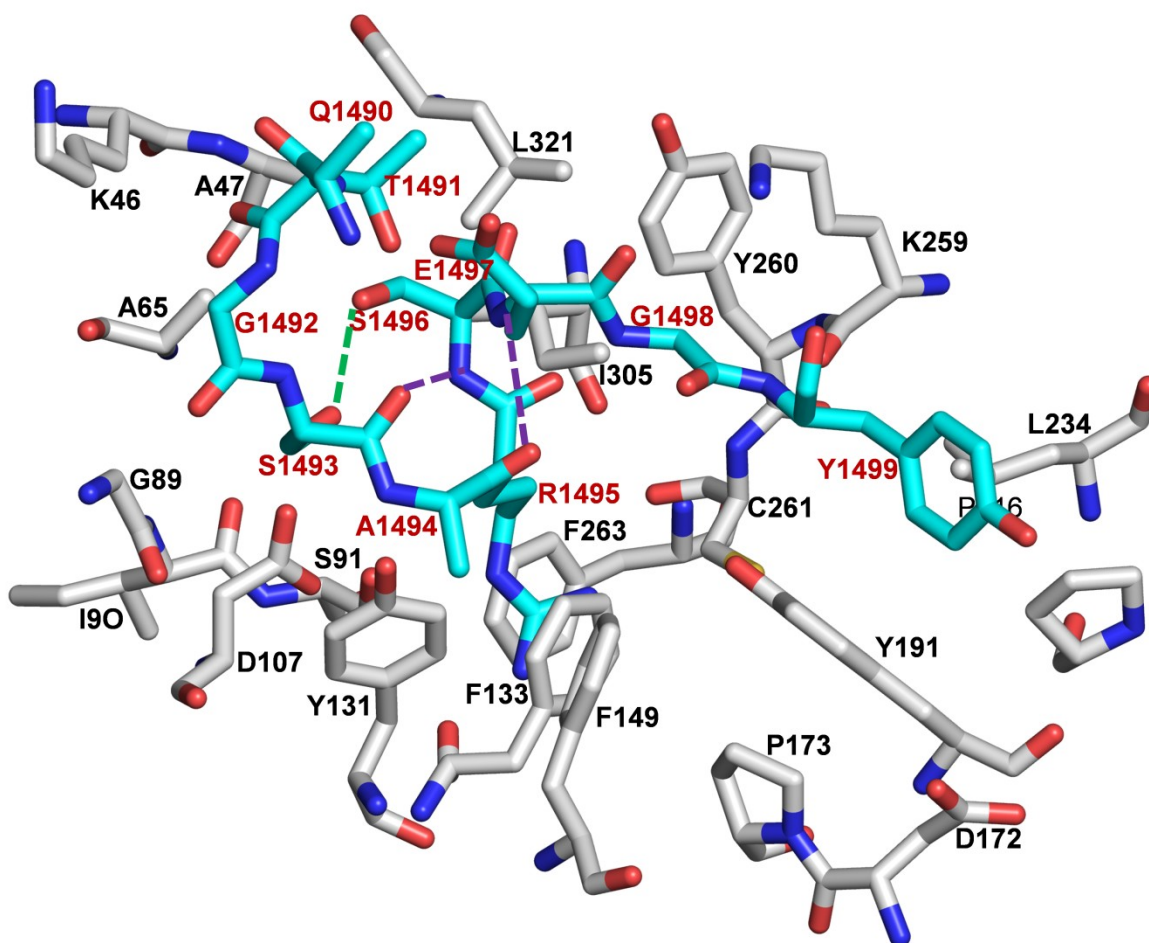
**Figure 4.4c: Structure of MLL3 *Win* motif peptide bound to WDR5**

The amino acid residues that form the interaction surface between WDR5 and MLL3 *Win* motif peptide are shown. WDR5 residues are shown as white sticks and the MLL3 *Win* motif peptide residues are shown as sticks in yellow. Intramolecular hydrogen bonds are shown as purple dotted lines ( $i$  to  $i+3$ ) and green dotted lines, respectively. WDR5 residues are labeled in black and MLL3 *Win* motif peptide residues are labeled in red.



**Figure 4.4d: Structure of MLL4 *Win* motif peptide bound to WDR5**

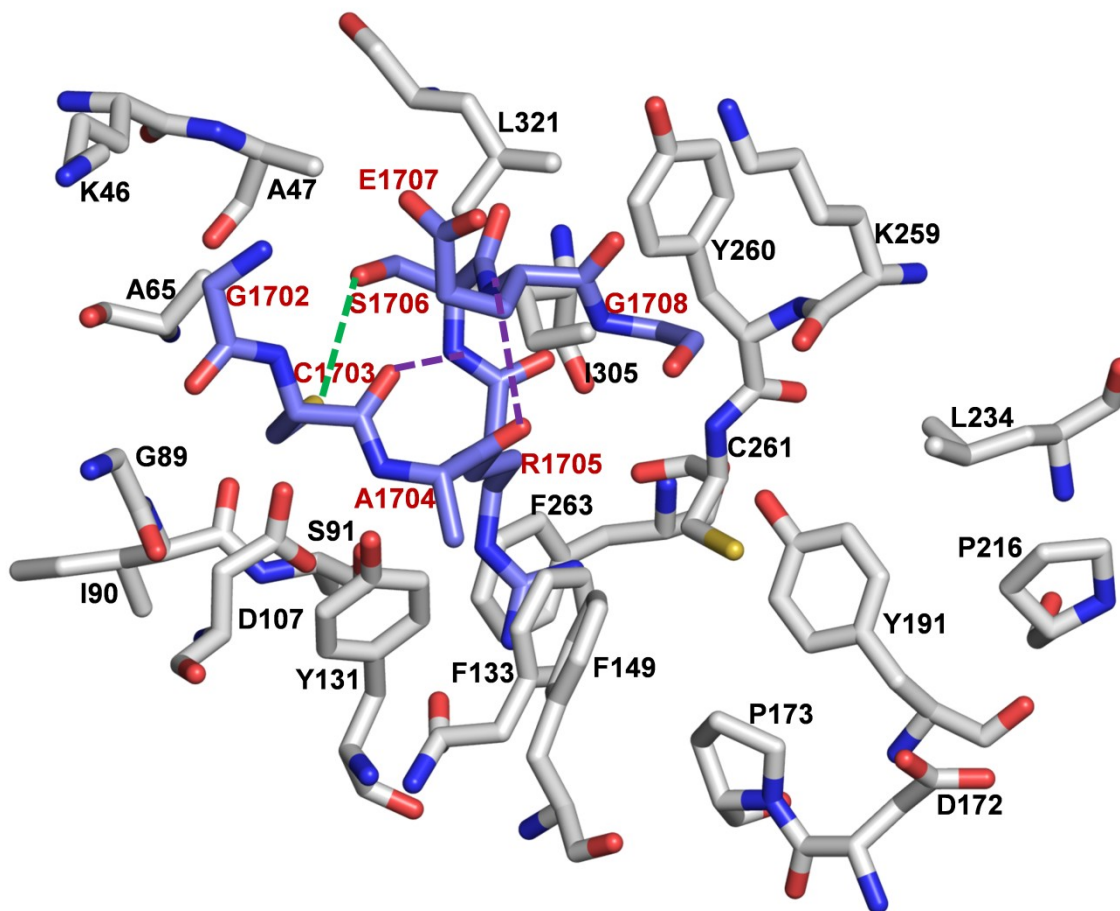
The amino acid residues that form the interaction surface between WDR5 and MLL4 *Win* motif peptide are shown. WDR5 residues are shown as white sticks and the MLL4 *Win* motif peptide residues are shown as sticks in orange. Intramolecular hydrogen bonds are shown as purple dotted lines (i to i+3) and pink dotted lines, respectively. WDR5 residues are labeled in black and MLL4 *Win* motif peptide residues are labeled in red.



**Figure 4.4e: Structure of SETd1a *Win* motif peptide bound to WDR5**

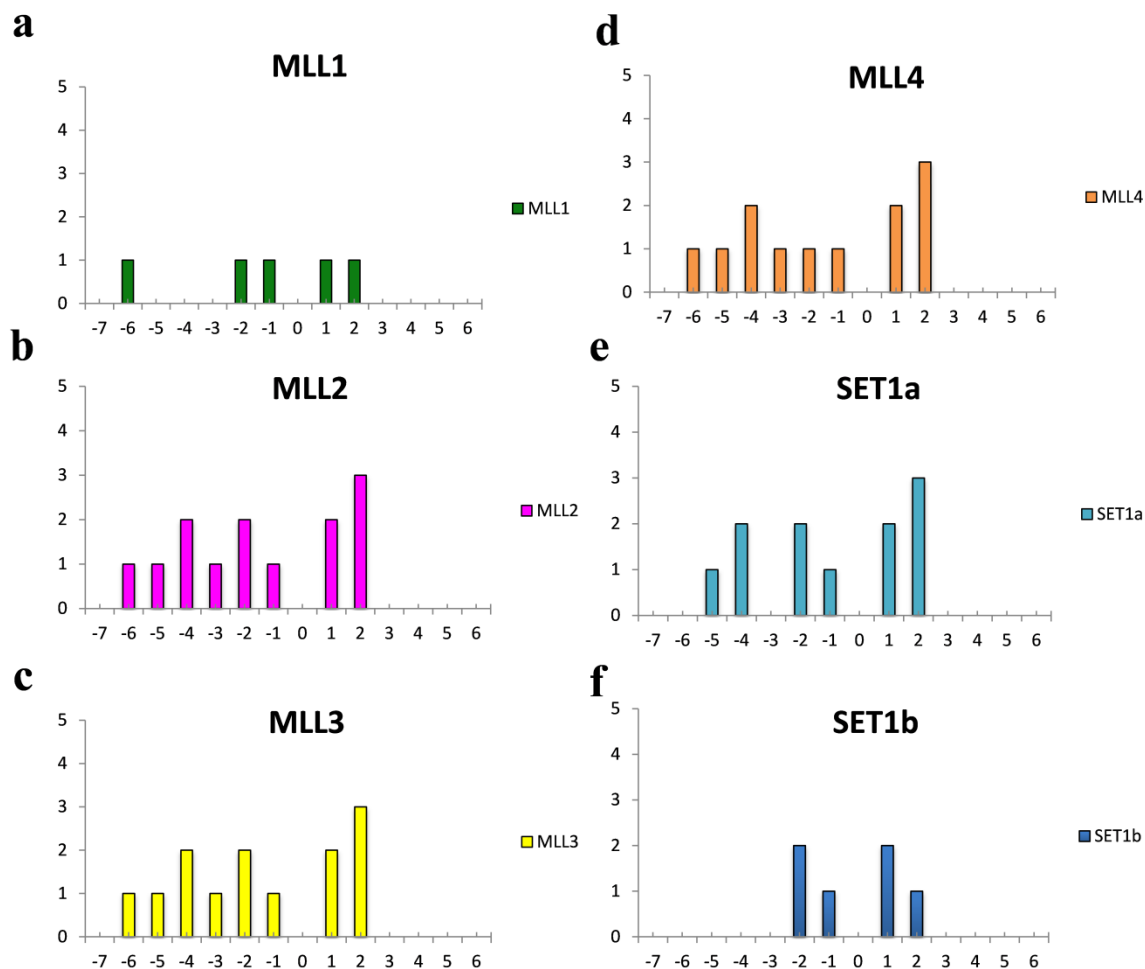
The amino acid residues that form the interaction surface between WDR5 and SETd1a *Win* motif peptide are shown. WDR5 residues are shown as white sticks and the SETd1a *Win* motif peptide residues are shown as sticks in cyan. Intramolecular hydrogen bonds are shown as purple dotted lines (i to i+3) and green dotted lines, respectively. WDR5 residues are labeled in black and SETd1a *Win* motif peptide residues are labeled in red.





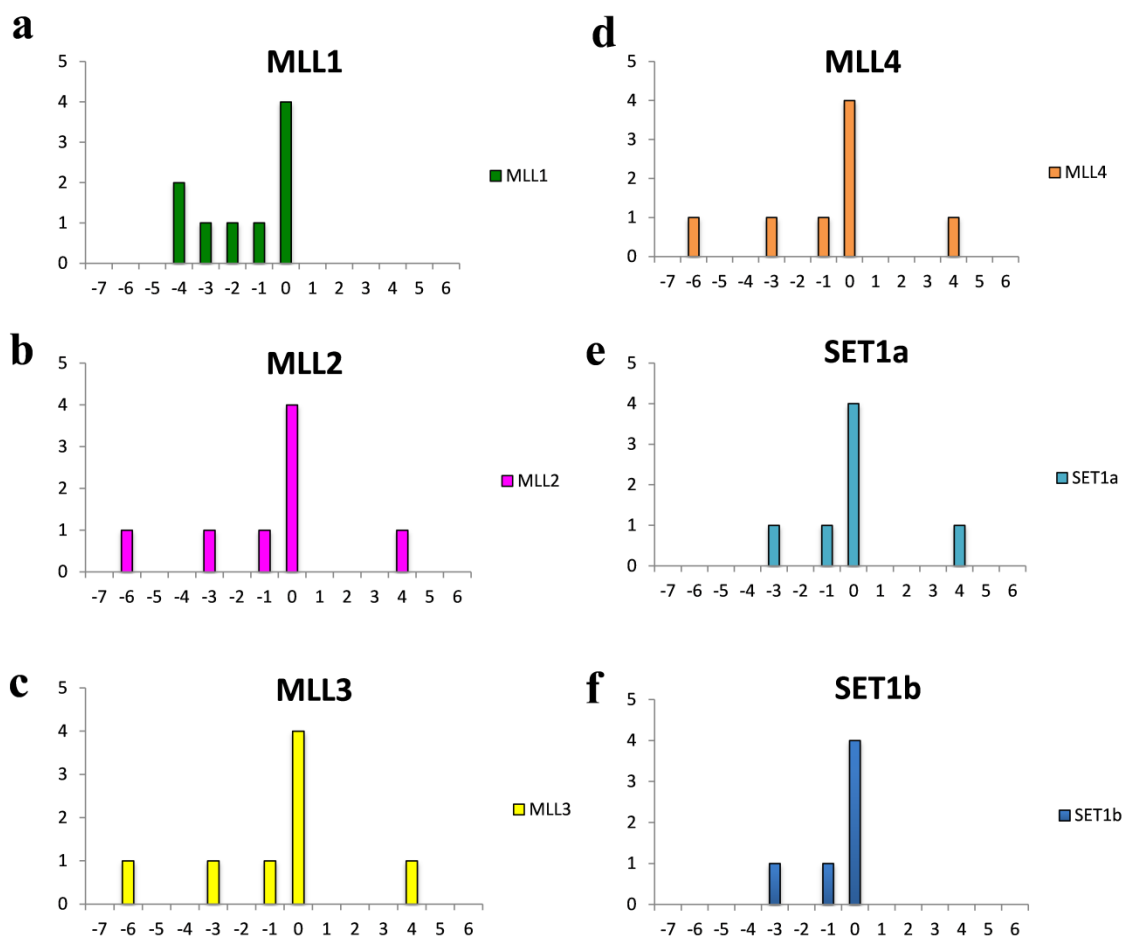
**Figure 4.4f: Structure of SETd1b *Win* motif peptide bound to WDR5**

The amino acid residues that form the interaction surface between WDR5 and SETd1b *Win* motif peptide are shown. WDR5 residues are shown as white sticks and the SETd1b *Win* motif peptide residues are shown as sticks in dark blue. Intramolecular hydrogen bonds are shown as purple dotted lines (i to i+3) and green dotted lines, respectively. WDR5 residues are labeled in black and SETd1b *Win* motif peptide residues are labeled in red.



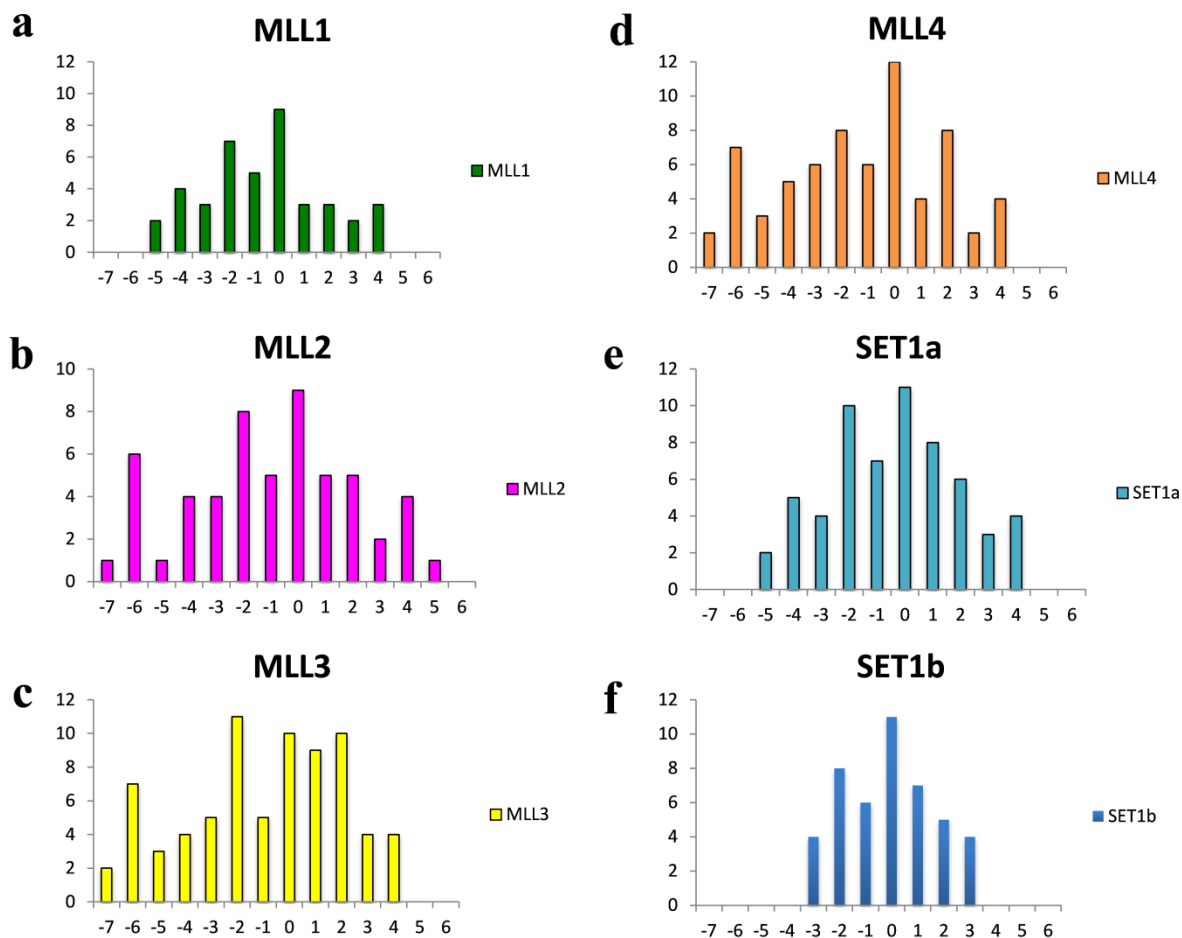
**Figure 4.5: Intramolecular hydrogen bonds within human SET1 family *Win* motif peptides**

**a-f.** Summary of number of intramolecular hydrogen bonds present in human SET1 family *Win* motif peptide structures. The hydrogen bonds are listed according to the residue position of the peptide with the conserved *Win* motif arginine residue designated as position 0. CCP4i suite was used to analyze the intramolecular hydrogen bond contacts present within the different *Win* motif structures using a cutoff distance range of 2.50-3.25 Å. Additional intramolecular hydrogen bonds are observed in the N-terminus of MLL2, MLL3 and MLL4 *Win* motif structures as compared to MLL1, SETd1a and SETd1b structures. Refer to text to more detail.



**Figure 4.6: Intermolecular hydrogen bonds between WDR5 and different SET1 family *Win* motif peptides**

**a-f.** Summary of number of intermolecular hydrogen bonds present in human SET1 family *Win* motif peptide structures. The hydrogen bonds are listed according to the residue position of the peptide with the conserved *Win* motif arginine residue designated as position 0. CCP4i suite was used to analyze the intermolecular hydrogen bond contacts present between WDR5 and different *Win* motif peptides using a cutoff distance range of 2.50-3.25 Å. All *Win* motif peptides participate in similar number of intermolecular hydrogen bonds. Refer to text for more detail.



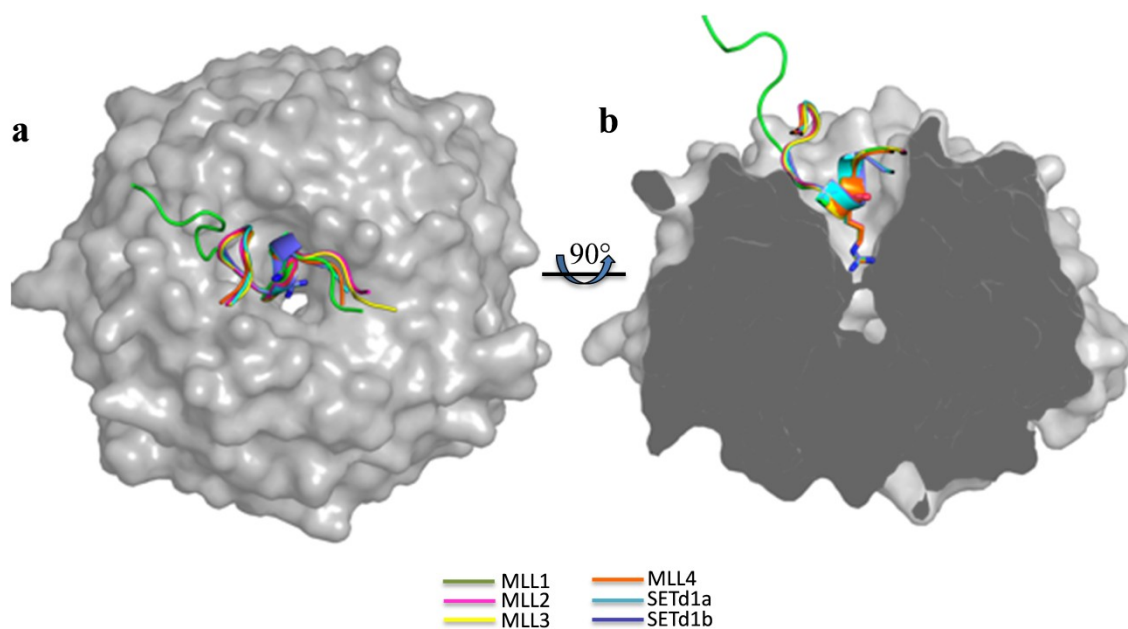
**Figure 4.7: Intra and intermolecular van der Waals contacts present in different human SET1 family *Win* motif peptides**

**a-f.** Summary of number of inter and intramolecular van der Waals contacts present in human SET1 family *Win* motif peptide structures. The van der Waals contacts are listed according to the residue position of the peptide with the conserved *Win* motif arginine residue designated as position 0. CCP4i suite was used to analyze inter- and intramolecular van der Waals contacts using a cutoff distance range of 3.3-4.0 Å. Refer to text to more detail.

## **STRUCTURAL ANALYSIS OF WDR5-HUMAN SET1 FAMILY *WIN* MOTIF PEPTIDE COMPLEXES**

### ***Human SET1 family Win motif peptides bind WDR5 using a similar conformation***

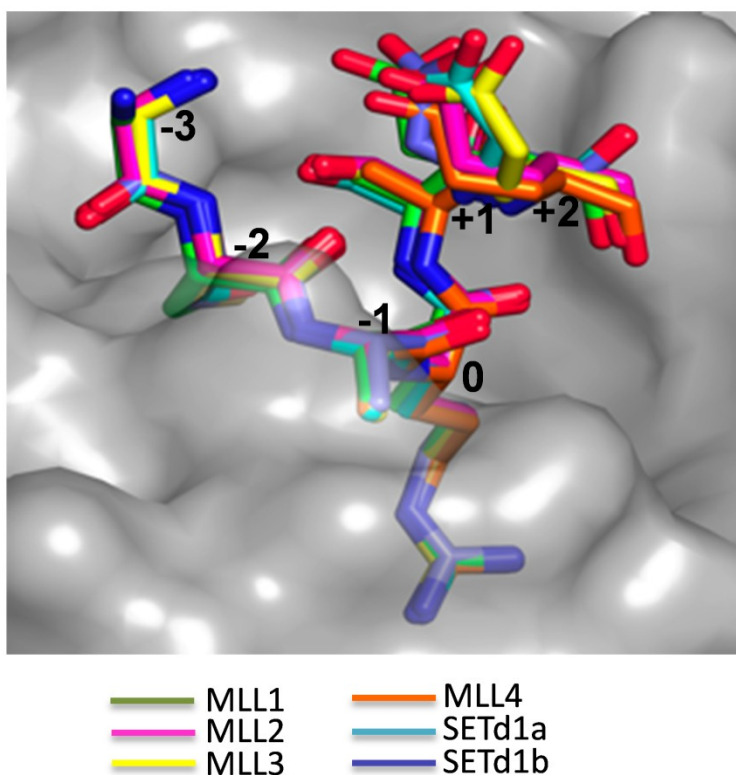
The overall structure of WDR5 in each co-crystal structure is topologically similar to that of previously reported structures of WDR5 bound to MLL1 *Win* motif peptides, or histone H3 peptides (60, 64-68). The WDR5 structure consists of a seven-blade  $\beta$ -propeller with a central cavity that traverses the center of the protein from top to bottom (Figure 4.8a and b). Initial electron density difference maps revealed that each *Win* motif peptide binds in the central opening at the top of WDR5, with the conserved arginine residue in each structure (designated as position 0) inserted into the central tunnel (Figure 4.8a and b).



**Figure 4.8. Human SET1 family *Win* motif peptides bind WDR5 using a similar  $3_{10}$  helical conformation**

Human SET1 family *Win* motif peptides bind WDR5 using the same arginine-binding pocket. (a-b) WDR5 molecule is shown as a surface representation (grey) and the different *Win* motif peptides are shown in different colors. a. Superposition of *Win* motif peptides shown with a cartoon representation. b. Cut-away view of the arginine-binding pocket of WDR5 is shown. All *Win* motif peptides insert the conserved arginine into the central tunnel in WDR5. The side chain of the conserved arginine is shown as a stick representation.

The six conserved residues of the *Win* motif (which ranges from amino acid -3 to +2, relative to the conserved arginine) are ordered in each co-crystal structure and are all highly similar (Figures 4.3 and 4.9). Superposition of all atoms in the conserved six-residue sequence in all peptides reveals an RMS deviation of 0.207 Å (Figure 4.9). All SET1 family *Win* motif peptides bind WDR5 by adopting a  $3_{10}$ -helical conformation (Figure 4.8) and an extensive network of direct and water-mediated hydrogen bonds and hydrophobic interactions stabilize the binding of all SET1 family *Win* motif peptides (Figures 4.5-4.7). The conserved *Win* motif residues (-3 to +2) all participate in the formation of the  $3_{10}$ -helix in each of the crystal structure and are stabilized by two intramolecular  $i$  to  $i+3$  main-chain hydrogen bonds between residues at -2 and +1 positions and between residues at -1 and +2 positions (Figures 4.4a-f).



**Figure 4.9 Conserved *Win* motif residues adopt structurally similar conformations**

Conserved *Win* motif residues adopt structurally similar conformations. a. Model superposition of the six conserved *Win* motif residues from different WDR5-*Win* motif peptide co-crystal structures. Conserved *Win* motif residues are indicated and numbered with respect to the conserved arginine (position 0). Refer to figure 4.1 for residue numbering. WDR5 molecule is shown as a surface representation (grey) and the different *Win* motif peptides are shown as sticks (color coded as in Figure 4.8).



Conserved *Win* motif residues Gly(-3), Ala(-1) and Arg (0) occupy similar positions in all the *Win* motif structures (Figure 4.9) and involve in several interactions with WDR5 that are identical (Figures 4.5-4.7). Similar to the previously determined structure of the 12-residue MLL1 *Win* motif peptide (60), the conserved Arg at position 0 in all the SET1 family *Win* motif structures makes direct and water-mediated hydrogen bonds with WDR5 residues: Ser91, Phe133, Ser175 and Cys261 (Figures 4.4a-f and figures 4.5-4.7). Additionally, the arginine side chain makes intermolecular van der Waals contacts with Ser49, Ser91, Asp92, Phe133, Tyr260, Cys261, Phe263 and Ile305 (Figures 4.4a-f and figures 4.5-4.7). This extensive network of interactions observed with the *Win* motif arginine explains its high sequence conservation within the different SET1 family *Win* motif peptides and provides a structural basis for why the mutation of the conserved arginine (R3765A) in MLL1 completely abolishes the interaction with WDR5 both in vitro (61) and in mammalian cells (Figure 3.1). In all the *Win* motif peptide structures, other conserved *Win* motif residues Glycine at -3 and Alanine at the -1 position participate in multiple hydrogen bonds (Gly89 and Asp107) and van der Waals contacts (Ala65, Gly89, Ser91, Asp107, Tyr131, Phe133 and Phe149) with WDR5 residues that are common to all *Win* motif structures (Figures 4.4a-f and figures 4.5-4.7). It is likely that these conserved sets of interactions observed within the conserved *Win* motif accounts for the strong interaction affinities of all human SET1 family *Win* motif peptides to WDR5 and explains their high amino acid sequence conservation within the SET1 family members (refer to Figure 1.2 and appendix Figures 1.1-1.6).

However, comparison of the different SET1 family *Win* motif peptide structures also reveal minor differences within the conserved *Win* motif residues. The largest difference is

observed at residues present in the -2 and +1 position. In three out of six structures (MLL2, MLL3, and SETd1b), the cysteine side chain at the -2 position forms an intramolecular hydrogen bond with the hydroxyl group of serine residue present at +1 position (Figures 4.4b, 4.4c and 4.4f). Likewise, the SETd1a *Win* motif peptide has a serine at both the -2 and +1 positions and can form a similar intramolecular side chain hydrogen bond (Figure 4.4e). In contrast, MLL1 and MLL4 *Win* motifs have a serine or an alanine at the -2 position and an alanine at the +1 position and do not have the ability to form the intramolecular side chain hydrogen bond (Figure 4.4a, and 4.4d). It is possible that this extra hydrogen bond present in the MLL2, MLL3, SETd1a, and SETd1b *Win* motif peptides contribute to their 5- to 50-fold increased affinities for WDR5 as compared to the MLL1 *Win* motif peptide. Interestingly, the MLL4 *Win* motif peptide that has an alanine at the -2 and +1 positions (hence no additional hydrogen bond) can still bind WDR5 with 30-fold greater affinity compared to MLL1. One possible explanation for this difference could be that an alanine side chain in MLL4 (-2 position) makes favorable hydrophobic interactions in the relatively hydrophobic pocket in WDR5 (formed by Ala 65, Ile90 and the alanine at -1 position of *Win* motif peptides) as compared to a hydrophilic side chain of serine present in MLL1 (Figures 4.4a and 4.4e).

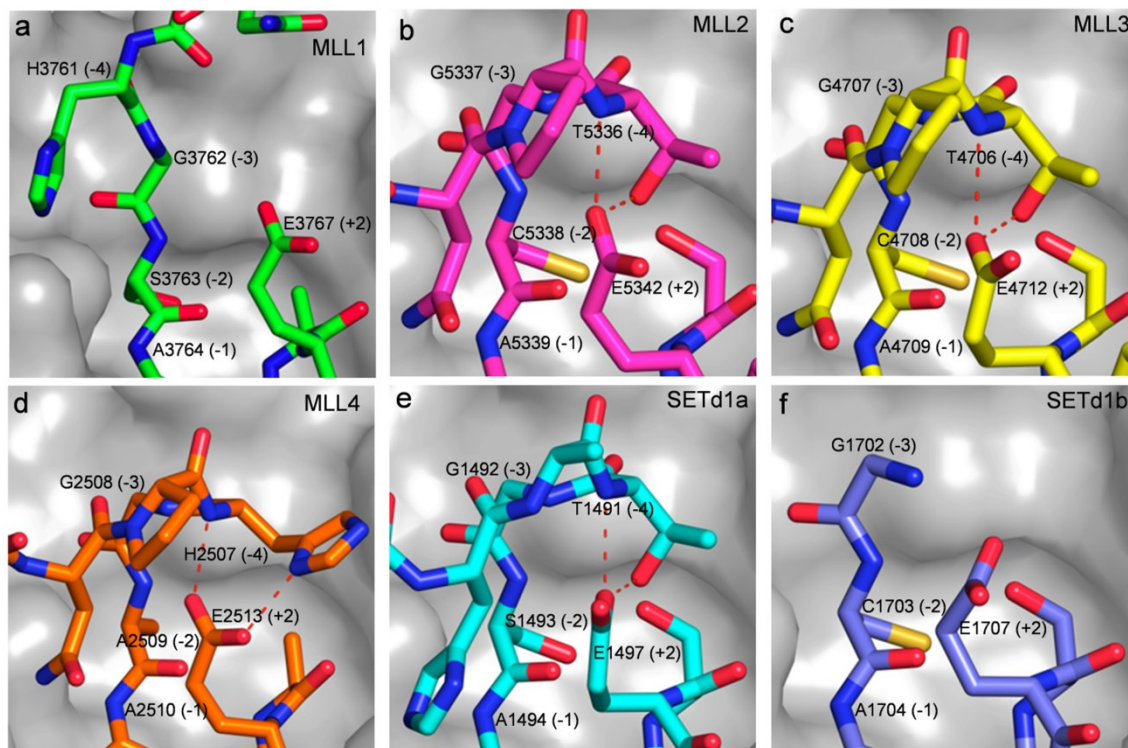
The second structural difference observed within the conserved *Win* motif is two residues C-terminal to the conserved arginine (position +2) where the methylene atoms of the conserved glutamate side chain adopt slightly different positions (Figure 4.9). However, the position of carboxylate of the +2 glutamate is in a relatively similar position in each structure. One other observed difference is the carbonyl of the glutamate in position +2 is oriented in the opposite direction in the SETd1a and SETd1b structures compared to that of the MLL1-4 structures (Figure 4.9). This change in orientation of the carbonyl oxygen in SETd1a and SETd1b

structures enables the carbonyl oxygen to participate in van der Waals contacts with the WDR5 residue K259. These observed differences within the conserved six-residue *Win* motif alone do not correlate with the 5- to 51-fold differences in affinity for peptide binding to WDR5, suggesting that the non-conserved residues that flank the *Win* motif may also be responsible for differences in affinity. Indeed, the largest differences in the six structures are observed in non-conserved the residues flanking the N- and C-termini of the *Win* motif.

### ***Structural differences at the N-terminus***

While there is some structural heterogeneity in the non-conserved residues flanking the N-terminus of the *Win* motif, in four out of the six structures, the N-terminal flanking residues (-4 to -7) adopt a U-shaped conformation in which the concave surface forms a pocket that interacts with the side chain carboxylate of the conserved +2 glutamate (Figure 4.10). In each of these structures, the +2 carboxylate forms two hydrogen bonds with the side chain and main chain atoms of the residue in the -4 position (either histidine or threonine) (Figure 4.10). These additional interactions may stabilize the bound conformation of the  $3_{10}$ -helix of the peptide when free in solution, which may enhance binding affinity. Similar N-terminal interactions may also be present in the other WDR5-*Win* motif complexes with MLL1 and SETd1b peptides in solution, but these interactions could not be observed in the crystal structures in this investigation. The N-terminal flanking residues of the SETd1b peptide co-crystal structure were disordered and could not be modeled (Figure 4.3f). In contrast, most of the N-terminal residues of the 19-residue MLL1 *Win* motif peptide were ordered, and the model reveals that they adopt a significantly extended conformation that is stabilized by interaction with another molecule in the

asymmetric unit (Appendix Figure 6). This extended conformation is therefore likely a crystallization artifact.



**Figure 4.10 Structural differences at the N-terminus**

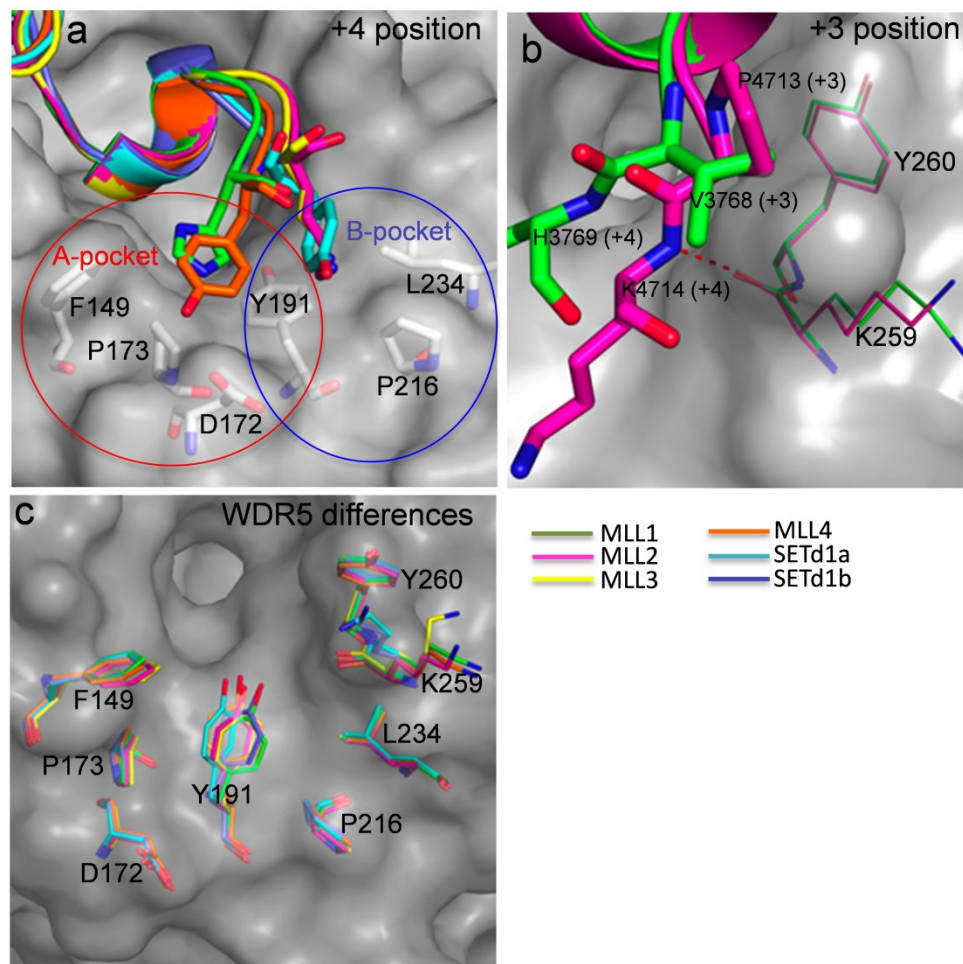
N-terminal residues flanking the conserved *Win* motif adopt structurally different conformations. (a-f). N-terminus (residues from -4 to -1) of MLL1 (a), MLL2 (b), MLL3 (c), MLL4 (d), SETd1a (e) and SETd1b (f) peptides are shown. Intramolecular hydrogen bonds between the -4 residue and +2 residue of MLL2, MLL3, MLL4 and SETd1a *Win* motif peptides are highlighted in red.

### ***Structural differences at the C-terminus***

In five out of the six structures, at least one residue flanking the C-terminal end of the *Win* motif could be modeled unambiguously (Figure 4.3). In the WDR5-MLL3 *Win* motif structure, all four C-terminal flanking residues could be modeled owing to its high resolution (1.22 Å). However, the last two residues (+5 and +6) do not appear to make any contacts with WDR5 (Figures 4.6-4.7). In contrast, in the WDR5-SETd1b structure, only one residue after the conserved +2 glutamate could be unambiguously modeled (Figure 4.3f).

A comparison of the C-terminal flanking residues among WDR5-*Win* motif peptide structures reveals that the +4 residue interacts with one of two shallow hydrophobic pockets on WDR5, which we arbitrarily designate as the A- and B-pockets (Figure 4.11a). The A-pocket is lined by WDR5 residues Y191, P173, F149 and D172. The +4 amino acid in the MLL1 and MLL4 peptides interacts with residues in the A-pocket. The B-pocket is lined by residues Y191, P216 and L234, which interact with the +4 amino acid side chain of the MLL2, MLL3, SETd1a *Win* motif peptides. In addition, B-pocket peptides form an additional main chain hydrogen bond between the amide nitrogen of the +4 amino acid and the carbonyl oxygen of WDR5 residue K259 (Figure 4.11b and Figure 4.6). This hydrogen bond is absent for peptides that bind in the A-pocket. All of the B-pocket *Win* motif peptides, albeit to varying degrees, have higher affinity for WDR5 compared to that of the MLL1 *Win* motif peptide. This increase in affinity may be due to the presence of this additional main chain hydrogen bond with the +4 amino acid in the B-pocket.

Differences in binding affinity is not correlated with the identity of the residue in the +4 position, as tyrosine is seen in both A- or B-pockets of the MLL4 and SETd1a structures, respectively (Figure 4.11a). Instead, A- or B-pocket binding appears to be correlated with the identity of the preceding residue in the +3 position. Common among A-pocket peptides is the presence of a valine in the +3 position of MLL1 and MLL4 structures. The valine side chain in the +3 position interacts with WDR5 residues Y260 and the aliphatic portion of the K259 side chain (Figure 4.11b). In contrast, B-pocket peptides have either proline (MLL2, MLL3) or glycine (SETd1a, SETd1b) in the +3 position. The +3 proline interacts more extensively with the Y260 side chain, and to a lesser extent with the K259 side chain (Figure 4.11b). As a result, the proline C $\alpha$  atom is shifted by 1.2-1.4 Å compared to that of the C $\alpha$  atom of valine in the MLL1 and MLL4 structures. Likewise, in the SETd1a and SETd1b structures, the C $\alpha$  atom of the +3 glycine forms van der Waals contacts with the C $\delta$ 1 and C $\epsilon$ 1 atoms of the Y260 side chain, correlating with a 2.2-2.4 Å shift in its position relative to that of the C $\alpha$  atom of valine in the MLL1 and MLL4 structures. These results suggest that it is the different interactions between WDR5 and the side chain or main chain atoms of the +3 amino acid of the *Win* motif that determines, at least in part, the location of the +4 amino acid.



**Figure 4.11 Structural differences at the C-terminus**

a-c. Superposition of SET1 family *Win* motif peptides showing interactions at the C-terminal *Win* motif flanking residues. In a, a closer look at the interactions at +4 position is shown. WDR5 residues that form the A- (red circle) and B- (blue circle) pockets are indicated as sticks (white). Different *Win* peptides are color coded as in Figure 4.8. In b, representative superposition of MLL1 (A-pocket) and MLL2 (B-pocket) peptides show the differences of interactions at the +3 position. The +3 and +4 amino acids from the MLL1 (green) and MLL2 (pink) structures are shown as stick representations. WDR5 residues (K259 and Y260) from the *Win* peptide structures are shown as lines in the same colors, respectively. The +4 residue in B-pocket peptides (MLL2, MLL3, SETd1a, and SETd1b) forms an intermolecular hydrogen bond (red dotted line) with the main chain carbonyl of WDR5 residue K259. c. superposition of A- (F149, D172, P173, and Y191) and B-pocket (Y191, P216, and L234) WDR5 residues is shown for all structures in the absence of *Win* peptides. The largest differences are observed in the side chains of Y191 and K259. The WDR5 residues are color coded as follows: green (MLL1), pink (MLL2), yellow (MLL3), orange (MLL4), cyan (SETd1a), and dark blue (SETd1b).



Despite differences in the positions of the C-terminal amino acids of the different *Win* motif peptides, there is remarkably little difference in the positions of WDR5 residues in the different complexes. Main chain atom positions of WDR5 are highly similar among the different *Win* peptide bound structures and superpose with RMS deviations of 0.1-0.4 Å (Table 4.3). The largest difference is a rotation around the C $\beta$ -C $\gamma$  bond of Y191, which rotates to accommodate interactions with the +4 amino acid of the different *Win* motif peptides (Figure 4.11c). In addition, the side chain of WDR5 residue K259 is oriented differently in the SETd1a and SETd1b *Win* motif peptide structures compared to that of the MLL1-4 structures (Figure 4.11c). The absence of a side chain in the +3 position of the SETd1a and SETd1b *Win* motif peptides likely allows a rotation around C $\alpha$ -C $\beta$  bond of WDR5 residue K259 bringing its side chain within van der Waals distance of the peptides.

Collectively, these structural analyses predict that the amino acid differences within and outside the conserved *Win* motif contribute in parts to the 50-fold difference in affinities for WDR5. To quantify the contribution of each of these differences to the overall binding affinities, we compared the binding properties of the 14-residue MLL1 and MLL4 *Win* motif peptides, which are highly similar in sequence (refer to Figure 1.2).

	MLL1	MLL2	MLL3	MLL4	SET1a	SET1b
MLL1	-	0.280	0.359	0.278	0.357	0.338
MLL2		-	0.343	0.156	0.391	0.386
MLL3			-	0.352	0.421	0.418
MLL4				-	0.339	0.318
SET1a					-	0.115
SET1b						-

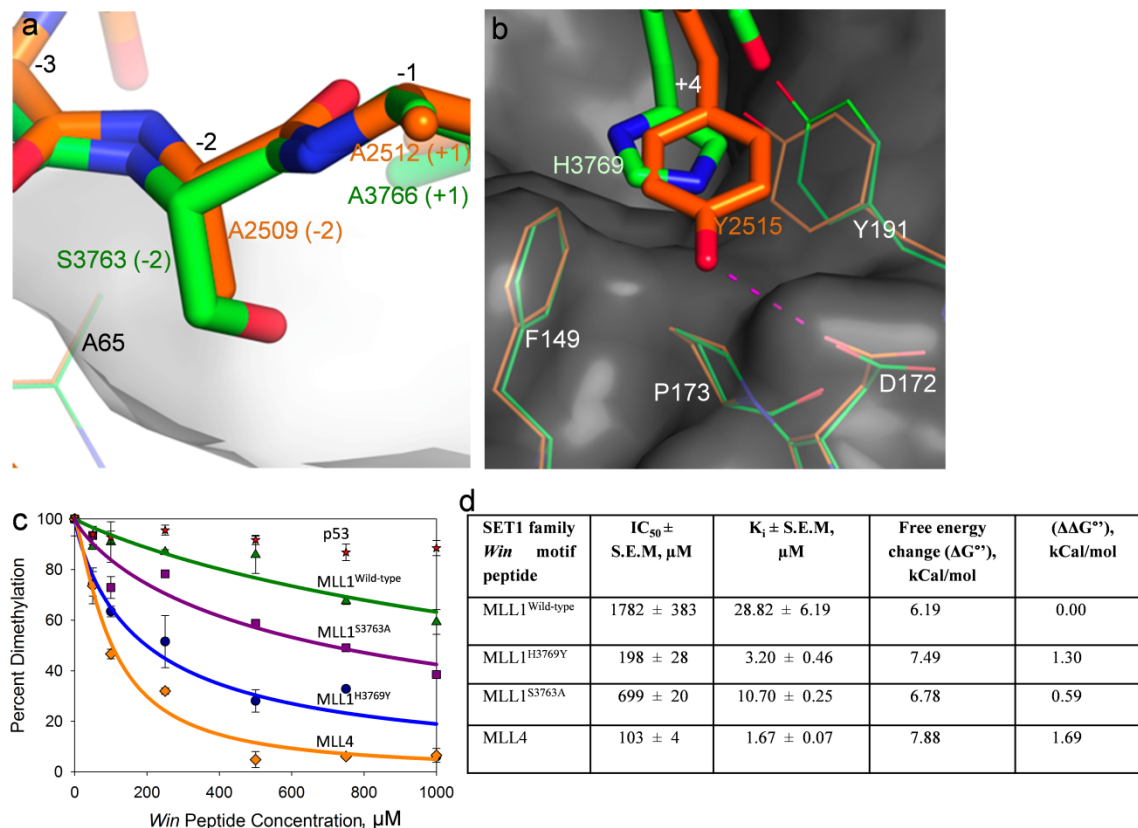
**Table 4.3: Pair-wise RMSD values for the superimposition of WDR5 molecules only from *Win* motif peptide bound structures**

***The -2 and +4 amino acids of SET1 family Win motifs contribute to high affinity binding to WDR5***

The amino acid sequences of the MLL1 and MLL4 peptides differ in only two positions (-2 and +4), yet the MLL4 peptide binds WDR5 with 31-fold greater affinity (Table 3.2). Comparison of the 3-dimensional structures reveals that the tyrosine hydroxyl atom of the +4 position in the MLL4 *Win* motif peptide forms a direct hydrogen bond with WDR5 residue D172 in the A-pocket (Figure 4.12b). A similar direct hydrogen bonding is absent in the MLL1 *Win* motif structure containing a histidine in the +4 position. Instead, a water-mediated hydrogen bond with D172 stabilizes the position of the histidine side chain (Figure 4.4a). These structural differences suggest that a direct hydrogen bond between the +4 amino acid of the *Win* motif peptide and D172 of WDR5 contributes significantly to the observed differences in affinity.

To test this hypothesis, peptides were synthesized that individually replaced the -2 and +4 residues in the MLL1 *Win* motif peptide with the corresponding residue in the MLL4 peptide, and were compared for their ability to inhibit the H3K4 dimethylation activity of the MLL1 core complex in vitro. When the -2 serine of the MLL1 *Win* motif peptide was replaced with alanine,  $IC_{50}$  decreases 3-fold compared to that of the 14-residue wild-type MLL1 *Win* motif peptide (Figure 4.12c and d). In contrast, when the +4 histidine residue of the MLL1 *Win* motif peptide is replaced with tyrosine,  $IC_{50}$  decreased 9-fold compared to that of the wild-type MLL1 *Win* motif peptide (Figure 4.12c and d). These results suggest that the tyrosine in the +4 position of the *Win* motif peptide contributes  $\sim 1.3$  kcal/mol to binding WDR5 within the context of the MLL1 core complex, likely due to the formation of a direct hydrogen bond between the +4 tyrosine hydroxyl and D172 in WDR5 (Figure 4.12d). In contrast, replacement of -2 serine with an alanine contributes  $\sim 0.6$  kcal/mol to binding (Figure 4.12d). We hypothesize, based on the crystal

structures of the MLL1 and MLL4 *Win* motif peptides that this contribution results from an increase in entropy associated with easier desolvation of the alanine in the -2 position compared to that of serine for insertion into a relatively hydrophobic pocket on WDR5 (Figure 4.12a).



**Figure 4.12: The -2 and +4 amino acids of SET1 family *Win* motifs contribute to high affinity binding to WDR5.**

Structural differences between MLL1 and MLL4 *Win* motif peptides contribute to the differences in binding affinity to WDR5. (a-b): Superposition of MLL1 and MLL4 *Win* motif peptides. a, Superposition showing the -2 position of MLL1 (green) and MLL4 (Orange) *Win* motif peptides. WDR5 residues are shown as line representations. b, Superposition showing the +4 amino acid in MLL1 (green) and MLL4 (orange) structures. MLL4 residue Y2515 (+4 position) sits in the A-pocket and forms a direct hydrogen bond (pink dotted line) with the WDR5 residue D172. c,  $\text{IC}_{50}$  plots for the inhibition of the H3K4 dimethylation activity of the fully assembled MLL1 core complex by the wild-type and mutant MLL1 *Win* motif peptides. For comparison, MLL4 and the control p53 peptide plots are also shown. The methylation reactions were carried out as described in Figure 3.6. Data presented is an average of two independent experiments. d,  $\text{IC}_{50}$ ,  $K_i$ , and the standard free energy of binding values derived from the fits to the data in c are summarized.  $\Delta\Delta G^\circ$  values represent the difference in free energy of binding between wild-type and mutant MLL1 *Win* motif peptides.

Together, the structures of all six human SET1 family *Win* motif peptides bound to WDR5 reaffirm the earlier predictions that all human SET1 *Win* motif based peptides bind WDR5 using the same arginine binding pocket in WDR5. While the WDR5-*Win* motif interactions are highly conserved among the different *Win* motif peptides, the structures reveal that residues flanking the N- and C-termini of the *Win* motif are structurally different-suggesting that these amino acid differences contribute to the majority of differences in WDR5 binding affinities. Collectively, these structural data provide a molecular basis for the high interaction affinities of all human SET1 family *Win* motif based peptides to WDR5 and suggest a framework for the design of additional *Win* motif based inhibitors that target this interaction.

## DISCUSSION

*Win* motif peptides from other human SET1 family members bind WDR5 with 5- to 51-fold greater affinity than that of a similar length MLL1 *Win* motif peptide. Three-dimensional structures reported in this chapter reveal that each peptide inserts the conserved arginine of the *Win* motif into the central tunnel of WDR5. Surprisingly, despite the large differences in affinity, the three-dimensional structure of WDR5 displays remarkably little variation when bound to the different peptides. Instead, most of the variation is observed in the non-conserved residues located C-terminal to the *Win* motif in SET1 family sequences. WDR5 appears to have evolved two separate hydrophobic pockets to accommodate different sequences in the +3 and +4 positions C-terminal to SET1 family *Win* motifs. In each case, high affinity binding appears to be conferred by an additional direct hydrogen bond either through main chain or side chain interactions between WDR5 and the +4 amino acid of SET1 family sequences. These

interactions likely account for the majority of the affinity differences among SET1 family peptides. However, structural analysis also suggests that subtle variation within the conserved six-residue *Win* motif sequence also contributes to the observed differences in binding energy. We speculate that binding may be facilitated if the conformation of the  $3_{10}$ -helix is stabilized in solution by a hydrogen bond between the side chain in the -2 and +1 positions of the *Win* motif. Indeed, The MLL2, MLL3 and SETd1b peptides possess such a hydrogen bond and they bind WDR5 with the greatest affinities. For those peptides lacking this hydrogen bond, binding may be facilitated by residues that are easier to desolvate when binding to WDR5. This may explain why replacing the -2 serine in the MLL1 peptide with alanine results in 3-fold greater affinity to WDR5.

A similar study was carried out by Zhang et al., (333) who reported structures of WDR5 bound to six 11-residue peptides containing *Win* motif sequences flanked by 5 additional residues on the C-terminus. Based on their crystal structures they proposed that amino acid differences C-terminal to the *Win* motif are important for affinity differences among peptides used in their investigation. The results presented in this chapter are consistent with that observation, but also suggests that subtle variation within the conserved *Win* motif may also contributes to binding energy differences. While the study by Zhang et al (333) showed that their MLL3 *Win* motif peptide has the weakest affinity for WDR5, the results presented as a part of Chapter 3 contrasts with that report and demonstrates that MLL3-like peptide used in this study binds WDR5 with the strongest affinity. Careful analysis revealed that the peptide used in the study by Zhang et al., (333) is different from the wild type human MLL3 sequence used in this investigation in the -2 and +1 positions and also lacks the four N-terminal *Win* motif-flanking residues. The mutant MLL3 peptide used by Zhang et al (333) contains a serine and alanine in

the -2 and +1 positions of the *Win* motif, respectively; whereas the wild-type MLL3 *Win* motif contains a cysteine and serine in the same respective positions. The three-dimensional structure of the wild type MLL3 *Win* motif presented here reveals that the -2 cysteine and +1 serine are within hydrogen bonding distance. This hydrogen bond is absent in the mutant MLL3 peptide, which may explain its reduced affinity for WDR5. It is also possible that reduced affinity for the mutant MLL3 peptide could be due to the absence of the four N-terminal flanking residues, which hydrogen bond with the conserved +2 glutamate. The presence of these features in the wild type MLL3 peptide may improve overall affinity by reducing the number of different conformations of the peptide that WDR5 must sample in solution.



## **Chapter 5: *Win* motif-based inhibitors: A novel class of methylation inhibitors with potential to treat Acute Myeloid Leukemia**

This chapter is a summary of preliminary studies that are aimed at identifying additional *Win* motif-based inhibitors and the initial experiments to test the effect of *Win* motif peptides in cells. The data presented in this chapter are unpublished and part of the work was done in collaboration with Dr. Ziwei Huang's group at the Upstate Cancer Institute. The Immunofluorescence experiments described in this chapter were done by Dr. Laila Kabrossy (our former postdoc) and were included in this thesis for completeness.

## INTRODUCTION

The extensive biochemical and structural characterization of human SET1 family *Win* motif peptides described in Chapters 3 and 4 have demonstrated that *Win* motif peptides or related peptidomimetic compounds can inhibit the H3K4 methyltransferase activity of the MLL1 core complex by disrupting the association of MLL1 with the WRAD sub-complex. Furthermore, the crystal structures of WDR5 bound to human SET1 family *Win* motif peptides (Chapter 4) have identified unique protein structural features that determine the high affinity binding. The structures also reveal that conserved residues from the *Win* motif (-3 to +2) form a cyclical shaped  $3_{10}$ -helix that fits snugly into the outer opening in WDR5 and raises the possibility that other cyclic peptidomimetic compounds might also bind WDR5 with similar or better inhibitory properties. These structural results provide a rationale for the design of new peptides and non-peptide mimetics with better inhibitory properties. We have addressed this possibility in Chapter 5 using a combination of systematic mutagenesis, chemical library screening, and phage display screens and have identified new *Win*-motif like inhibitors that bind WDR5 with  $< 10$  nM affinities. These peptides are also 10-fold better inhibitors of MLL1 core complex activity compared to the MLL3 *Win* motif peptide, which was identified as the best inhibitor in my initial analyses reported in Chapters 3 and 4. In addition, the results of a randomized phage display screen presented in this chapter have identified six additional peptides that are also highly potent inhibitors of MLL1 core complex activity. Based on the crystal structures of peptides derived from the phage display, we have identified novel protein structural features that contribute to high WDR5 interaction affinities. Furthermore, using a chemical library screen we have also identified four chemical compounds that mimic *Win* motif binding to WDR5 and inhibit the H3K4 dimethylation activity of MLL1 core complex.

The efficacy of *Win* motif based inhibitors in down regulating the H3K4 methylation activity of MLL1 core complex *in vitro* has yet to be demonstrated *in vivo*. However, based on the *in vitro* studies it is expected that the *Win* motif based peptides will have the ability to down regulate MLL1 core complex dependent H3K4 di/trimethylation *in vivo*. We have tested this hypothesis in Chapter 5 using a fluorescent tagged MLL3 *Win* motif based peptide that has a cell penetrating sequence linked to its C-terminus. The results show that this peptide is readily taken up by mammalian cells and the peptide is localized at euchromatic regions. The cells treated with the MLL3 peptide display distinct nuclear defects and decreased global levels of H3K4 trimethylation. However, a control mutant MLL3 peptide wherein the conserved *Win* motif arginine is mutated to alanine does not inhibit H3K4 trimethylation. These results suggest that MLL3 *Win* motif peptide is a highly potent inhibitor of H3K4 trimethylation in cells.

#### ***Targeting WDR5-Win motif interaction as a novel strategy for treating MLL1 amplifications and MLL1-PTDs***

Growing evidence suggests that increased HOXA gene expression associated with MLL1 amplifications and MLL1-PTDs underlie the pathogenesis of these gain-of-function form of leukemias (79, 308). HOXA gene expression is dependent on the histone methyltransferase activity of MLL1 (54), and given the increased H3K4 methylation observed in MLL1-PTDs (78, 79), inhibitors that down-regulate the histone methyltransferase activity of MLL1-PTD have the potential to reverse this aberrant epigenetic program. Until recently, inhibitors that specifically target MLL1's H3K4 methyltransferase activity have been difficult to develop due to lack of a clear understanding of MLL1's structure and function. However, recent findings have established that H3K4 methylation levels are precisely regulated by the function of two

independent methyltransferases: the MLL1 SET domain and a novel multi-subunit enzyme, WRAD, that lacks amino acid sequence homology to known methyltransferases (32, 33). Because multiple methylation on H3K4 is catalyzed by two different enzymes, it is interesting to speculate that inhibitors that prevent the association of two methyltransferases will have clinical significance in the treatment of acute leukemias that are characterized by aberrant H3K4 methylation (78, 79). Binding and inhibition studies reported in this dissertation demonstrate that human SET1 family *Win* motif peptides and related peptidomimetic compounds can target the histone methyltransferase activity of MLL1. These *Win* motif peptides hence have the potential to down regulate the increased H3K4 dimethylation, and aberrant HOX gene expression associated with MLL1-PTDs and gene amplification mutations in MLL1. Therefore, the *Win* motif peptides represent a group of novel “first in class” inhibitors that are expected to have the ability to down- regulate H3K4 dimethylation levels in the cell without perturbing H3K4 monomethylation, or the methylation activities of other H3K4 methyltransferases.

## METHODS

### *Peptide synthesis*

Ac-10-mer ( $_{\text{Ace}}\text{ARTEVHLRKS}_{\text{NH}_2}$ ) and the six residue ( $_{\text{Ace}}\text{ARTEVY}_{\text{NH}_2}$ ) *Win* motif peptides used in this chapter were synthesized by Genscript. Both peptides were synthesized with an acetyl- and amide-capping group at the N- and C-terminus, respectively, to eliminate the contributions of unnatural N- and C-terminal charges on binding. The peptides identified from the initial Phage display screen (Ph. D <sup>TM</sup> Phage Display Screen kit from New England BioLabs) (refer to table 5. 3 for peptide sequences) were synthesized by Dr. Ziwei Huang’s laboratory and do not have capping groups on the N- and C-termini. MLL3<sup>WT-FITC-TAT</sup> peptide

(<sub>Ace</sub>VNPTGCARSEPK-Ahx-K(FITC)-GRKKRRQRRRAP<sub>NH2</sub>) and the control arginine mutant peptide MLL3<sup>R4710A-FITC-TAT</sup> peptide (<sub>Ace</sub>VNPTGCAASEPK-Ahx-K(FITC)-GRKKRRQRRRAP<sub>NH2</sub>) used in the *in vivo* experiments were also synthesized by Dr. Ziwei Huang's laboratory. The MLL3<sup>FITC-TAT</sup> peptides used in the *in vivo* experiments has only the first 12 residues (VNPTGCARSEPK) of the original MLL3 *Win* motif peptide used in Chapters 3 and 4 because the last two C-terminal residues (methionine and serine) did not appear to make any contacts with WDR5 in the crystal structure. A Fluorescent IsoThioCynate (FITC) moiety was attached to the C-terminus of MLL3 *Win* peptide using an Ahx (6-Aminohexanoic acid, C<sub>6</sub>H<sub>13</sub>NO<sub>2</sub>) linker and a cell penetrating sequence derived from the HIV-TAT protein (RKKRRQRRRAP) was also attached to the C-terminus of this peptide to enable intracellular uptake. Chemical compounds NP7 (2-chloro-N-[2-(4-methylpiperazin-1-yl)-5-nitrophenyl] benzamide), OBW (methyl 3-[(3-methoxybenzoyl) amino]-4-(4-methylpiperazin-1-yl) benzoate), WD1 and WD2 used in the inhibition assays were also provided by Dr. Ziwei Huang's laboratory.

### ***Isothermal Titration Calorimetry***

Isothermal Titration Calorimetry (ITC) experiments using the Ac-10-mer and the six residue *Win* motif peptides were carried out as described previously in Chapter 3. Prior to ITC, all peptides and proteins were dialyzed against sample buffer to minimize variations in sample preparations. Individual ITC experiments for each of the peptides were carried out by titrating a known concentration of *Win* motif peptide (0.425-0.490 mM) (determined by amino acid analysis at the KECK proteomics facility at Yale University) diluted in the sample buffer into a sample cell containing a known concentration of full-length WDR5 (0.050mM) in the same

buffer. Binding stoichiometry ( $N$ ), dissociation constant ( $K_d$ ), standard enthalpy ( $\Delta H$ ), entropy ( $\Delta S$ ), and free energy ( $\Delta G$ ) changes associated with *Win* peptide binding to WDR5 were derived by fitting the binding isotherm to an one-site binding model (Origin 7.0).

### ***MLL1 core complex inhibition assays***

MALDI-TOF mass spectrometry based methyltransferase assays were performed as described previously in Chapter 3. Methylation assays for the Ac-10-mer and MLL3-FITC-TAT peptides were carried out by incubating 7.3  $\mu\text{M}$  of the MLL1 core complex in the absence and presence of increasing concentrations of peptide (0-250  $\mu\text{M}$ ). For the peptides identified from phage display screen (Peptides 1-8) and the chemical compounds (NP7, OBW, WD1 and WD2), 7.3  $\mu\text{M}$  MLL1 core complex was incubated with 250  $\mu\text{M}$  inhibitor and the amount of unmodified, mono-, di-, and trimethylated histone H3 peptides present after 12 hours were quantified by MALDI-TOF mass spectrometry. Curve-fitting and analysis were carried out as described previously in Chapter 3.

### ***Crystallization and Structure Determination***

Crystals of the Ac-10-mer ( ${}_{\text{Ace}}\text{ARTEVHLRKS}_{\text{NH}_2}$ ) *Win* peptide, Peptide 3 (GARTEVPFLTIF) and Peptide 6 (SARVAIEYNTAR) from the phage display screen were obtained as described in Chapter 4. Immediately before crystallization, a 13 mg/ml stock solution of  $\Delta\text{N}$ -WDR5 in the sample buffer was mixed with a stock solution of each of the peptide dissolved in the same buffer. The final concentration of  $\Delta\text{N}$ -WDR5 and the individual peptides were 11.7 mg/ml and 1 mM, respectively. The mother liquor used for crystallization contained 20-30 mM ammonium sulfate, 25-27% polyethylene glycol (PEG)-3350, and 100 mM HEPES (pH 7.3-7.5). The crystals were flash frozen in the mother liquor containing 40% PEG3350. The diffraction data

were collected at the Cornell High Energy Synchrotron Source (CHESS) at the F1 beamline and processed as previously described in Chapter 4. The structures of Ac-10-mer and Peptide 6 were solved by difference fourier analysis in CNS using as a search model the coordinates of the 12-residue MLL1 *Win* motif peptide structure (PDB code:3EG6), which also crystallized in C222<sub>1</sub> space group with similar unit cell dimensions (60). Structure of Peptide 3 was solved by molecular replacement using as a search model the coordinates of the apo-WDR5 crystal structure (PDB code: 2H68). Molprobity was used to analyze the protein geometry and steric clashes in final refined structures and the statistics are reported as molprobity score in Table 5.2 (329, 330). Data collection and refinement statistics for the different peptide structures are summarized in Table 5.2.

### ***Immunofluorescence***

In cell experiments with the MLL3<sup>WT-FITC-TAT</sup> and MLL3<sup>R4710A-FITC-TAT</sup> peptides were carried out in P19 cells (mouse embryonal carcinoma). Approximately,  $0.5 \times 10^6$  cells were plated on a 6-well plate along with 2ml of DMEM media and were allowed to divide till ~60% confluence. For the peptide localization immunofluorescence experiments, MLL3<sup>WT-FITC-TAT</sup> peptide was diluted to 50 $\mu$ g/ml in 2ml of DMEM media and added to P19 cells attached in a 6-well plate. After 30min, 4 hr and 24 hours of peptide treatment, the cells were washed 2X with phosphate buffer saline (PBS) and fixed by adding 70% ethanol for 20 min at room temperature. The cells were washed 3X with PBS to remove excess ethanol and stained with DAPI stain (final concentration of 1  $\mu$ g/ml) for 10 min at room temperature. The cells were washed 3X with PBS and mounted on the cover slip using 30  $\mu$ l of mounting media. Immunofluorescence images were acquired using a Zeiss Axioplan2 microscope (Carl Zeiss, Inc., NY). Images were acquired with a Hamamatsu

Orca II C4742-98 dual scan-cooled CCD camera (Hamamatsu Photonics), an X-Cite 120 light source (EXFO America), and filter sets for FITC (excitation wavelength=499 nm), and DAPI (excitation wavelength=345 nm) fluorescence. MetaMorph imaging software (Universal Imaging Corp., PA) was used to control illumination shutters, camera exposure and image acquisition. The fluorescence images were collected with the help of Dr. Torsten Wollert in Dr. Langford's lab at Department of Biology, Syracuse University.

### ***Western Blotting***

Western blots to measure the global levels of H3K4 mono- and trimethylation in the presence or absence of MLL3<sup>FITC-TAT</sup> peptides were carried out in P19 cells. Approximately,  $2 \times 10^6$  P19 cells were plated on a 6-well plate in the presence of 2 ml of DMEM media and allowed to divide to <50% confluence. The cells were then treated with increasing concentrations (0-100  $\mu\text{g/ml}$ ) of MLL3<sup>WT-FITC-TAT</sup> or MLL3<sup>R4710A-FITC-TAT</sup> peptides in 2 ml of fresh media for 3 hours. After 3 hours of peptide treatment, 0.5  $\mu\text{l}$  of retinoic acid (RA) was added to induce differentiation in P19 cells. After 24 hours of RA treatment, the cells were lysed in 100  $\mu\text{l}$  of RIPA lysis buffer and the whole cell extracts were separated on a 4-12% SDS PAGE. The protein bands were then transferred to a nitrocellulose blot for 1-3 hours at 4 °C. Blocking was done using bovine serum albumin for 1 hour at 37°C followed by washing 3X with PBS. Western blots were carried out using antibodies specific for mono- and trimethylated H3K4 (H3K4 me1 and me3), trimethylated H3K27 (H3K27me3), histone H3, and tubulin for 1 hr at 37 °C. The blot was then washed 3X with PBS to remove unbound antibodies and then incubated with the HRP-conjugated anti-rabbit secondary antibody for 1 hr at 37 °C. After the final wash, the protein bands were visualized using chemiluminescence (ECL kit, GE Health care).



## RESULTS AND DISCUSSION

### *5.1 Identification of key structural elements required for binding WDR5: Analysis of the binding of MLL1 Win motif peptide to WDR5*

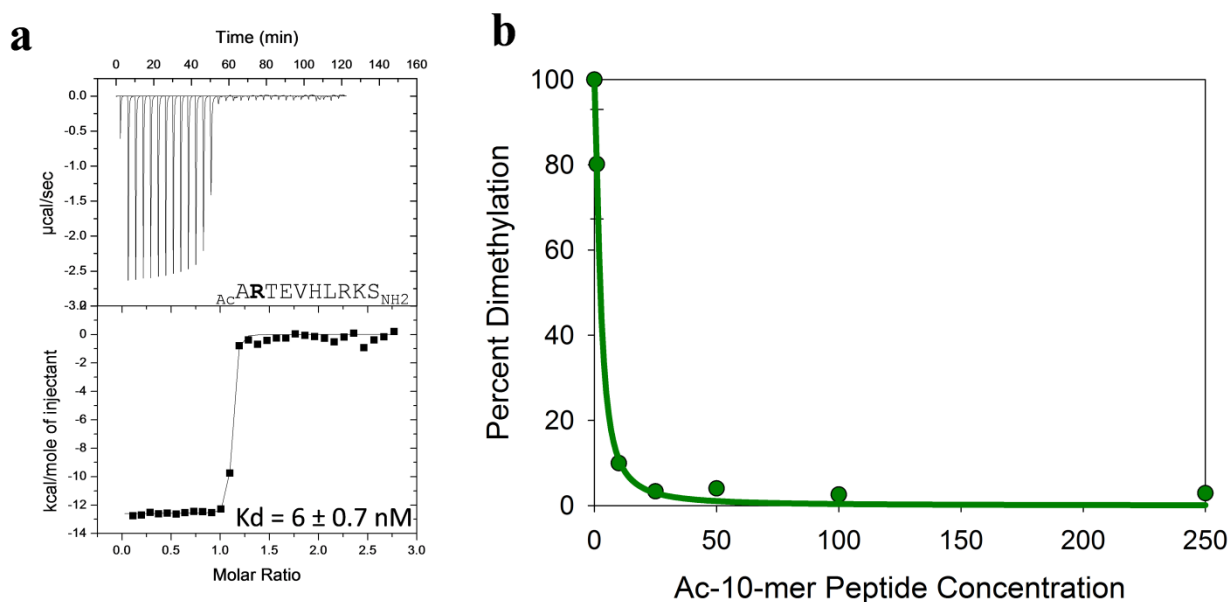
To facilitate the design of small-molecule inhibitors that target MLL1-WDR5 interaction, Karatas et al (310) performed a systematic deletion analysis on the original 12-residue MLL1 *Win* motif peptide (3762-3773) and showed that a three residue sequence composed of  $\text{Acetyl-ARA-NH}_2$  is the minimal motif that is required for its interaction with WDR5 (310). The acetyl group on the N-terminus is essential as its removal results in the loss of binding. The  $\text{Acetyl-ARA-NH}_2$  peptide binds WDR5 in peptide competition experiments with similar affinity ( $K_i = \sim 160$  nM) to that of the original 12 residue MLL1 *Win* motif peptide ( $K_i = \sim 120$  nM) (61, 310). Moreover, systematic mutagenesis and molecular dynamics simulations further suggest that  $\text{Acetyl-ARA-NH}_2$  peptide could recapitulate the two intramolecular  $i \rightarrow i+3$  main chain hydrogen bonds that are present in the original 12-residue MLL1 *Win* motif peptide. Absence of one or both of these hydrogen bonds significantly weakens the interaction affinity between the 3-residue  $\text{Acetyl-ARA-NH}_2$  peptide and WDR5 (310). These results suggest that  $\text{Acetyl-ARA-NH}_2$  peptide binds WDR5 by adopting a partial  $3_{10}$ -helical conformation similar to that of the 12-residue MLL1 *Win* motif peptide and participates in similar sets of interactions (61, 310). However, the most potent inhibitor derived from the MLL1 *Win* motif peptide sequence based on their analysis is the acetyl-10mer ( $\text{acetyl-ARA EVHLRKS-NH}_2$ ) encompassing residues 3764-3773, which binds WDR5 (in peptide competition experiments) with a  $K_i = 3$  nM (310). These results suggest that sequences outside of the  $\text{-ARA-}$  sequence contribute to the affinity and would likely increase specificity. In addition, replacement of alanine at the +1 position (1 residue C-terminal to R3765) with a threonine in the acetyl-10mer ( $\text{acetyl-ARTEVHLRKS-NH}_2$ ) or 3-mer ( $\text{Acetyl-ART-NH}_2$ ) results

in a 3-to-6-fold increase in binding affinity to WDR5, respectively (310). In summary, these experiments suggest that the  $_{\text{Acetyl}}\text{-ARA-NH}_2$  motif anchors the peptide to the arginine binding pocket of WDR5 and that the  $3_{10}$ -helical conformation is crucial for the interaction. In addition, sequences N- and C-terminal to the ARA motif are important for increasing binding specificity. It remains to be determined if the  $_{\text{Acetyl}}\text{-ARA-NH}_2$  peptide actually inhibits the WDR5-MLL1 interaction within the context of the assembled MLL1 core complex.

### ***5.1.1 Binding and inhibition studies for the Acetyl-10-mer peptide***

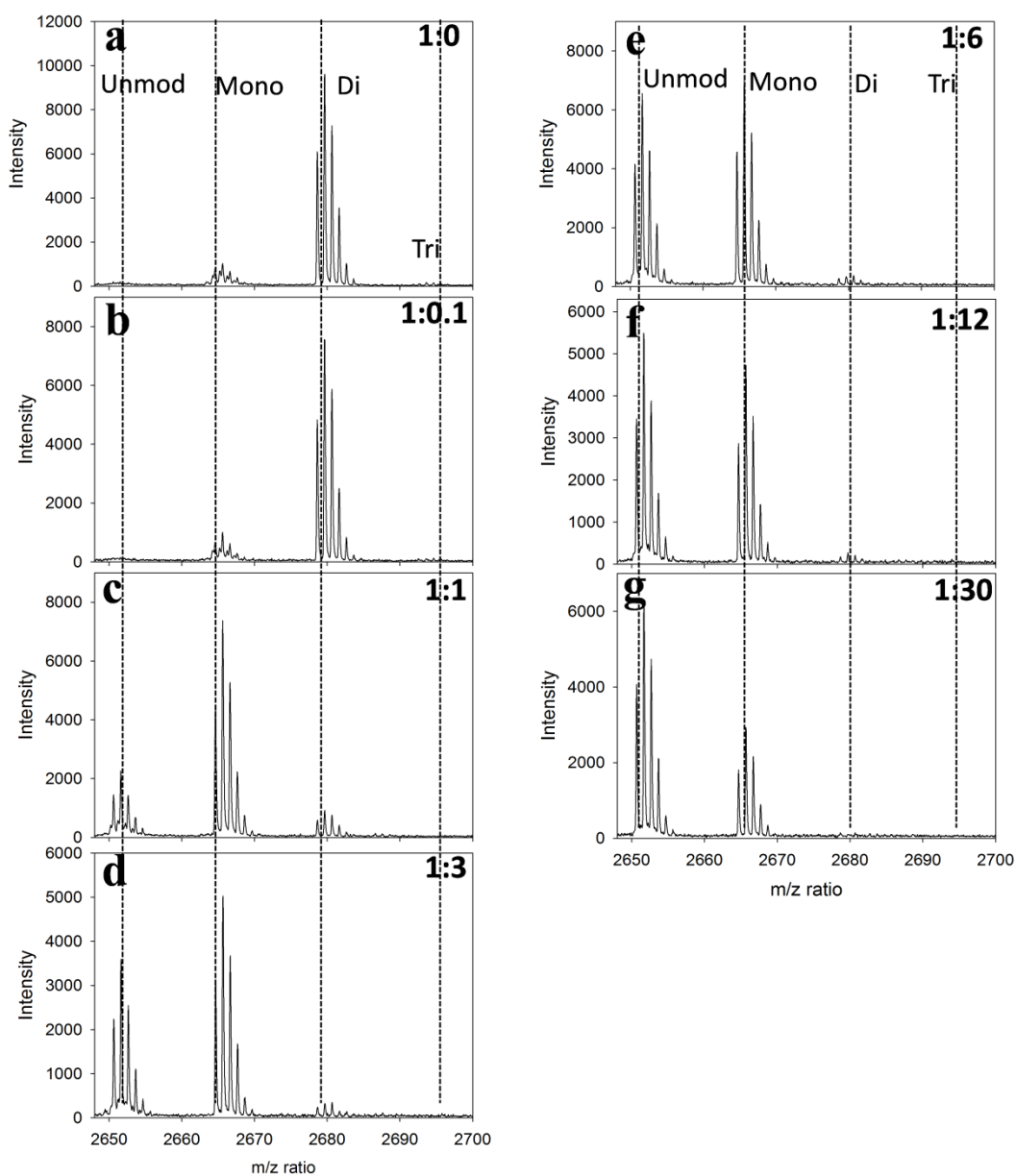
Karatas et al (310) has identified the  $_{\text{Ace}}\text{-ARTEVHLRKS-NH}_2$  peptide (designated Ac-10-mer) as the best inhibitor in their peptide competition experiments. To test if the Ac-10-mer can bind WDR5 and inhibit the MLL1 core complex dependent H3K4 dimethylation activity under our assay conditions, we synthesized a similar Ac-10-mer peptide with acetyl and amide capping groups at the N- and C-termini, respectively. The results of ITC based binding experiments show that the Ac-10-mer peptide binds WDR5 with a dissociation constant ( $K_d$ ) of 6 nM (Figure 5.1a). This binding affinity is 10-fold and 460-fold better than the affinities reported for MLL3 (54 nM) and MLL1 (2762 nM) *Win* motif peptides, respectively (refer to Table 3.2). While this binding constant is in correlation with the binding affinity ( $K_i=1$  nM) reported for the Ac-10-mer peptide in the study by Karatas et al., (310), it has not been shown if the Ac-10-mer peptide does indeed inhibit the H3K4 dimethylation activity of MLL1 core complex with greater efficiency. To test this hypothesis, we next tested the Ac-10-mer peptide for the inhibition of MLL1 core complex activity using MALDI-TOF mass spectrometry. Expectedly, the Ac-10-mer peptide inhibited the H3K4 dimethylation activity with a high efficiency (Figure 5.2). Significant inhibition was seen at a peptide concentration that is in a 1:1 molar ratio with the MLL1<sup>3745</sup> protein (refer to Figure

5.1b and Figure 5.2c). The  $IC_{50}$  value derived from the fit is  $2.5\mu\text{M}$  (Table 5.1), which is 10-fold lower than that reported for the MLL3 *Win* motif peptide ( $25\mu\text{M}$ ) in Chapter 3. These results indicate that the Ac-10-mer is a highly specific inhibitor of H3K4 dimethylation activity. However, it should be noted that the sequence of Ac-10-mer peptide (ARTEVHLRKS) resembles closely the sequence of histone H3 substrate (ARTKQTARKS) and could possibly compete with histone H3 peptide for the binding the active site in MLL1 core complex. Such a possibility would also result in loss of H3K4 methylation but due to a mechanism that differs from that proposed for *Win* motif peptides. However, monomethylation catalyzed by the MLL1 core complex (in the presence or absence of Ac-10-mer peptide) is not significantly altered suggesting that the histone H3 peptide substrate can still bind to the MLL1 core complex active site (Figure 5.2). Additional experiments are required to test the binding of the Ac-10-mer peptide to the MLL1 SET domain or the WRAD active site. Nevertheless, the Ac-10-mer peptide still represents a highly potent inhibitor of MLL1 core complex activity.



**Figure 5.1** Ac-10-mer peptide ( $\text{Ac-ARTEVHLRKS}_{\text{NH}_2}$ ) binds WDR5 with the strongest affinity

a-b. Binding and inhibition data for Ac-10-mer peptide. a, Isothermal Titration Calorimetry (ITC) data for the binding of Ac-10-mer peptide with WDR5. Upper panels show heat of binding plotted as a function of time. Lower panels show the binding isotherms fit to a one-site binding model. Ac-10-mer peptide sequence and the dissociation constants ( $K_d \pm \text{S.E.M.}$ ) derived from the fit are indicated. S.E.M values are derived from two independent experiments. b, H3K4 dimethylation inhibition data for Ac-10-mer peptide. MALDI-TOF assays were carried out as described in 3.6. Relative H3K4 dimethylation levels in the presence of increasing concentrations of Ac-10-mer peptide are shown. Error bars represent the variation observed in two independent experiments.  $\text{IC}^{50}$  values and inhibition constants ( $K_i$ ) were derived as described in Figure 3.5 and reported in Table 5.1.



**Figure 5.2: Ac-10-mer peptide specifically inhibits the H3K4 dimethylation activity of MLL1 core complex**

Ac-10-mer peptide inhibits the H3K4 dimethylation activity of the MLL1 core complex. a-g. MALDI-TOF assays for the H3K4 methylation activity catalyzed by the MLL1 core complex. MALDI-TOF assays were carried out with  $7.3 \mu\text{M}$  MLL1 core complex for 12 hours in the absence or presence of increasing concentrations of Ac-10-mer peptide. Relative levels of unmodified, mono-, di-, and trimethylation on H3K4 in the absence (a) or presence of increasing concentrations of Ac-10-mer peptide (b-g) are shown. Fold-excess of the Ac-10-mer peptide compared to MLL<sup>3745</sup> in each reaction is indicated within the panels.

Peptide	Sequence	Dissociation Constant, $K_d$ $\pm$ S.E.M, nM	IC50 $\pm$ S.E.M, $\mu$ M	$K_i$ $\pm$ S.E.M, $\mu$ M
Ac-10-mer	$_{Ac}$ ARTEVHLRKS $_{NH_2}$	6 $\pm$ 0.7	2.49 $\pm$ 0.98	0.04 $\pm$ 0.01
Ac-6-mer	$_{Ac}$ ARTEVY $_{NH_2}$	9 $\pm$ 4	N.D	N.D
MLL1	$_{Ac}$ LNPHGSARAEVHLR $_{NH_2}$	2762 $\pm$ 338	1782 $\pm$ 383	28.82 $\pm$ 6.19
MLL2	$_{Ac}$ INPTGCARSEPKIL $_{NH_2}$	75 $\pm$ 5	45 $\pm$ 0	0.73 $\pm$ 0.00
MLL3	$_{Ac}$ VNPTGCARSEPKMS $_{NH_2}$	54 $\pm$ 5	25 $\pm$ 12	0.40 $\pm$ 0.19
MLL4	$_{Ac}$ LNPHGAARAEVYLR $_{NH_2}$	88 $\pm$ 16	103 $\pm$ 4	1.67 $\pm$ 0.07
SETd1a	$_{Ac}$ EHQTGCARSEGYP $_{NH_2}$	541 $\pm$ 46	132 $\pm$ 1	2.14 $\pm$ 0.02
SETd1b	$_{Ac}$ EHVTGCARSEGFYT $_{NH_2}$	103 $\pm$ 14	108 $\pm$ 5	1.75 $\pm$ 0.09

N.D= Not Determined

**Table 5.1: Summary of thermodynamic binding parameters for the Ac-10-mer and 6-mer peptides shown in comparison with human SET1 family *Win* motif peptides**

### 5.1.2 Structural basis for the high affinity binding of Ac-10-mer peptide to WDR5

To understand the molecular basis for the 10-fold better affinity of the Ac-10-mer peptide compared to that of the MLL3 *Win* motif peptide, we determined the x-ray structure of the Ac-10-mer peptide bound to WDR5 at 1.6 Angstroms. X-ray data collection and refinement statistics are summarized in Tables 5.2 and 5.3. Six of the ten residues could be modeled unambiguously into the electron density map, while the last four C-terminal residues (LRKS) are disordered (Figure 5.3a). The structure reveals that the Ac-10-mer peptide binds WDR5 using a  $3_{10}$  helical conformation (Figure 5.3b) as previously shown for human SET1 family *Win* motif peptides (Figure 4.8a). Similar to the other *Win* motif peptide structures, the side chain of the conserved arginine (at 0 position) is inserted into the central water-filled tunnel in WDR5 and is stabilized by an extensive network of hydrogen bonds, van der Waals interactions and hydrophobic contacts (Figures 5.4 and 5.5). The  $3_{10}$ -helical structure is stabilized by two intramolecular  $i$  to  $i+3$  hydrogen bonds between the -2 (acetyl carbonyl) and +1 (threonine amide), and -1 (alanine carbonyl) and +2 (glutamate amide) residues, respectively (Figures 5.4 and 5.6). Interestingly, the carbonyl oxygen of the acetyl capping group at -2 position is within hydrogen bonding distance to the amide nitrogen and side chain hydroxyl of +1 threonine residue (Figure 5.6). This network of hydrogen bonds involving the acetyl capping group could stabilize the bound  $3_{10}$ -helical conformation of the Ac-10-mer peptide and can explain why deletion of the acetyl capping group significantly lowers the binding affinity for WDR5 (310). In the MLL1 *Win* motif structure (from which the Ac-10-mer peptide is derived), serine and alanine residues are present at the identical -2 and +1 positions, respectively, (Figures 4.4a and 5.6a) and hence do not have the ability to form the intramolecular bond present in the Ac-10-mer peptide structure. The presence of these additional intramolecular hydrogen bonds at the N-terminus could stabilize the

$3_{10}$ -helical conformation of Ac-10-mer peptide and potentially contribute to its increased binding affinity for WDR5. The MLL3 *Win* motif peptide structure also reveal an intramolecular hydrogen bond between the side chains of -2 and +1 residues (Figure 5.6b); however, we hypothesize that the structure of the  $3_{10}$ -helix is more closed in the Ac-10-mer structure due to the presence of a smaller residue (acetyl cap) at the -2 position, which in the MLL3 peptide is occupied by a cysteine. In addition, the methyl moiety of the acetyl capping group sits in a hydrophobic pocket formed by A65 and I90 and makes more favorable hydrophobic interactions as compared to that of a cysteine in the identical position in the MLL3 peptide structure. Interestingly, the MLL4 *Win* motif peptide has an alanine at the -2 position that could mimic the methyl moiety of the acetyl capping group. The presence of a hydrophobic residue at the -2 position contributes 3-fold to the observed differences in affinity between MLL1 (serine) and MLL4 (alanine) *Win* motif peptides that have either a hydrophilic or hydrophobic residue at this position. These novel features observed in the N-terminus of Ac-10-mer peptide could contribute to its 10-fold better affinity as compared to MLL3.

The sequence of the Ac-10-mer peptide is identical to the sequence of MLL1 *Win* motif peptide in the residue positions that are C-terminal to the conserved *Win* motif (+3 to +8) (Figure 5.6c), which are expected to participate in similar sets of interactions compared to that of the MLL1 *Win* motif peptide. Similar to that of the MLL1 *Win* motif peptide structure (Figures 4.4a and 5.6c), the +3 (valine) and +4 (histidine) residues of the Ac-10-mer peptide participate in a number of interactions with WDR5 (Figure 5.4 and 5.5). For example, the +4 histidine in the Ac-10-mer structure is located in the A- pocket (Figure 5.4) similar to that MLL1 *Win* motif peptide and participates in van der Waals and hydrophobic contacts with WDR5 residues F149, Y191,



and P173. It is likely that the structural similarity at the C-terminus and the presence of additional hydrogen bonds at the N-terminus of Ac-10-mer peptide contributes to the ~460-fold increase in affinity compared to that of the MLL1 *Win* motif peptide.

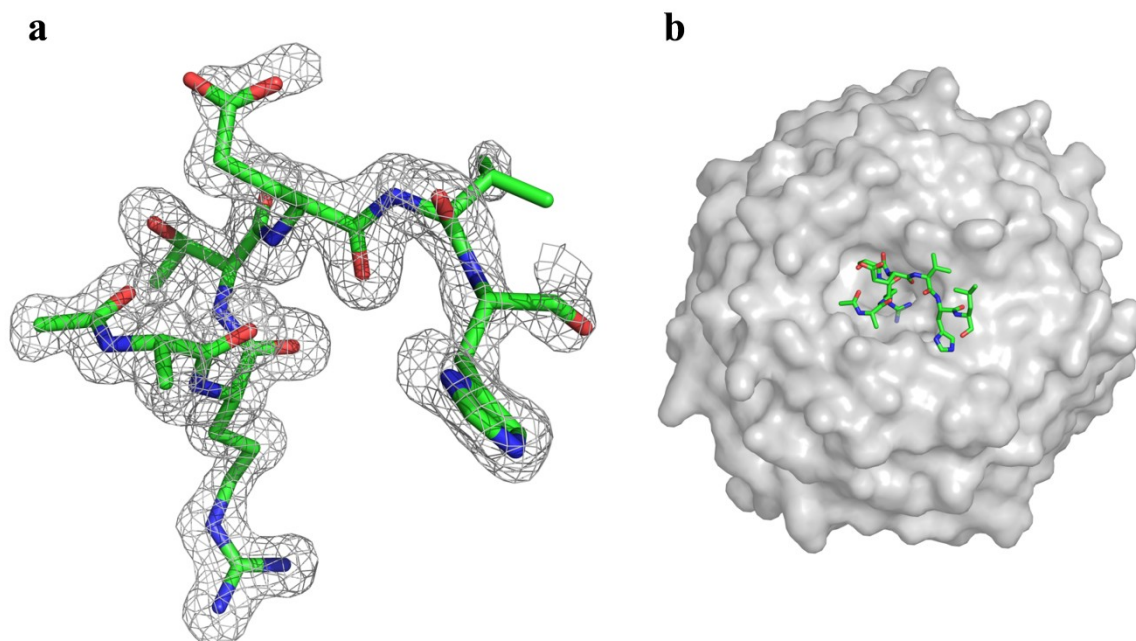
	<b>Ac-10mer</b>	<b>Peptide 3</b>	<b>Peptide 6</b>
Sequence	${}_{\text{Ac}}\text{ARTEVHLRKS}_{\text{NH}_2}$	${}_{\text{Acc}}\text{GARTEVPFLTIF}_{\text{NH}_2}$	${}_{\text{Acc}}\text{SARVAIEYNTAR}_{\text{NH}_2}$
Space group	C222 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C222 <sub>1</sub>
Cell dimensions a, b, c (Å°)	78.413 , 98.888, 80.169	48.756, 67.408, 91.033	78.507, 98.584, 80.175
Angles $\alpha, \beta, \gamma$ (°)	90, 90 90	90, 90, 90	90, 90, 90
Resolution (Å°) <sup>a</sup>	50-1.60(1.63-1.60)	50-1.40 (1.42-1.40 )	50-1.40(1.42-1.4)
Redundancy <sup>a</sup>	14.3 (11.4)	12.4 (5.2)	12.6 (5.6)
Completeness (%) <sup>a</sup>	100 (100)	99 (89.4)	98.2 (83.1)
I/ $\sigma$ <sup>a</sup>	56.03(9.87)	48.52 (2.6)	56.54 (4.89)
R <sub>merge</sub> <sup>a</sup>	11.4 (42.7)	10.3 (72.1)	10.9 (41)

<sup>a</sup> Values in parentheses are for the highest resolution shell.

**Table 5.2: Summary of the X-ray data collection statistics for Ac-10-mer, Peptide-3 and Peptide-6 structures**

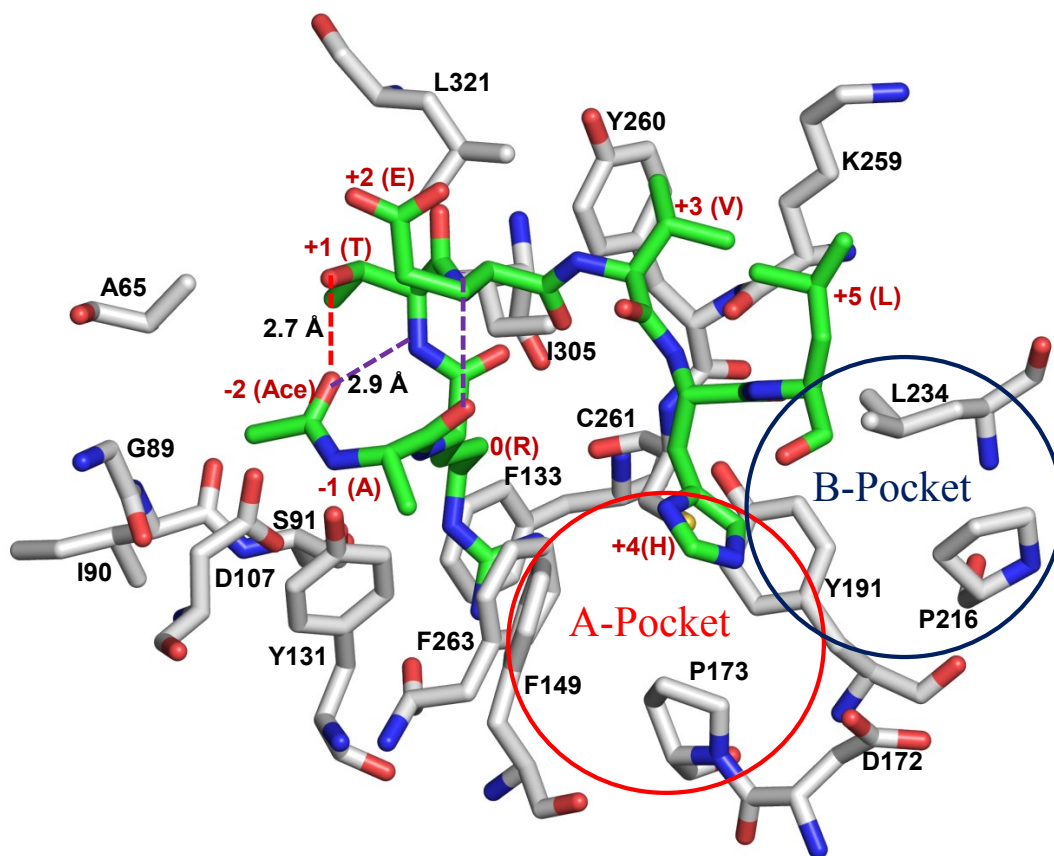
	<b>Ac-10-mer</b>	<b>Peptide 3</b>	<b>Peptide 6</b>
Number of molecules in the asymmetric unit	1	1	1
Resolution (Å°)	1.6	1.40	1.40
No.reflections	29227	59085	58901
R <sub>work</sub> / R <sub>free</sub>	19.39/21.83	16.68/19.37	20.46/22.23
<b>No.atoms</b>			
Protein	2356	2356	2356
Peptide	60	94	68
Water	213	510	338
<b>B-factors</b>			
Protein	12.91	14.38	12.55
Peptide	20.41	14.03	20.97
Water	20.65	26.39	24.68
<b>root mean square (r.m.s) deviations</b>			
Bond lengths (Å°)	0.005	0.006	0.005
Bond angles (°)	1.49	1.193	1.52
<b>Ramachandran plot</b>			
% most favored	87.5	96.47	86.8
% allowed	12.5	3.53	13.2
% outliers	0.0	0.0	0.0

**Table 5.3: Summary of the refinement statistics for Ac-10-mer, Peptide-3 and Peptide-6 structures**



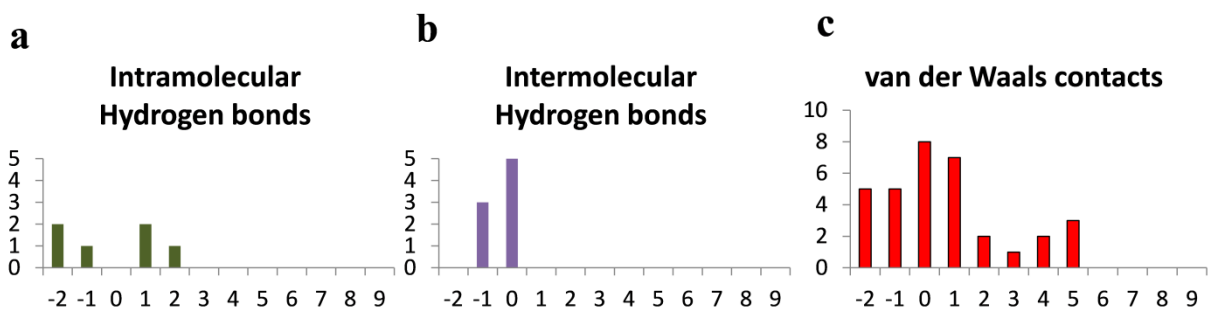
### Figure 5.3 Crystal structure of Ac-10-mer peptide bound to WDR5

a-b. Structure of Ac-10-mer peptide in complex with WDR5. In a, Simulated annealing (SA) omit map contoured at 3  $\sigma$  showing the Ac-10-mer peptide bound to WDR5. The peptide is shown as sticks and the map is shown as grey mesh. b, Ac-10-mer peptide binds at the top of WDR5 using a  $3_{10}$  helical conformation. Ac-10-mer peptide is found in the same arginine-binding pocket that was shown to be occupied by Human SET1 family Win motif peptides. WDR5 molecule is shown as a surface representation (grey) and the Ac-10-mer peptide is shown as sticks (green). Electron density for the last four C-terminal residues (LRKS) were ambiguous and hence these residues were not modeled in the final structure.



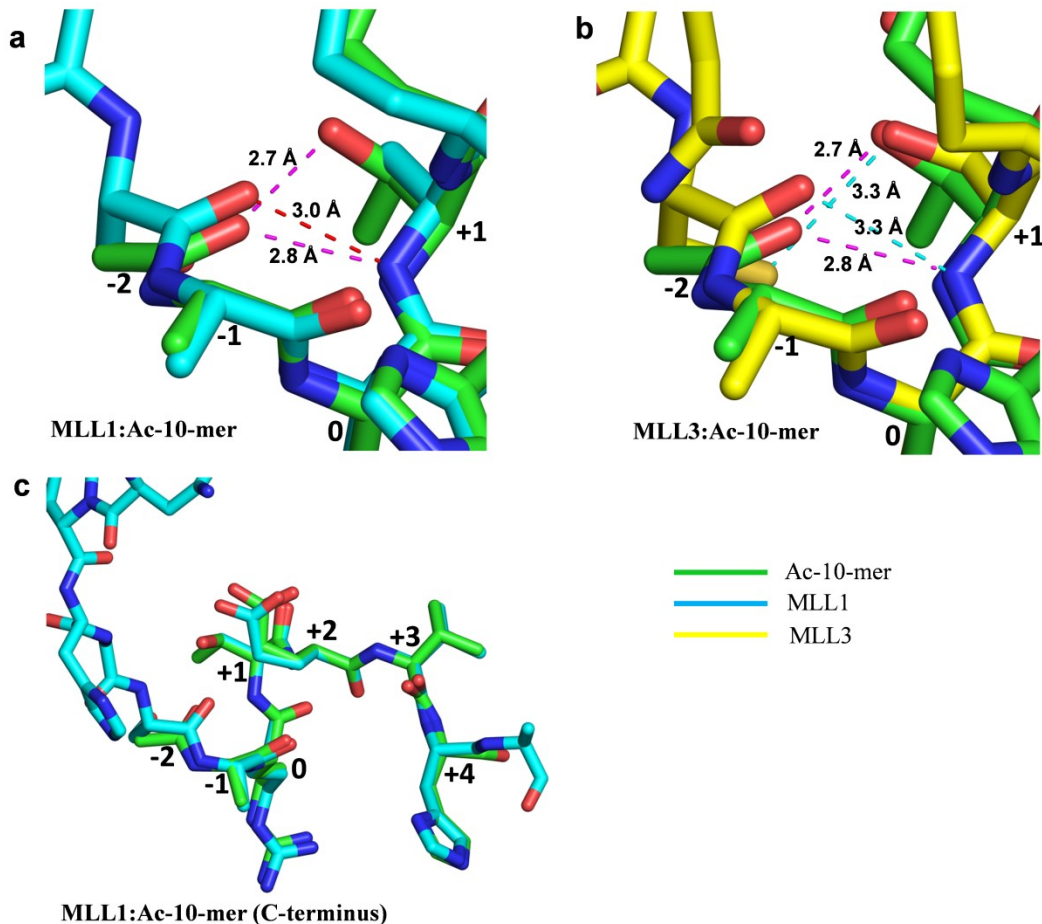
**Figure 5.4** Crystal structure of Ac-10-mer peptide bound to WDR5

The amino acid residues that form the interaction surface between WDR5 and Ac-10-mer peptide are shown. WDR5 residues are shown as white sticks and the Ac-10-mer peptide residues are shown as sticks in green. Intramolecular  $i \rightarrow i+3$  (purple dotted lines) and intermolecular (red dotted lines) hydrogen bonds present within the Ac-10-mer peptide structure are shown with the bond distances in Å. A (red)- and B (blue)-hydrophobic pockets present in WDR5 are circled. The +4 histidine residue of Ac-10-mer peptide is found in the A-pocket. WDR5 residues are labeled in black and Ac-10-mer peptide residues are labeled in red.



**Figure 5.5 Summary of interactions observed in Ac-10-mer peptide:WDR5 co-crystal structure**

**a-c.** Summary of number of inter and intramolecular hydrogen bonds and van der Waals contacts present in Ac-10-mer peptide structure. The interactions are listed according to the residue position of the peptide with the conserved *Win* motif arginine residue designated as position 0. CCP4i suite was used to analyze inter- and intramolecular hydrogen bonds (a and b) using a cutoff distance range of 2.5-3.25 Å and van der Waals contacts (c) using a cutoff distance range of 3.3-4.0 Å. Refer to text to more detail.



**Figure 5.6 Structural comparisons between Ac-10-mer and MLL1/MLL3 *Win* motif peptides**

a-c. Structural analyses of the similarities and differences between Ac-10-mer and MLL1 or MLL3 *Win* motif peptides. MLL1 (cyan), MLL3 (yellow) and Ac-10-mer (green) peptides are shown as stick representations. a, N-terminal structural differences between MLL1 and Ac-10-mer peptides showing the intramolecular hydrogen bonding network. The hydrogen bonding network of Ac-10-mer peptide is indicated in pink and that of the MLL1 *Win* motif peptide is indicated in red. b, N-terminal structural differences between MLL3 and Ac-10-mer peptides showing the intramolecular hydrogen bonding network. The hydrogen bonding network of Ac-10-mer peptide is indicated in pink and that of the MLL3 *Win* motif peptide is indicated in cyan. c, Structural superposition of the C-termini of MLL1 *Win* motif and Ac-10-mer peptides. C-termini of MLL1 and Ac-10-mer peptides are highly similar.

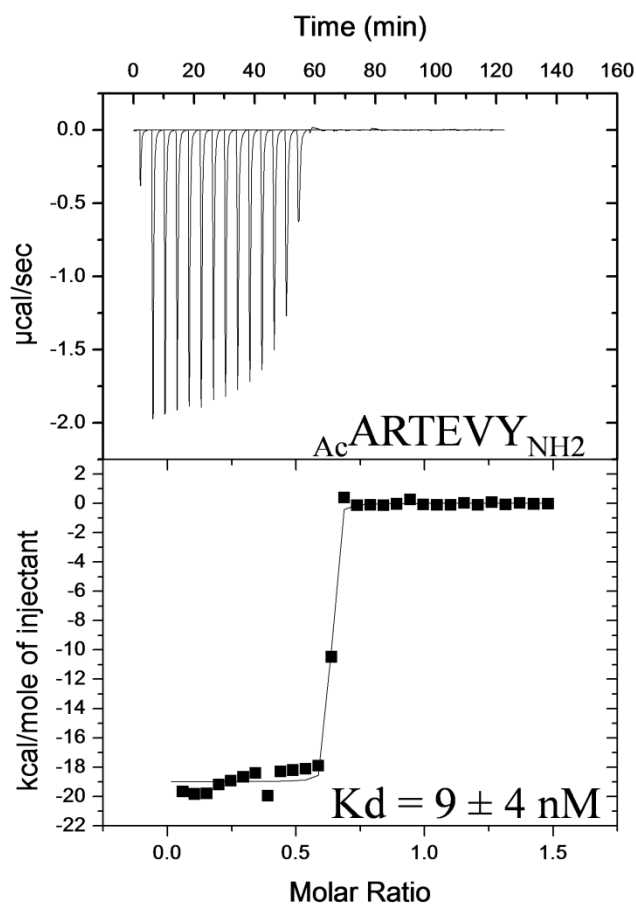
### 5.1.3 Structure based design of a six residue *Win* motif peptide

The crystal structure of Ac-10-mer peptide reveals that the presence of acetyl capping group (at -2 position) and a hydroxyl side chain (threonine in Ac-10-mer and serine in MLL2/MLL3/SETd1b) could significantly enhance WDR5 binding affinity. Moreover, the crystal structure of MLL4 *Win* motif peptide also shows that a tyrosine residue at the +4 position forms a direct hydrogen bond with the carboxylate side chain of WDR5 residue D172. Thermodynamic analysis suggests that this hydrogen bond contributes to ~9-fold observed differences in affinities between MLL4 and MLL1 (has a histidine at +4 position) *Win* motif peptides. Based on these crystal structures, we hypothesized that combining these structural features could further enhance the binding affinity of *Win* motif-based peptides to WDR5. To test this hypothesis, we designed a six residue *Win* motif based peptide ( $_{\text{Ac}}\text{ARTEVY}_{\text{NH}_2}$ ) that has an acetyl capping group at -2, threonine at +1 and a tyrosine at +4, and measured its affinity for WDR5 using ITC.

The ITC data shows that the six residue *Win* motif peptide binds WDR5 with a dissociation constant of 9 nM (Figure 5.7), which is similar to that of the Ac-10-mer peptide (Table 5.1). While this binding affinity is 6-fold and 300-fold better than the MLL3 and MLL1 *Win* motif peptides, respectively; it is similar in binding affinity of that of the Ac-10-mer peptide. These results suggest that the six-residue peptide captures the majority of interactions present in the Ac-10-mer peptide and that the presence of tyrosine at the +4 position does not contribute to the increased affinity. However, it should be noted that the Ac-10-mer peptide has four additional residues (LRKS) on the C-terminus compared to the six residue peptide. It is possible that these additional residues contribute to increased WDR5 affinity and hence could negate the



positive contribution of having a tyrosine at +4 position in the six residue peptide. Indeed, a previous study showed that deletion of lysine and serine residues (+7 and +8) at the C-terminus of MLL1 *Win* motif peptide results in a 2-fold decrease in WDR5 binding affinity (310). One other possibility for the similar interaction affinities observed for both the Ac-10-mer and the six-residue *Win* motif peptide is that in the absence of additional residues on the C-terminus, the +4 tyrosine in the six residue peptide may not adopt a conformation similar to the MLL4 peptide structure. Three dimensional structure of six residue peptide bound to WDR5 is required to distinguish these possibilities. It is important to note that the binding affinities reported for the Ac-10-mer and six-residue peptides are very close to the lower limits ( $K_a = 10^9$  to  $10^{10}$ ) (334) of the dynamic range of measurements possible with ITC. Therefore, it is possible that the similar dissociation constants reported here for Ac-10-mer and six-residue peptides are biased due to experimental limitations of ITC and the binding affinities could be different for the two peptides. Nevertheless, based on these studies the six-residue *Win* motif peptide was identified as the smallest length peptide that can recapitulate majority of interactions required for high affinity binding to WDR5. Hence, the six residue peptide represents an excellent starting point for the design of additional small molecule inhibitors that target MLL1-WDR5 interaction.



**Figure 5.7: Six-residue *Win* motif-based peptide ( $\text{AcARTEVYNH}_2$ ) binds WDR5 with low nanomolar affinity**

Isothermal Titration Calorimetry (ITC) data for the binding of six-residue *Win* motif-based peptide to WDR5. Upper panels show heat of binding plotted as a function of time. Lower panels show the binding isotherms fit to a one-site binding model. Six-residue *Win* motif-based peptide sequence and the dissociation constants ( $K_d \pm \text{S.E.M.}$ ) derived from the fit are indicated. S.E.M values reported are derived from two independent experiments.

## ***5.2 Ongoing efforts to identify of additional Win-motif like inhibitors with better inhibitory properties:***

### ***5.2.1 Approach 1: Phage display screen (in collaboration with Dr. Ziwei Huang's group at Upstate Cancer Institute)***

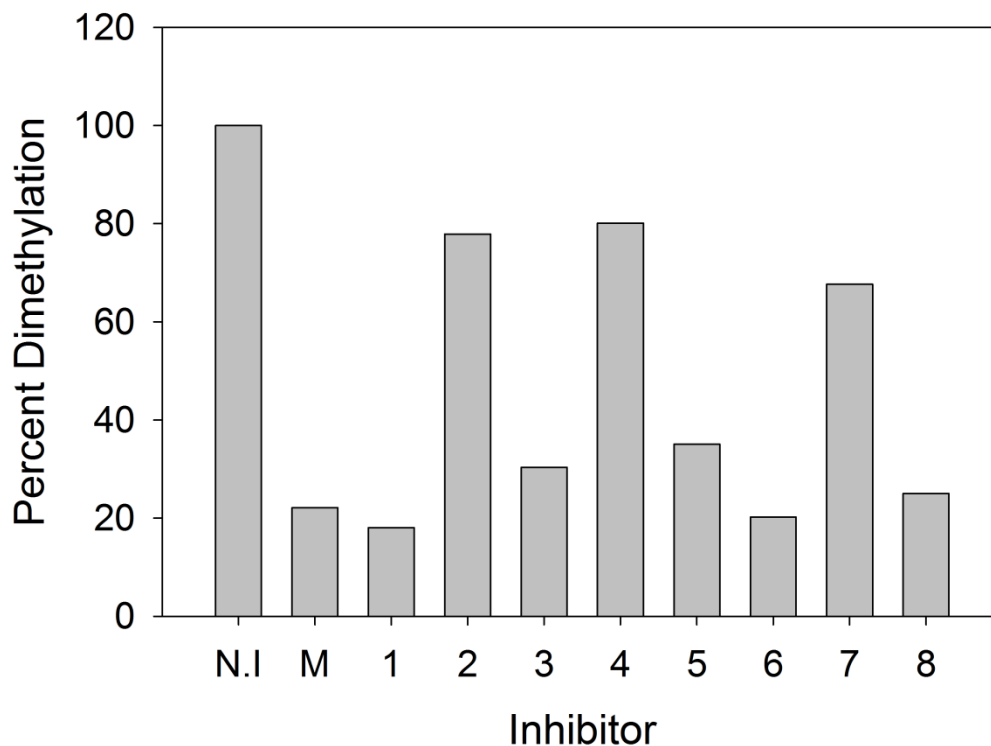
Binding and inhibition studies reported for the Ac-10-mer and six-residue peptides suggest that these inhibitors bind WDR5 with low nanomolar affinities and the Ac-10-mer peptide is a highly potent inhibitor of MLL1 core complex dependent H3K4 dimethylation activity. The efficacy of using these *Win* motif-based peptides as potential inhibitors *in vivo* could be greatly enhanced by designing peptides or small molecule compounds that inhibit the MLL1 core complex activity with more stringent efficiencies. To facilitate the design of better *Win* motif inhibitors, we utilized a random group of peptides derived from a phage display screen (Ph. D<sup>TM</sup> screen New England BioLabs). The premade libraries (~10<sup>9</sup> clones) consists of a mixture of linear (heptapeptide, Ph.D.-7 and dodecapeptide, Ph.D.-12) and cyclic (heptapeptide, Ph.D.-C7C) peptides whose sequences are completely randomized. The randomized segment of the cyclic peptides is flanked by a pair of cysteine residues, which are oxidized during phage assembly to form a disulfide linkage, resulting in the displayed peptides being presented to the target as loops. We then used this random library of phage display peptides and tested their ability to bind the target protein WDR5. Using this approach, we (Yan Xu in Dr. Huang's Lab) identified eight peptides (numbered 1-8) of random length (7-12 residues) and sequence (refer to Table 5.4) that bound to the target protein using competition based fluorescence polarization anisotropy experiments (FPA experiments were carried out by Yan Xu in Dr. Huang's laboratory). Interestingly, the sequence of all eight peptides identified from this random library screen contains an arginine residue within a sequence context similar to that of *Win* motif (Table 5.4).

Peptide number	Sequence	IC <sub>50</sub> , μM
	-5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10	
1	S V R T D P A M H W S R	1
2	L R T D L V H N S L V Q	>100
3	G A R T E V P F L T I F	3
4	S A A S R T I D M W A T	>100
5	D R T T H Y Y L G K L E	1.4
6	S A R V A I E Y N T A R	2
7	I S A S M R T T M N N S	>100
8 (cyclic peptide)	C R T A P M P H C	7

Note: IC<sub>50</sub> values were determined using fluorescence polarization anisotropy

**Table 5.4 Sequence and IC<sub>50</sub> values for peptides identified using Phage display screen**

Using this bank of peptides, we tested these peptides for their ability to inhibit the H3K4 dimethylation activity of the MLL1 core complex using our MALDI-TOF mass spectrometry based methylation assays. The results show that in the presence of peptides-1, 3, 5, 6, and 8, the dimethylation activity of the MLL1 core complex is decreased 5-fold (~20% remaining) (Figure 5.8) compared to that observed in the absence of any inhibitor (normalized to 100%). The amount of inhibition observed with peptides-1, 3, 5, 6, and 8 is similar to that of MLL3 *Win* motif peptide (Figure 5.8). On the contrary, peptides- 2, 4, and 7 did not inhibit the H3K4 dimethylation activity to similar levels (Figure 5.8). These results are similar to the  $IC_{50}$  values determined from the fluorescence polarization experiments (Table 5.4). Furthermore, these inhibition studies suggest that peptides -1, 3, 5, 6, and 8 would bind WDR5 with similar affinities compared to the MLL3 *Win* motif peptide. Since the sequences of the peptides are quite different, analysis of three-dimensional structures of these peptides- (1, 3, 5, 6 and 8) could provide useful structural information that could facilitate the process of structure-based drug design.



**Figure 5.8 MALDI-TOF based inhibition data for the peptides identified from phage display screen**

H3K4 dimethylation inhibition data for peptides identified from phage display screen. MALDI-TOF assays were carried out as described in 3.6. H3K4 methylation assays were carried out in the absence (N.I=No inhibitor) or presence of 250  $\mu$ M of MLL3 (M) *Win* motif or other peptides (1-8) derived from phage display. Relative H3K4 dimethylation levels in the presence of different peptides are shown. Percent dimethylation level in the absence of any inhibitor (N.I) is normalized to 100%. The data presented is from a single experiment. Compounds 1, 3, 5, 6, and 8 are as good as MLL3 *Win* motif peptide in inhibiting the H3K4 dimethylation activity of MLL1 core complex.

### ***5.2.2 Structural basis for the interaction between WDR5 and phage display library derived peptides***

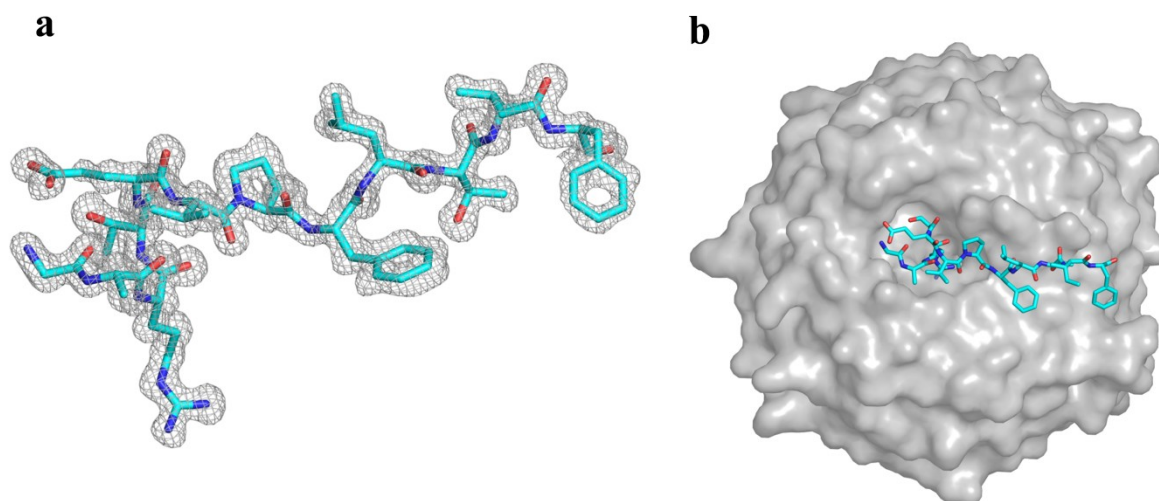
Based on the results on the inhibition studies described earlier, we attempted to determine the crystal structures of WDR5 in complex with peptides- 1, 3, 5, 6, and 8, respectively. Crystal structures of WDR5:Peptide-3 and WDR5:Peptide-6 binary complexes were determined at 1.4 Å resolution. However, crystallization trials with peptides- 1, 5 and 8 did not produce diffraction quality crystals.

#### ***Crystal structure of WDR5-Peptide3 binary complex***

To understand the molecular basis for the inhibitory efficiency of Peptide-3, we determined the x-ray structure of Peptide-3 (GARTEVPFLTIF) bound to WDR5 at 1.4 Å resolution. The x-ray data collection and refinement statistics are summarized in Tables 5.2 and 5.3. All 12 residues could be modeled unambiguously into the electron density map (Figure 5.9a). The structure reveals that Peptide-3 binds WDR5 using a  $3_{10}$ -helical conformation (Figure 5.9b) similar to that of Ac-10-mer (Figure 5.3b) or human SET1 family *Win* motif peptides (Figure 4.8a). Similar to the other *Win* motif peptide structures, the side chain of the arginine (at 0 position) is inserted into the central water-filled tunnel in WDR5 and is stabilized by an extensive network of hydrogen bonds, van der Waals interactions and hydrophobic contacts (Figures 5.9, 5.10 and 5.11). The  $3_{10}$ -helical structure is stabilized by two intramolecular *i* to *i*+3 hydrogen bonds between the -2 (glycine carbonyl) and +1 (threonine amide), and -1 (alanine carbonyl) and +2 (glutamate amide) residues, respectively (Figure 5.10). In addition, the carbonyl oxygen of the glycine residue at -2 position is within hydrogen bonding distance to the side chain hydroxyl of the +1 threonine residue. Moreover, the carbonyl oxygen of -1 alanine is also within hydrogen bonding distance to the amide nitrogen of the +3 valine (Figure 5.10). The presence of these

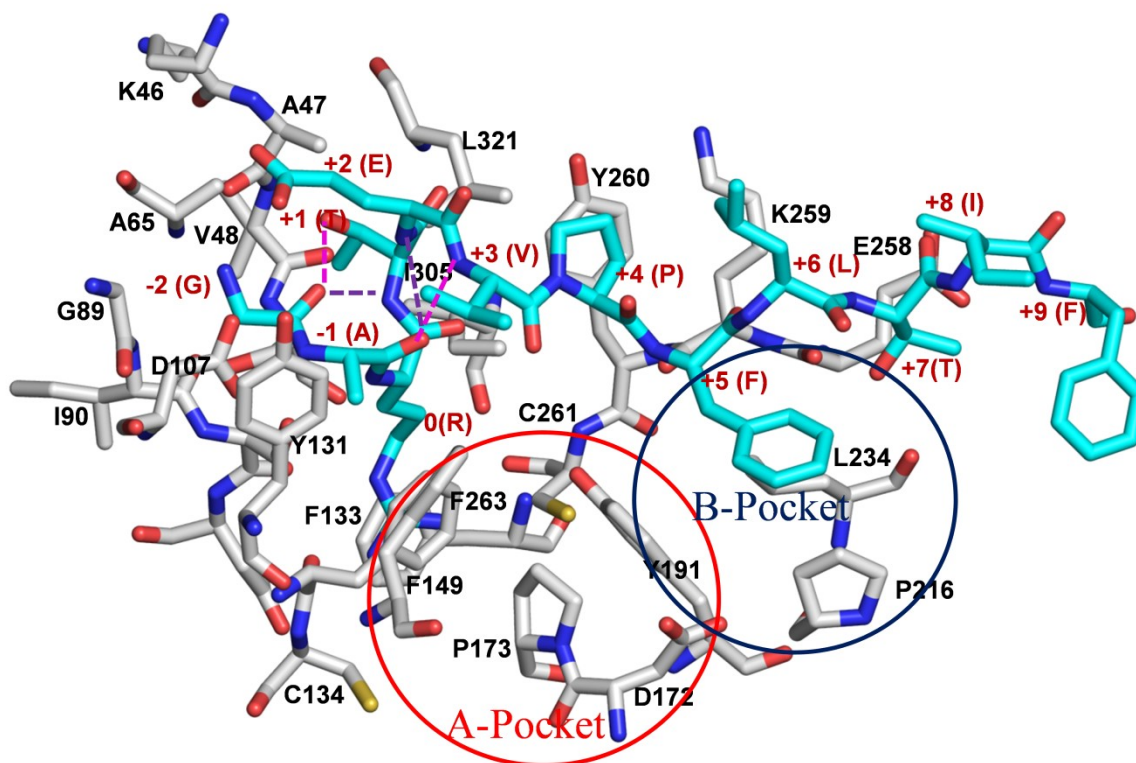
additional intramolecular hydrogen bonds involving residues -2 to +3 could lock the peptide in a favorable  $3_{10}$ -helical conformation and potentially contribute to its increased binding affinity for WDR5.





**Figure 5.9 Crystal structure of Peptide-3 bound to WDR5**

a-b. Structure of Peptide-3 (GARTEVPFLTIF) in complex with WDR5. In a, Simulated annealing (SA) omit map contoured at  $3\sigma$  showing peptide 3 bound to WDR5. Peptide-3 is shown as sticks (blue) and the map is shown as grey mesh. All 12 residues could be modeled into the electron density for the peptide. b, Peptide-3 binds at the top of WDR5 using a  $3_{10}$  helical conformation. Peptide-3 is found in the same arginine-binding pocket. WDR5 molecule is shown as a surface representation (grey) and the peptide is shown as sticks (blue).

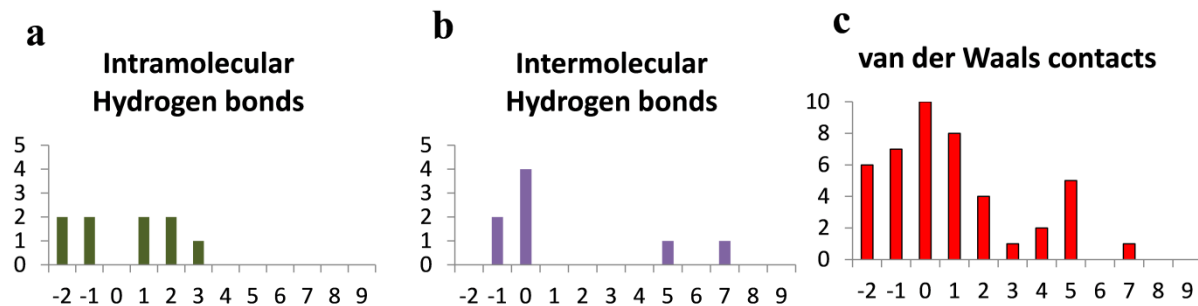


**Figure 5.10** Crystal structure of Peptide-3 bound to WDR5

The amino acid residues that form the interaction surface between WDR5 and Peptide-3 are shown. WDR5 residues are shown as white sticks and Peptide-3 residues are shown as sticks in blue. Intramolecular hydrogen bonds present in the Peptide-3 structure are shown as purple (I to i+3) and pink dotted lines. A (red)- and B (blue)-hydrophobic pockets present in WDR5 are circled. The +5 phenylalanine residue of Peptide-3 is found in the B-pocket. WDR5 residues are labeled in black and the peptide residues are labeled in red.

Amino acid residues C-terminal to the arginine (+2 to +9) display significant structural heterogeneity in the Peptide-3 structure as compared to Ac-10-mer or other human SET1 family *Win* motif peptide structures. A major difference in the Peptide-3 structure is seen at +3, +4 and +5 positions. In the Peptide-3 structure, the +3 and +4 positions are occupied by valine and proline, respectively. The presence of proline in +4 position reorients the side chain of +3 valine in the opposite direction compared to that of the MLL1 and MLL4 *Win* motif structures each of which have a valine at +3 position (Figures 4.4a, 4.4d and 5.10). The presence of an intramolecular hydrogen bond with the carbonyl oxygen of alanine at the -1 position also helps to reorient the valine side chain away from WDR5 residue Y260 (Figure 5.10). These features enable the +4 proline ring from the Peptide-3 structure to make extensive van der Waals contacts with the aromatic ring of Y260 in WDR5 (Figure 5.10). MLL2 and MLL3 *Win* motif peptides also have a proline, but in the +3 position and engage in similar van der Waals interactions with Y260. The next residue in the Peptide-3 sequence, phenylalanine at the +5 position, is found in the B-pocket in WDR5 and engages in van der Waals and hydrophobic contacts with WDR5 residues L234, P216 and Y191 (Figure 5.10). In addition, the amide nitrogen of the +5 phenylalanine residue also forms a direct hydrogen bond with the main chain carbonyl oxygen of WDR5 residue K259. The interactions observed for the +5 phenylalanine residue are similar to the interactions observed for the +4 lysine residue in MLL2 and MLL3 *Win* motif peptide structures, which also bind to the B-pocket. The presence of this intermolecular hydrogen bond could explain why both MLL3 and Peptide-3 have similar  $IC_{50}$  values in WDR5 binding assays (Table 5.4). One other observed difference in the Peptide-3 structure is that the carbonyl oxygen of the threonine residue at +7 position is within hydrogen bonding distance to the carboxylate side chain of WDR5 residue E258. A similar hydrogen bond could be present in the all peptides

that have a +7 residue, however, the +7 residue is disordered in most of the crystal structures and hence could not be modeled.

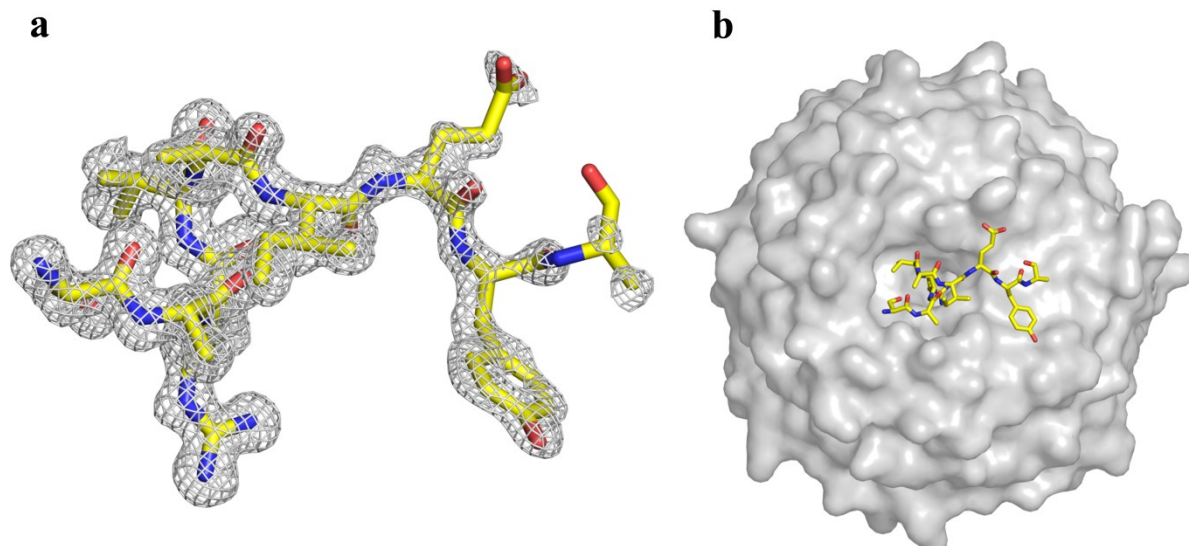


**Figure 5.11: Summary of interactions observed in Peptide-3: WDR5 co-crystal structure**

**a-c.** Summary of number of inter and intramolecular hydrogen bonds and van der Waals contacts present in Peptide- 3:WDR5 co-crystal structure. The interactions are listed according to the residue position of the peptide with the arginine residue designated as position 0. CCP4i suite was used to analyze inter- and intramolecular hydrogen bonds (a and b) using a cutoff distance range of 2.5-3.25 Å and van der Waals contacts (c) using a cutoff distance range of 3.3-4.0 Å. Refer to text to more detail.

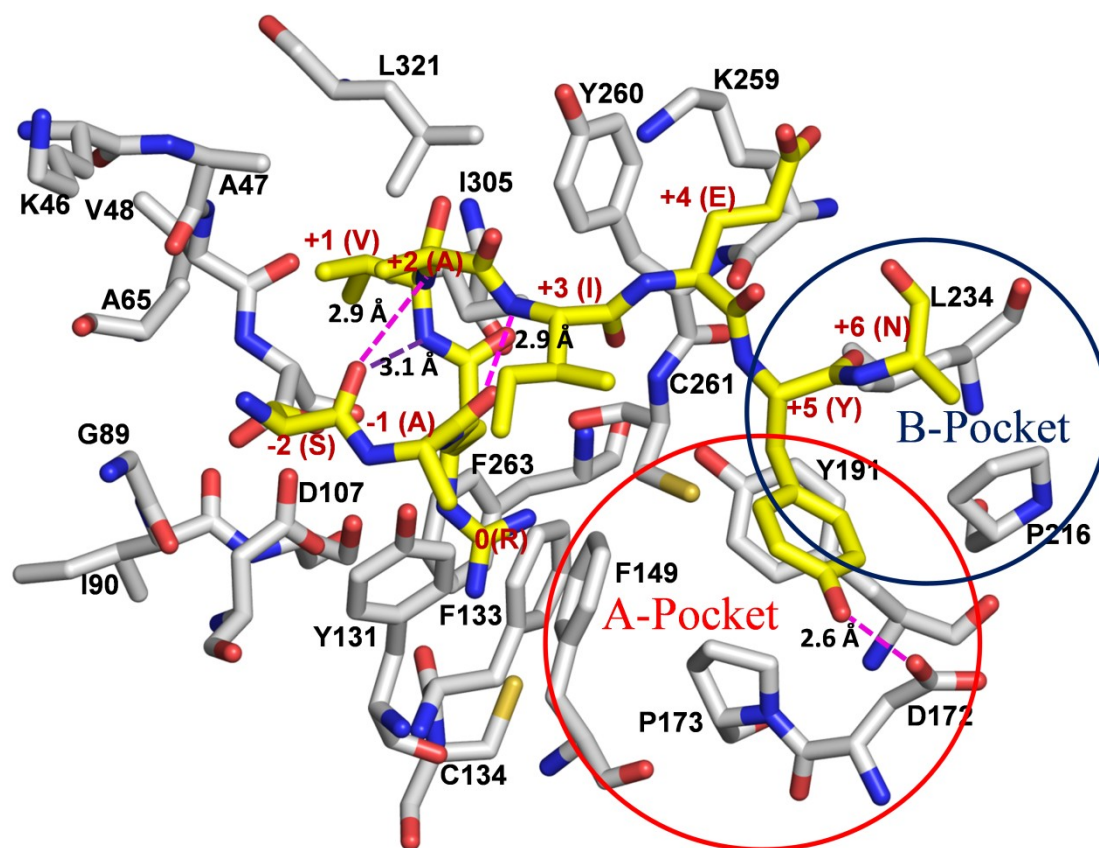
### ***Crystal structure of WDR5:Peptide-6 binary complex***

We next determined the x-ray three-dimensional structure of Peptide-6 (SARVAIEYNTAR) bound to WDR5 at 1.4 Å resolution. The x-ray data collection and refinement statistics are summarized in Tables 5.2 and 5.3. Eight residues could be modeled unambiguously into the electron density map with the last four C-terminal residues being disordered (Figure 5.12a). The structure reveals that Peptide-6 binds WDR5 using a  $3_{10}$ -helical conformation (Figure 5.12b) similar to that of Ac-10-mer (Figure 5.3b), Peptide-3 (Figure 5.9b) or human SET1 family *Win* motif peptides (Figure 4.8a). Similar to the other peptide structures reported in this study, the side chain of the arginine (at 0 position) is inserted into the central water-filled tunnel in WDR5 and is stabilized by an extensive network of hydrogen bonds, van der Waals interactions and hydrophobic contacts (Figure 5.14). The  $3_{10}$ -helical structure is stabilized by an intramolecular  $i$  to  $i+3$  hydrogen bonds between the -2 (serine carbonyl) and +1 (valine amide residues (Figure 5.13). In addition, the main chain carbonyl oxygen of the serine residue at the -2 position is within hydrogen bonding distance to the main chain amide of +2 alanine residue (Figure 5.13). Moreover, the carbonyl oxygen of alanine at the -1 position also forms a direct intramolecular hydrogen bond with the amide nitrogen of +3 valine residue (Figure 5.13). These interactions at the N-terminus are quite similar to that of the Peptide-3 structure.



**Figure 5.12 Crystal structure of Peptide-6 bound to WDR5**

a-b. Structure of Peptide-6 (SARVAIEYNTAR) in complex with WDR5. In a, Simulated annealing (SA) omit map contoured at 3  $\sigma$  showing the Peptide-6 bound to WDR5. Peptide-6 is shown as sticks (yellow) and the map is shown as grey mesh. Only the first 9 residues could be modeled into the electron density for the peptide. b, Peptide-6 binds at the top of WDR5 using a 3<sub>10</sub>-helical conformation. Peptide-6 is found in the same arginine-binding pocket. WDR5 molecule is shown as a surface representation (grey) and the peptide is shown as sticks (yellow).



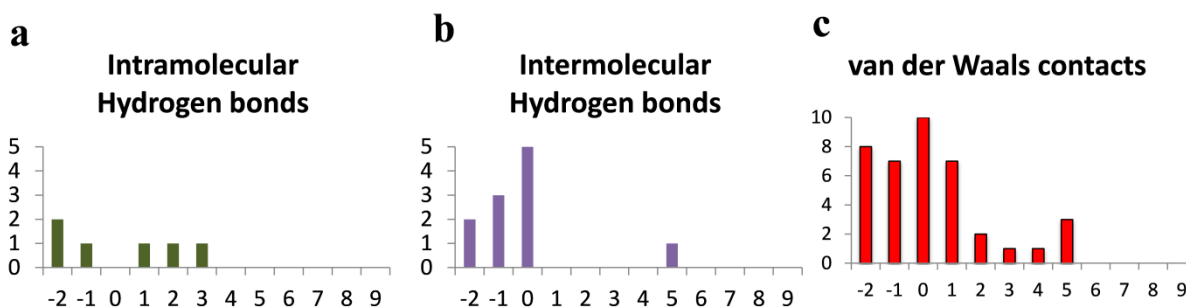
**Figure 5.13** Crystal structure of peptide 6 bound to WDR5

The amino acid residues that form the interaction surface between WDR5 and Peptide-6 are shown. WDR5 residues are shown as white sticks and Peptide-6 residues are shown as sticks in yellow. Intramolecular hydrogen bonds present in the Peptide-6 structure are shown as purple ( $i$  to  $i+3$ ) and pink dotted lines. A (red)- and B (blue)-hydrophobic pockets present in WDR5 are circled. The +5 tyrosine residue of Peptide-6 is found in the A pocket and forms a direct hydrogen bond with WDR5 residue D172. WDR5 residues are labeled in black and the peptide residues are labeled in red.



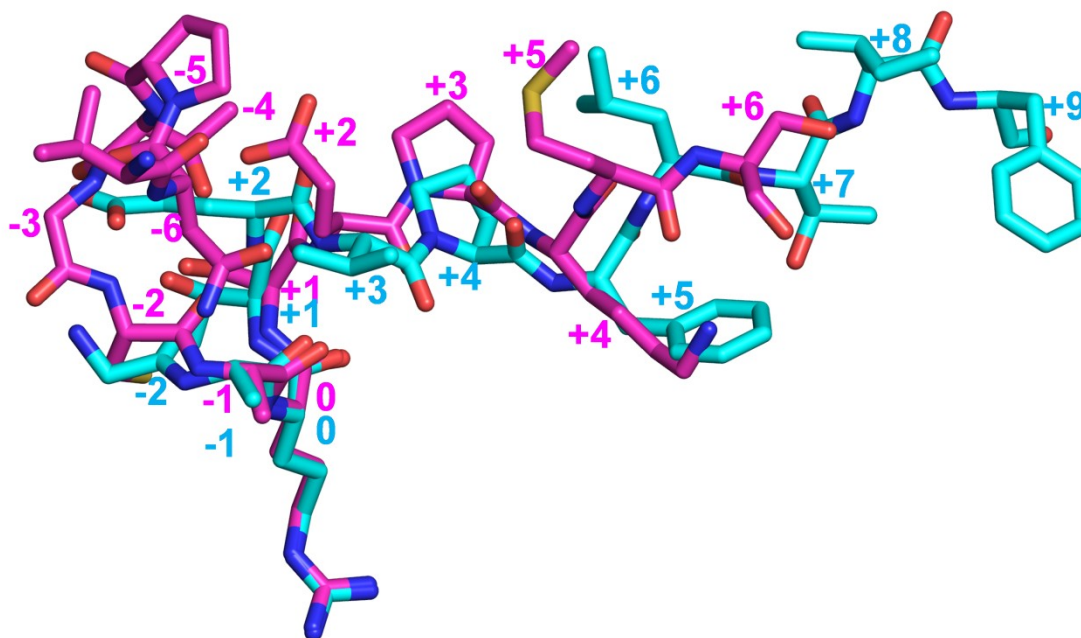
Amino acid residues C-terminal to the arginine (+2 to +5) displays significant structural heterogeneity in the Peptide-6 structure compared to that of Ac-10-mer or other human SET1 family *Win* motif peptide structures. Similar to the Peptide-3 structure, the side chain of the isoleucine residue at the +3 position is oriented away from the aromatic ring of WDR5 residue Y260 (Figure 5.13). However, in contrast to the Peptide-3 structure which has a proline at the +4 position that stacks against Y260, the glutamate residue at the +4 position in the Peptide-6 structure makes extensive van der Waals contacts with the WDR5 residue K259 (Figure 5.11). These interactions are similar to the +3 valine interactions observed in the MLL1 and MLL4 *Win* motif peptides that bind in the A-hydrophobic pocket. Similar to MLL1 and MLL4 *Win* motif structures, the +5 tyrosine residue in the Peptide-6 structure also binds in the A-pocket and engages in van der Waals and hydrophobic contacts with WDR5 residues F149, D172, P173 and Y191 (Figure 5.13). The hydroxyl group of the +5 tyrosine in the Peptide-6 structure forms a direct hydrogen bond with the carboxylate group of WDR5 residue D172 (Figure 5.12), an interaction that is also observed in the MLL4 *Win* motif structure. It is possible that this extra hydrogen bond contributes significantly to the WDR5 binding affinity of Peptide-6, which is expected to be similar to MLL3 and MLL4 *Win* motif peptides.

In summary, the structures of Peptide-3:WDR5 and Peptide-6:WDR5 binary complexes reveal that both peptides recapitulate the majority of salient interactions present in the MLL3 (Figures 5.15 and 5.16) or MLL4 *Win* motif structures through an intermolecular hydrogen bond between the +5 residue and WDR5 residues K259 or D172.



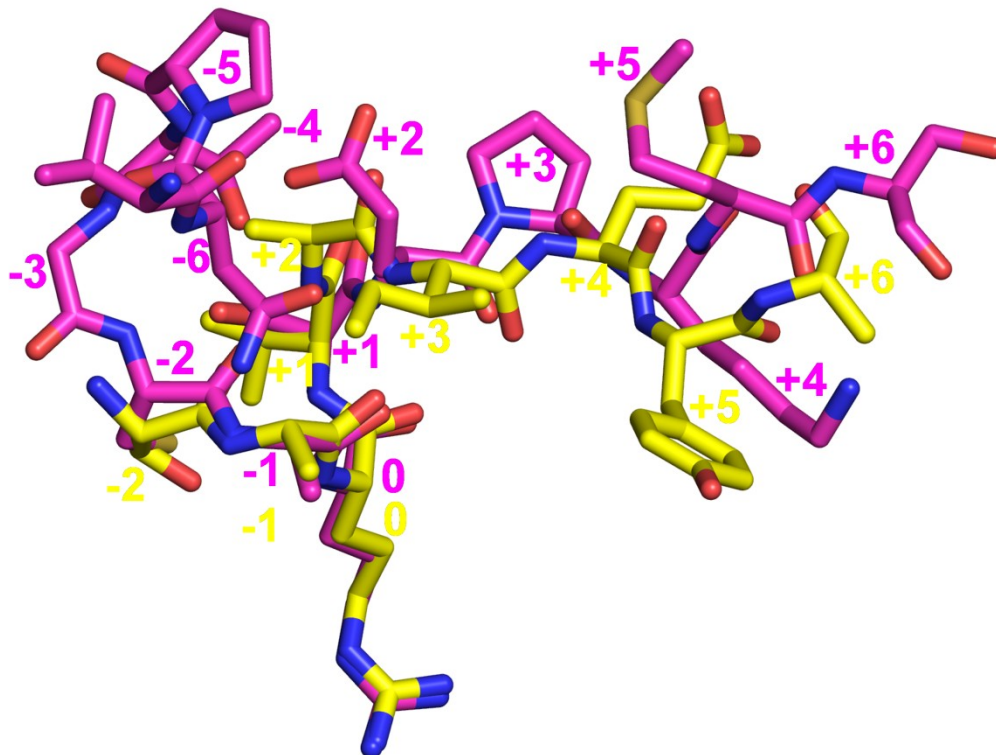
**Figure 5.14 Summary of interactions observed in Peptide-6:WDR5 co-crystal structure**

**a-c.** Summary of number of inter- and intramolecular hydrogen bonds and van der Waals contacts present in the Peptide-6:WDR5 co-crystal structure. The interactions are listed according to the residue position of the peptide with the arginine residue designated as position 0. CCP4i suite was used to analyze inter- and intramolecular hydrogen bonds (a and b) using a cutoff distance range of 2.5-3.25 Å and van der Waals contacts (c) using a cutoff distance range of 3.3-4.0 Å. Refer to text to more detail.



**Figure 5.15: Structural superposition of Peptide-3 and MLL3 *Win* motif peptide structures**

Structural overlay of MLL3 (VNPTGCARSEPKMS) and Peptide-3 (GARTEVPFLTIF) structures highlight the similarities and differences between the two structures. Peptide-3 (Cyan) and MLL3 *Win* motifs (Magenta) are shown as stick representations. The residue positions of the MLL3 *Win* motif are labeled in magenta and that of the Peptide-3 structure are labeled in cyan. MLL3 *Win* motif has additional residues modeled at the N-terminus (-3 to -6) while Peptide-3 has additional residues modeled at the C-terminus (+5 to +9). Amino acid residue at the +5 position in the Peptide-3 structure is present in the B-pocket occupied by the +4 amino acid residue in MLL3 *Win* motif peptide structure.



**Figure 5.16: Structural superposition of Peptide-6 and MLL3 *Win* motif peptide structures**

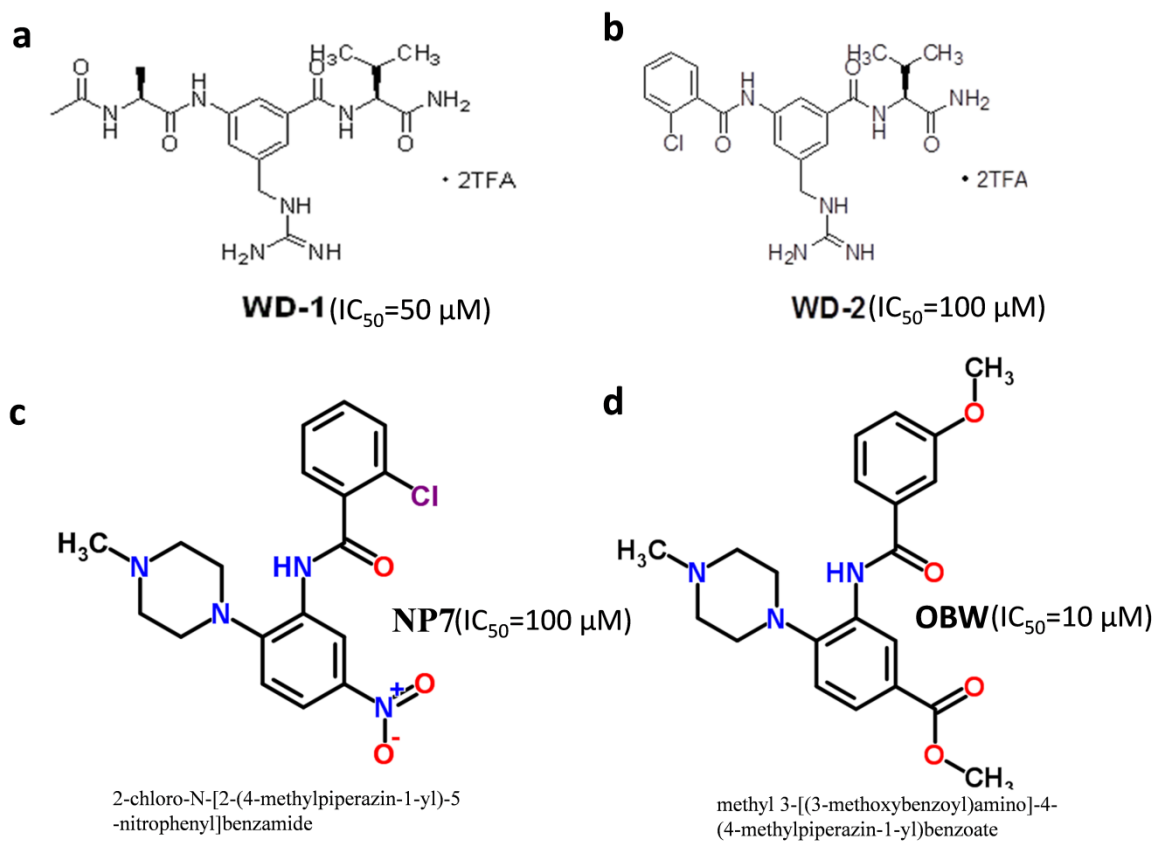
Structural overlay of MLL3 (VNPTGCARSEPKMS) and Peptide-6 (SARVAIEYNTAR) structures highlight the similarities and differences between the two structures. Peptide-6 (Yellow) and MLL3 *Win* motifs (Magenta) are shown as stick representations. The residue positions of the MLL3 *Win* motif are labeled in magenta and that of the Peptide-6 structure are labeled in yellow. MLL3 *Win* motif has additional residues modeled at the N-terminus (-3 to -6) while Peptide-6 has additional residues modeled at the C-terminus (+5 to +6). Amino acid residue at the +5 position in the Peptide-6 structure is present in the A-pocket while the +4 amino acid residue in MLL3 *Win* motif peptide structure is present in the B-pocket of WDR5.

### ***5.2.3 Identification of novel Win motif-like inhibitors using a chemical library screen (This work is done in collaboration with Dr. Huang's lab at Upstate Cancer Institute)***

The binding and inhibition data presented in Chapters 3, 4 and early part of Chapter 5 have shown that *Win* motif peptides or related peptide mimics are excellent candidates that disrupt the interaction between WDR5 and MLL1 and inhibit the H3K4 dimethylation activity of MLL1 core complex. While these peptidomimetics would have enhanced potency against the MLL1 core complex in cells, peptide inhibitors are hard to develop because of poor membrane solubility, large size (few kDa) and lower stability (degradation by proteases in cells). Small-molecule chemical inhibitors (that target a specific interaction such WDR5-*Win* motif) can overcome these drawbacks due to their enhanced membrane solubility and stability. The next section in this dissertation is aimed at screening a library of natural compounds to identify small molecules that can bind WDR5 with increased affinities.

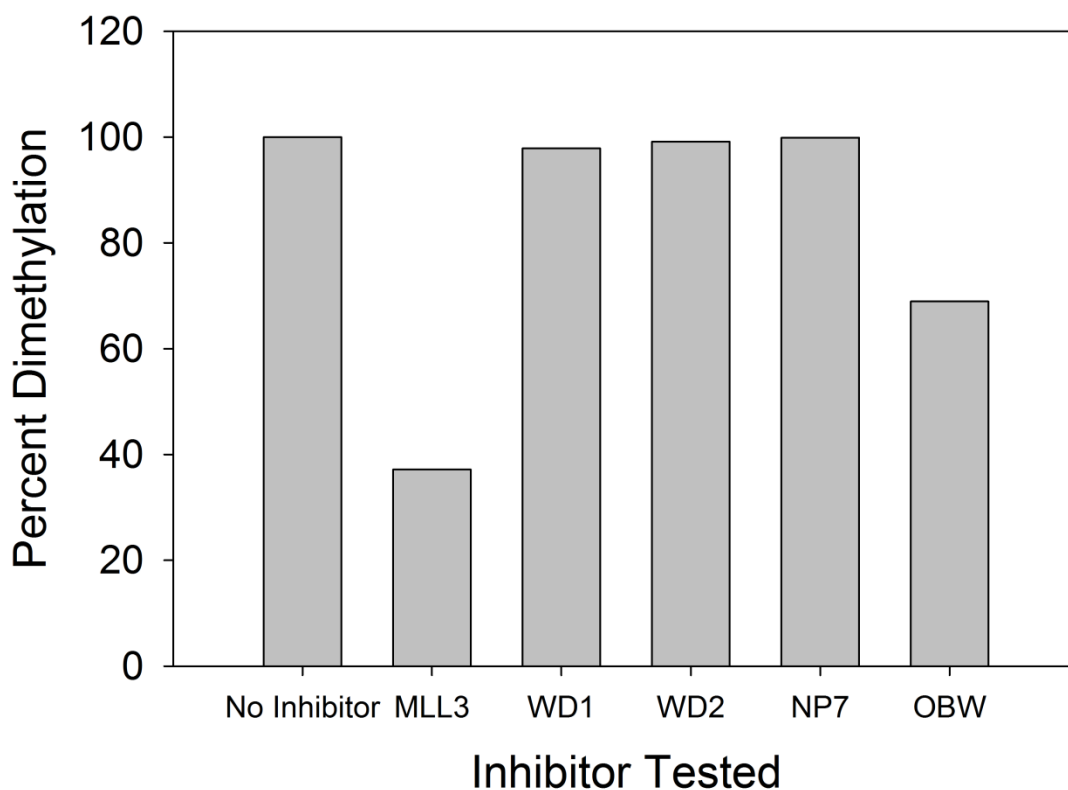
Using an array of chemical compounds obtained from the National Cancer Institute (NCI) database, we sought to identify novel chemical molecules that bind WDR5 with equal or greater affinity than that of the MLL3 *Win* motif peptide. However, our initial screens did not identify any chemical molecules that bind WDR5 with high affinity in competition based fluorescence polarization experiments (work done by Yan Xu in Dr. Huang's lab). Further screening trials led to the identification of two molecules, WD1 and WD2, which bind WDR5 with ~10-fold weaker affinity than the MLL3 *Win* motif peptide (refer to Figure 5.17a and b for the chemical structure). While this work was in progress, the structure of WDR5 in complex with two other chemical compounds NP7 (2-chloro-N-[2-(4-methylpiperazin-1-yl)-5-nitrophenyl]benzamide) and OBW (methyl 3-[(3-methoxybenzoyl)amino]-4-(4-methylpiperazin-1-yl)benzoate) were determined by the structural biology consortium and were included in my analysis.

Using these four chemical compounds (WD1, WD2, NP7 and OBW), we asked if these compounds can inhibit the H3K4 dimethylation activity of MLL1 core complex comparable to the MLL3 *Win* motif peptide. MALDI-TOF based H3K4 methylation assays show that at 30-fold excess (compared to MLL1<sup>3745</sup>), H3K4 dimethylation activity of the MLL1 core complex is inhibited ~80% in the presence of MLL3 *Win* motif peptide. However, little (OBW) to no inhibition was observed for WD1, WD2 and NP7 under the same assay conditions (Figure 5.18). These results suggest that all four chemical compounds are not as good as the MLL3 *Win* motif peptide in inhibiting MLL1 core complex activity under these assay conditions. Further screening trials using a larger library of chemical compounds would help identify small molecules that are highly potent inhibitors of MLL1 core complex.



**Figure 5.17 Chemical structure of compounds identified in virtual screen**

The chemical compounds WD1, WD2, NP7 and OBW were synthesized by Dr. Ziwei Huang's lab as per their lab chemical synthesis protocols and were handed to me.



**Figure 5.18 MALDI-TOF based inhibition assays for the chemical compounds identified from virtual screen**

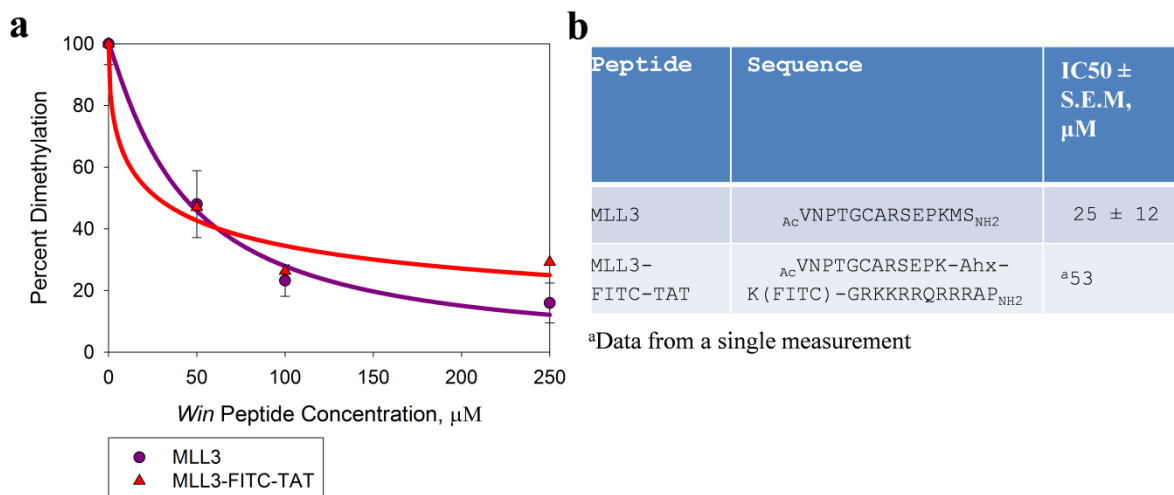
H3K4 dimethylation inhibition data for the chemical compounds identified from virtual screen. MALDI-TOF assays were carried out as described in 3.6. H3K4 methylation assays were carried out in the absence (No inhibitor) or presence of 250  $\mu\text{M}$  of MLL3 *Win* motif or other compounds derived from chemical synthesis. Relative H3K4 dimethylation levels in the presence of different inhibitors are shown. Percent dimethylation level in the absence of any inhibitor (N.I) is normalized to 100%. The data presented is from a single experiment. Compounds WD1, WD2, NP7 and OBW8 are not as good as MLL3 *Win* motif peptide in inhibiting the H3K4 dimethylation activity of MLL1 core complex.



### ***5.3 MLL3 Win motif peptide inhibits the H3K4 methylation activity of MLL1 core complex in cells (this work was done with assistance from a former postdoc Dr. Laila Kabrossy)***

#### ***5.3.1 Design of a fluorescent tagged MLL3 Win motif peptide for in vivo experiments***

To test the efficacy of *Win* motif peptides in inhibiting the H3K4 methylation activity of MLL1 core complex in cells, we synthesized a peptide derived from the *Win* motif in MLL3 and conjugated it to a fluorescent moiety (FITC). The peptide designated MLL3<sup>FITC-TAT</sup> (<sub>Ac</sub>VNPTGCARSEPK-Ahx-K(FITC)-GRKKRRQRRRAP<sub>NH2</sub>) also has a cell penetrating sequence (derived from HIV-TAT) attached to its C-terminus (refer to methods for more details). Before testing the MLL3-FITC-TAT peptide in cells, we first asked if the addition of FITC group or TAT sequence interferes with the ability of MLL3 *Win* motif sequence to inhibit MLL1 core complex activity. MALDI-TOF based methylation assays carried out with the MLL3<sup>WT-FITC-TAT</sup> peptide shows that there is only a slight increase (2-fold) in the IC<sub>50</sub> value for the MLL3<sup>WT-FITC-TAT</sup> peptide compared to that of the MLL3 peptide (VNPTGCARSEPKMS) (Figure 5.19). Since these values are not significantly different, the MLL3<sup>WT-FITC-TAT</sup> peptide was chosen for the further *in vivo* studies described in the next section.

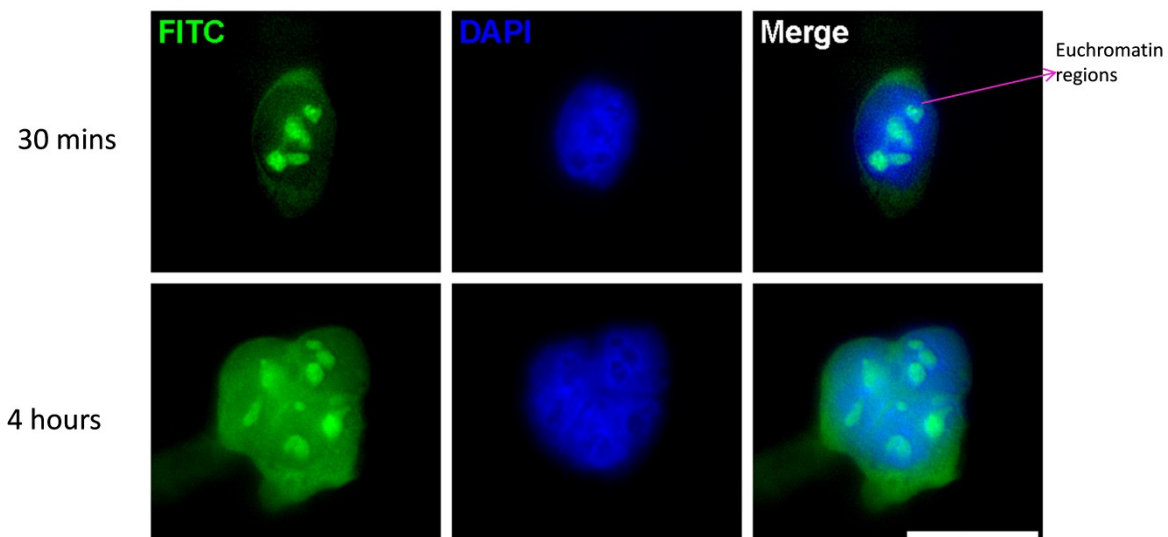


**Figure 5.19 MALDI-TOF based inhibition data for MLL3<sup>WT-FITC-TAT</sup> peptide**

a-b. H3K4 dimethylation inhibition data for MLL3<sup>WT-FITC-TAT</sup> peptide. MALDI-TOF assays were carried out as described in 3.6. a, Relative H3K4 dimethylation levels in the presence of increasing concentrations of MLL3<sup>WT-FITC-TAT</sup> peptide are shown. The data presented for MLL3<sup>WT-FITC-TAT</sup> peptide is from a single experiment. b. IC<sub>50</sub> values derived from fitting the data (refer to 3.6 for curve fitting) are reported.

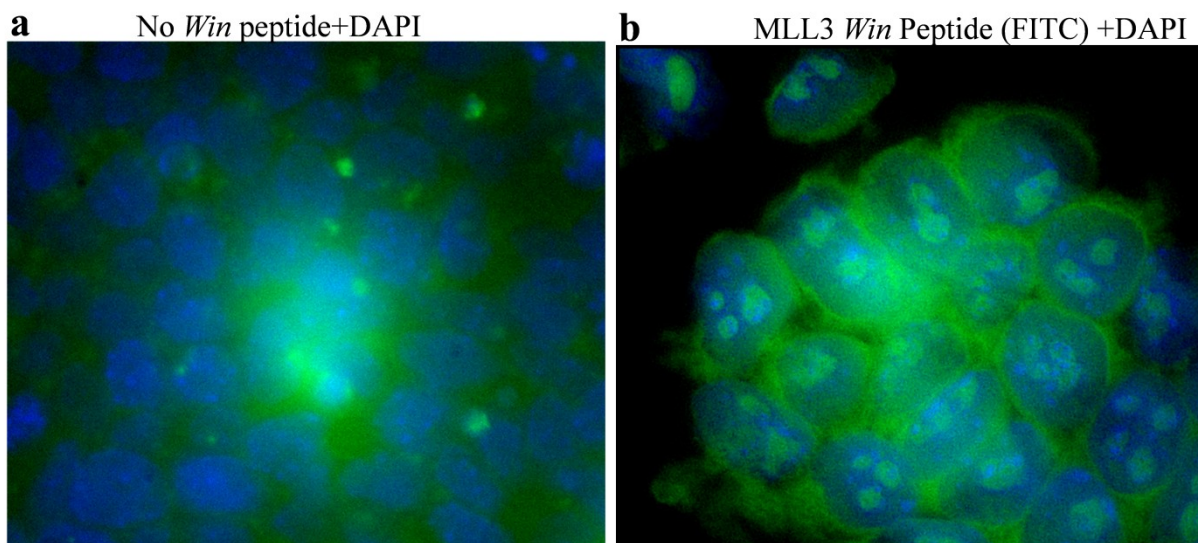
### 5.3.2 *MLL3-FITC-TAT peptide is readily taken up by mammalian cells*

To test if the  $MLL3^{WT-FITC-TAT}$  peptide can cross the cell membrane, we incubated the peptide with mouse P19 cells and checked for FITC fluorescence after various times between 30 min to 24 hours. The results show that the  $MLL3^{WT-FITC-TAT}$  peptide is readily taken up by P19 cells after 30 minutes (Figure 5.20) and FITC fluorescence does not disappear even after 24 hours suggesting that the peptide is stable inside the cell. Interestingly, the peptide localizes to distinct regions within the nucleus that do not overlap with DAPI staining. These regions represent actively transcribing euchromatin regions. Intriguingly, cells exposed to the  $MLL3^{WT-FITC-TAT}$  peptide for 4 hours start to display distinct nuclear defects (Figure 5.21b) compared to cells that were not treated with the peptide (Figure 5.21a). In the cells that are treated with peptide, the nuclei are larger, and microscopic examination revealed that there are fewer mitotic cells. These results suggest that treatment with  $MLL3^{WT-FITC-TAT}$  peptide induces cell-cycle arrest. Additional experiments using a control peptide ( $MLL3^{R4710A-FITC-TAT}$ , wherein the *Win* motif arginine is mutated to an alanine) are required to address this possibility and to distinguish potential peptide toxicity effects.



**Figure 5.20** MLL3<sup>WT-FITC-TAT</sup> peptide is readily taken up by P19 cells

Fluorescence microscopy images showing that MLL3<sup>WT-FITC-TAT</sup> peptide is readily taken up and translocated to the nucleus of mouse P19 cells. In the left panel, FITC fluorescence indicates the localization of MLL3<sup>WT-FITC-TAT</sup> peptide. In the middle panel, DAPI fluorescence shows the nucleus of P19 cells. In the right panel, a merge shows that the MLL3<sup>WT-FITC-TAT</sup> peptide is present in the nucleus and localizes to euchromatin regions. Representative immunofluorescence images after 30 min (upper panels) and 4 hours (lower panels) post-transfection with the MLL3<sup>WT-FITC-TAT</sup> peptide are shown.



**Figure 5.21** MLL3<sup>WT-FITC-TAT</sup> peptide induces nuclear defects in P19 cells

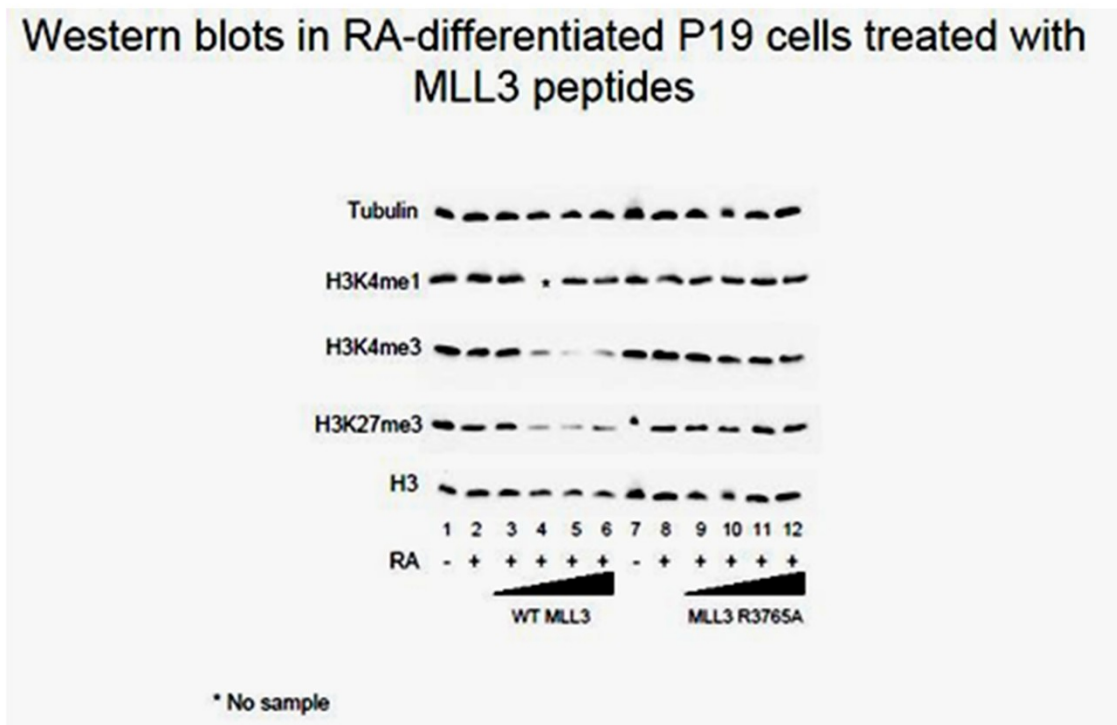
Fluorescence microscopy images showing that MLL3<sup>WT-FITC-TAT</sup> peptide induces nuclear defects in mouse P19 cells. a, DAPI/FITC fluorescence merge of P19 cells in the absence of MLL3<sup>WT-FITC-TAT</sup> peptide (treated with DMEM media for 24 hours). b, DAPI/FITC fluorescence merge shows that nuclei are significantly larger in P19 cells treated with the MLL3<sup>WT-FITC-TAT</sup> peptide for 24 hours.

### 5.3.3 MLL3-FITC-TAT peptide inhibits global H3K4 trimethylation levels in mammalian cells

Inhibition studies reported in Chapters 3-5 suggest that *Win* motif peptides are highly potent inhibitors of H3K4 dimethylation activity *in vitro*. To determine if *Win* motif peptides inhibit the H3K4 di/trimethylation activity in cells, we measured the global levels of H3K4 trimethylation in P19 cells that are treated with either a wild-type (MLL3<sup>WT-FITC-TAT</sup>) or control (MLL3<sup>R4710A-FITC-TAT</sup>) peptides. Western blots using antibodies specific for trimethylated H3K4 shows that the global levels of H3K4 trimethylation is decreased in cells treated with MLL3<sup>WT-FITC-TAT</sup> peptide (Figure 5.22, row 3, compare lanes 3-6 versus lane 2 that was not treated with the peptide).

However, the relative levels of H3K4 monomethylation, is unchanged (Figure 5.22, row 2, lanes 2-6). These results suggest that MLL3<sup>WT-FITC-TAT</sup> peptide inhibits only the H3K4 trimethylation activity of the MLL1 core complex possibly by disrupting MLL1 *Win* motif-WDR5 interaction. Furthermore, a control peptide (MLL3<sup>R4710A-FITC-TAT</sup>) wherein the *Win* motif arginine is replaced with an alanine did show any changes in H3K4 trimethylation levels (Figure 5.22, row 3, lanes 8-12) suggesting that the effects are specific for the MLL3<sup>WT-FITC-TAT</sup> peptide. The observed decrease is not due to the decrease in the amount of total histone H3 as western blot using anti-histone H3 antibody reveals similar amounts of histone H3 across different samples (Figure 5.22, row 5). Unexpectedly, the levels of H3K27me3 are also decreased in cells treated with MLL3<sup>WT-FITC-TAT</sup> peptide (Figure 5.22, row 3, lanes 3-6). One possible explanation is that MLL1 (and other SET1 family complexes) could be regulating the expression of enzymes that catalyze H3K27me3, loss of MLL1 activity in the presence of MLL3<sup>WT-FITC-TAT</sup> peptide would therefore result in an indirect loss of H3K27 methylation. An alternate explanation could be that MLL3<sup>WT-FITC-TAT</sup> peptide could also be targeting enzymatic complexes that catalyze H3K27 trimethylation. Further experiments are required to distinguish these possibilities. In summary,

the *in vivo* results suggest that MLL3<sup>WT-FITC-TAT</sup> peptide is a highly potent inhibitor for H3K4 di/trimethylation in cells and the inhibition is dependent on the conserved *Win* motif arginine. These *in vivo* results are highly promising and future experiments will test the efficiency of *Win* motif peptides and related compounds in inhibiting H3K4 di/trimethylation, Hox gene expression and aberrant cell proliferation in leukemia cells.



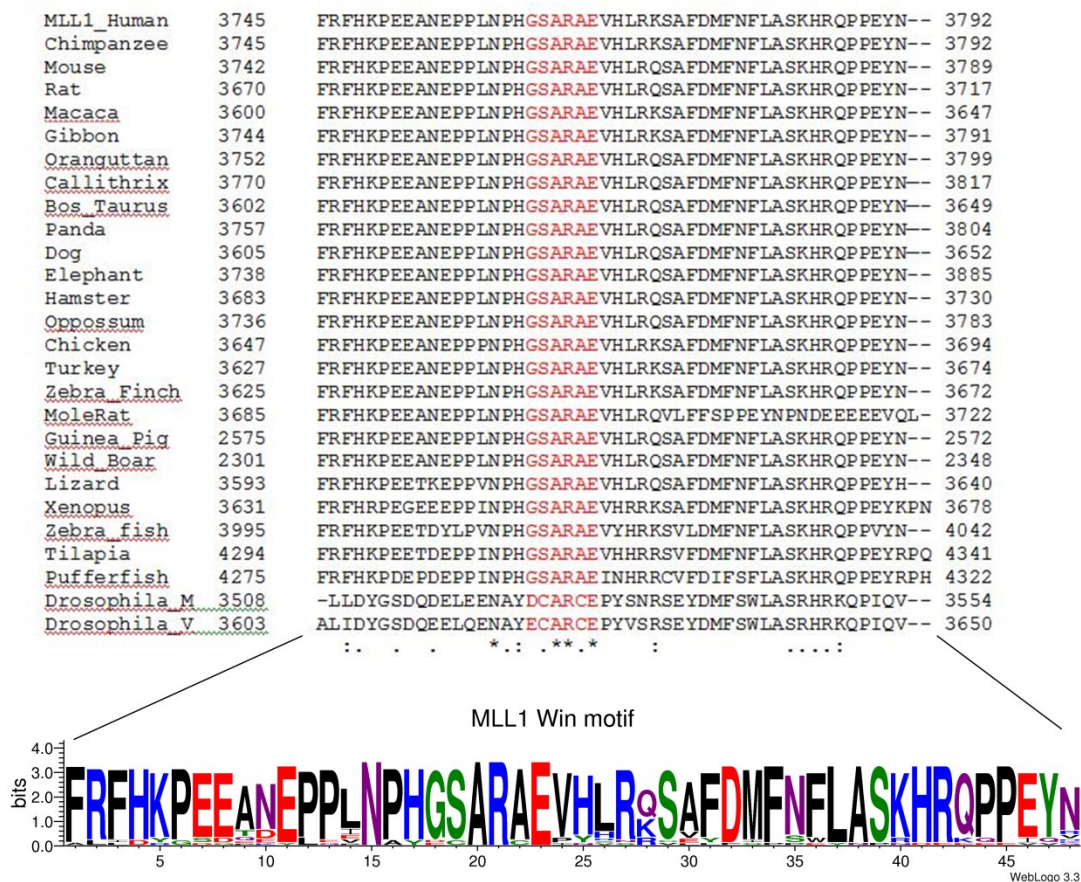
**Figure 5.22 Global levels of H3K4 trimethylation is decreased in the presence of MLL3<sup>WT-FITC-TAT</sup> peptide**

Western blots showing that the wild-type (MLL3<sup>WT-FITC-TAT</sup>) but not the mutant (MLL3<sup>R4710A-FITC-TAT</sup>) peptide inhibits global levels of H3K4 trimethylation. Mouse P19 cells were treated with retinoic acid (RA) to induce differentiation in the absence or presence of MLL3<sup>FITC-TAT</sup> (wild-type or R4710A mutant) peptides. Western blots using antibodies specific for mono (H3K4me1) (row 2)-, or trimethylated (H3K4me3) (row 3) histone H3K4 shows that H3K4me3 levels are decreased in the presence of MLL3<sup>WT-FITC-TAT</sup> peptide (lanes 2-6). H3K27me3 levels are shown for comparison (row 4). Total histone H3 present in each sample is shown (row 5). Tubulin is used as a loading control (row 1).



***Conclusions and future perspectives- Win motif peptides as potential candidates for targeted molecular therapy for MLL1-associated leukemia?***

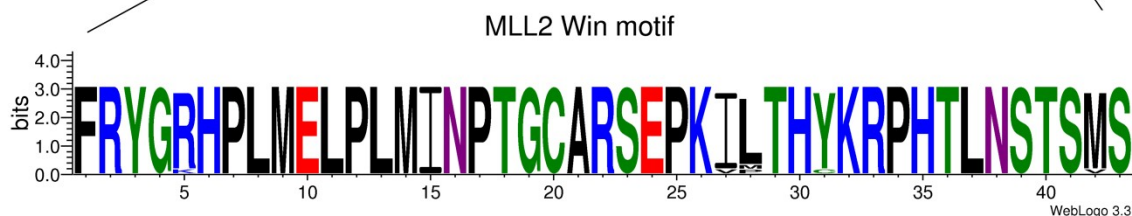
In conclusion, this dissertation summarizes the detailed biochemical and structural studies that were carried out to characterize the interaction between WDR5 and human SET1 family *Win* motifs. The findings from this research have contributed significantly to our understanding of how WDR5 recognizes different *Win* motif-based peptides and have led to the identification of additional peptides that bind WDR5 with greater than 450-fold better affinities. This thesis also summarizes the variety of approaches that could be undertaken to identify novel *Win* motif-based therapeutics. We have also presented promising preliminary data that shows that MLL3 *Win* motif peptide is a highly potent inhibitor of H3K4 trimethylation in cells. Collectively, the data presented in this thesis have identified *Win* motif-like inhibitors as a promising target for MLL1-related leukemias. It is expected that such an MLL1-targeted therapy will have enhanced pharmacological potential compared to the existing broadly-based chemotherapeutics for MLL1-associated leukemias and will hopefully result in better treatment outcomes. In the future, similar biochemical, structural and genetic studies will be instrumental in identifying additional molecular targets that will form the basis for novel treatment strategies.



### Appendix Figure 1a. *Win* motif in MLL1 is highly conserved in metazoan orthologs

Amino acid sequence conservation within the *Win* motif of MLL1 orthologs is shown. *Top*, ClustalW (70) multiple sequence alignment for all the MLL1 orthologs found in NCBI. The conserved residues of the *Win* motif are highlighted in red. *bottom*, a bit-score representation of the amino acid conservation within the sequences used in ClustalW alignment. Weblogo server (weblogo.berkeley.edu) (335) was used to generate the consensus motif with respect to residue positions in the alignment. Degree of conservation is indicated by the size of the letters.

Human_MLL2	5319	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	5361
Bos_Taurus	5285	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	5327
Macaca	5268	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	5310
Panda	5265	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	5307
Orangutan	5341	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	5383
Gibbon	5189	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	5231
Callithrix	5071	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	5113
Dog	5334	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	5376
Wild_Boar	4938	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	4980
Mouse	5370	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	5412
Rat	5325	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	5367
Guinea_Pig	5359	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	5401
Ferret	797	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	839
Hamster	1257	FRYGRHPLMELPLMINPTGCARSEPKILTHYKRPHTLNSTSMS	1299
Tilapia	4689	FRYGRHPLMELPLMINPTGCARSEPKVPTHCKRPHTLNSTSVS	4731
Xenopus	5801	FRYGKHPLMELPLMINPTGCARSEPKIMTHYKRPHTLNSTSMS	5843



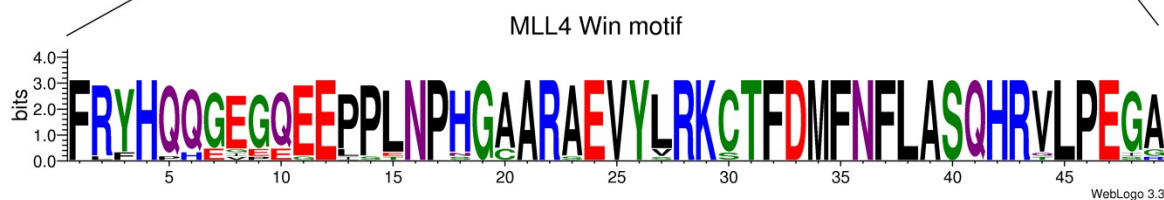
### Appendix Figure 1b. *Win* motif in MLL2 is highly conserved in metazoan orthologs

Amino acid sequence conservation within the *Win* motif of MLL2 orthologs is shown. *Top*, ClustalW (70) multiple sequence alignment for all the MLL2 orthologs found in NCBI. The conserved residues of the *Win* motif are highlighted in red. *bottom*, a bit-score representation of the amino acid conservation within the sequences used in ClustalW alignment. Weblogo server (weblogo.berkeley.edu) (335) was used to generate the consensus motif with respect to residue positions in the alignment. Degree of conservation is indicated by the size of the letters.



Human_MLL4	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Callithrix	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Chimpanzee	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Orangutan	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Macaca	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Gibbon	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Panda	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Bos taurus	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Wild boar	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Dog	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Rat	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Mouse	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Mole rat	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Guinea pig	FRYHQGGEGQEEPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLP	49
Anolis	FRYHQHEGEEELPLNPHGCARAEVYSRKCTFDMFNFLASQHRVLP	49
Xenopus	FLFHPQEEVEEGEISTNPSGCARSEVYVRKSTFDMFNFLASQHRVLP	49
Zebra fish	FRFHQHEIPEEELPENPNGCARAEVYVRKSTFDMFNFLASQHRVLP	49

\* : \* : : \* . \*\* \* . \*\* : \*\*\* \*\* . \*\*\*\*\* \*\*

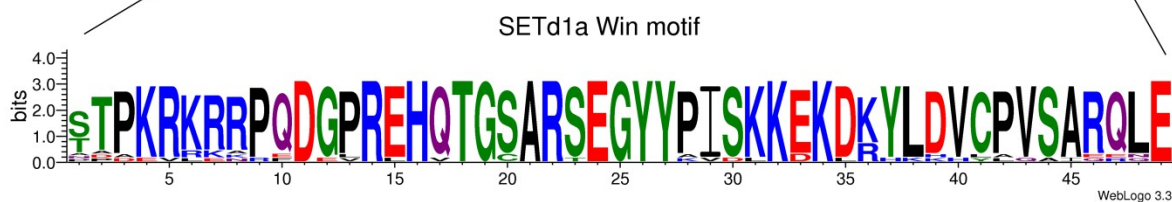


#### Appendix Figure 1d. *Win* motif in MLL4 is highly conserved in metazoan orthologs

Amino acid sequence conservation within the *Win* motif of MLL4 orthologs is shown. *Top*, ClustalW (70) multiple sequence alignment for all the MLL4 orthologs found in NCBI. The conserved residues of the *Win* motif are highlighted in red. *bottom*, a bit-score representation of the amino acid conservation within the sequences used in ClustalW alignment. Weblogo server (weblogo.berkeley.edu) (335) was used to generate the consensus motif with respect to residue positions in the alignment. Degree of conservation is indicated by the size of the letters.

Human_SETd1a	1474	-----TTPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDKYLDVCPVSARQLE--	1522
Chimpanzee	1474	-----TTPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDKYLDVCPVSARQLE--	1522
Gibbon	1442	-----TTPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDKYLDVCPVSARQLE--	1490
Bos_taurus	1475	-----STPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDRYLDVCPVSARQLE--	1523
Callithrix	1475	-----TTPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDKYLDVCPVSARQLE--	1523
Elephant	1470	-----NTPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDKYLDVCPVSARQLE--	1518
Panda	1360	-----STPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDRYLDVCPVSARQLE--	1408
Dog	1480	-----STPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDRYLDVCPVSARQLE--	1528
Wild boar	1317	-----STPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDRYLDVCPVSARQLE--	1365
Horse_XP_001915080	1474	-----STPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDRYLDVCPVSARQLE--	1522
Opposum	1473	-----TTPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDKYLDVCPVSARQLE--	1521
Mouse	1483	-----STPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDKYLDVCPVSARQLE--	1531
Mole_rat	1269	-----PADEVLEAPEDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDKYLDVCPVSARQLE--	1317
Hamster	1490	-----STPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDKYLDVCPVSARQLE--	1538
Guinea_pig	1469	-----STPKRRRPQDGP <del>REHQT</del> <b>GSARSE</b> GYYPISKKEKDKYLDVCPVSARQLE--	1517
Xenopus	1585	-----ADPKRRNRQDGL <del>REHQT</del> <b>GCARSE</b> GYAISKDKDKYLKVLVAEEQE--	1633
Jewel_wasp	1443	-----SPAKRRKRDEVRLHVT <b>GCARTE</b> GYKVDLKDCLKKHHHYAQSIQRNED--	1490
Coprinopsis	1362	-----SAPEPSEPTSSRKHLT <b>GSARTE</b> GFYKITKQDKAMYLEQYKSKAAAEEVIP	1410
Puccinia	733	-----AQSESKGKANEVGRHST <b>GSARTE</b> GYYHISSSQKAIYLPQ-RNKAIIDLGS	782
S_pombe	680	FKNGDVKYGDTAILPEPKGYFRSNTS <b>GSARSE</b> GYIIPTEKSLYLPLRNR-----	730

: \* : : \* : \* : \* : \* : \* :

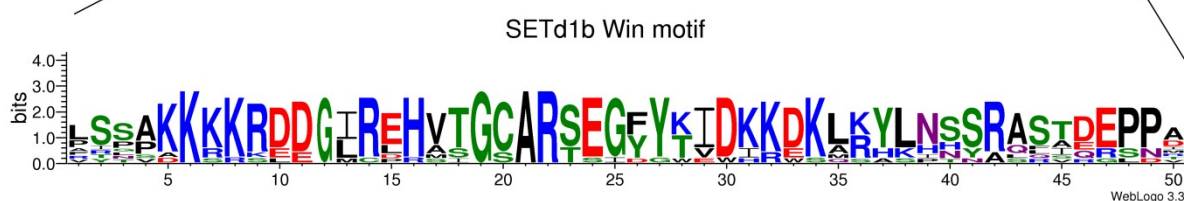


### Appendix Figure 1e. *Win* motif in SETd1a is highly conserved in metazoan orthologs

Amino acid sequence conservation within the *Win* motif of SETd1a orthologs is shown. *Top*, ClustalW (70) multiple sequence alignment for all the SETd1a orthologs found in NCBI. The conserved residues of the *Win* motif are highlighted in red. *bottom*, a bit-score representation of the amino acid conservation within the sequences used in ClustalW alignment. Weblogo server (weblogo.berkeley.edu) (335) was used to generate the consensus motif with respect to residue positions in the alignment. Degree of conservation is indicated by the size of the letters.

Human	1684	LSSAKKKKRDDGIREHVTGCARSEGFYTIIDKKDKLRYLNSSRASTDEPPA-	1733
Bos_Taurus	1620	LSSAKKKKRDDGIREHMTGCARSEGFYTIIDKKDKLRYLNSSRASTDEPPT-	1669
Dog	1681	LSSAKKKKRDDGIREHVTGCARSEGFYTIIDKKDKLRYLNSSRASTDEPPA-	1730
Mouse	1746	LSSAKKKKREDDGIREHVTGCARSEGFYTIIDKKDKLRYLNSSRASTDEPPM-	1795
Chicken	1769	FSTPKKKKRDDGMRHVTGCARSEGYKIDKKDKLKYLNNSRAFAEPPA-	1818
Mole_rat	1248	-ILPDKKKRDDGIREHVTGCARSEGFYTIIDKKDKLRYLNSSRASTDEPPAD	1297
Xenopus	1699	VYSPKKKKRDDGLREHVTGCARSEGYKIDKKDKLKYLNNSRADEPPI-	1748
Zebrafish	1607	VTGVKKKRKEDGIRDHVTGCARSEGYKIDKKDKMXYLNSSRLQSEEPDV-	1656
Ant	1260	-PSPAKKKRKRDELRLHASGSARTEGYKVDIREKAKHKHHYAQSIQRSNDV	1309
Carpenter_Ant	1169	-PSPAKKKRKRDELRLHASGSARTEGYKVDIREKAKHKHHYAQSIQRSNDV	1218
Barrel_clover	976	-RRNSAKSKSLDLCPRSIGCARTSIDGWEWHKWSQSASPTSRARVRGLPRL	1025

\* : : : \*.\*\*. : .

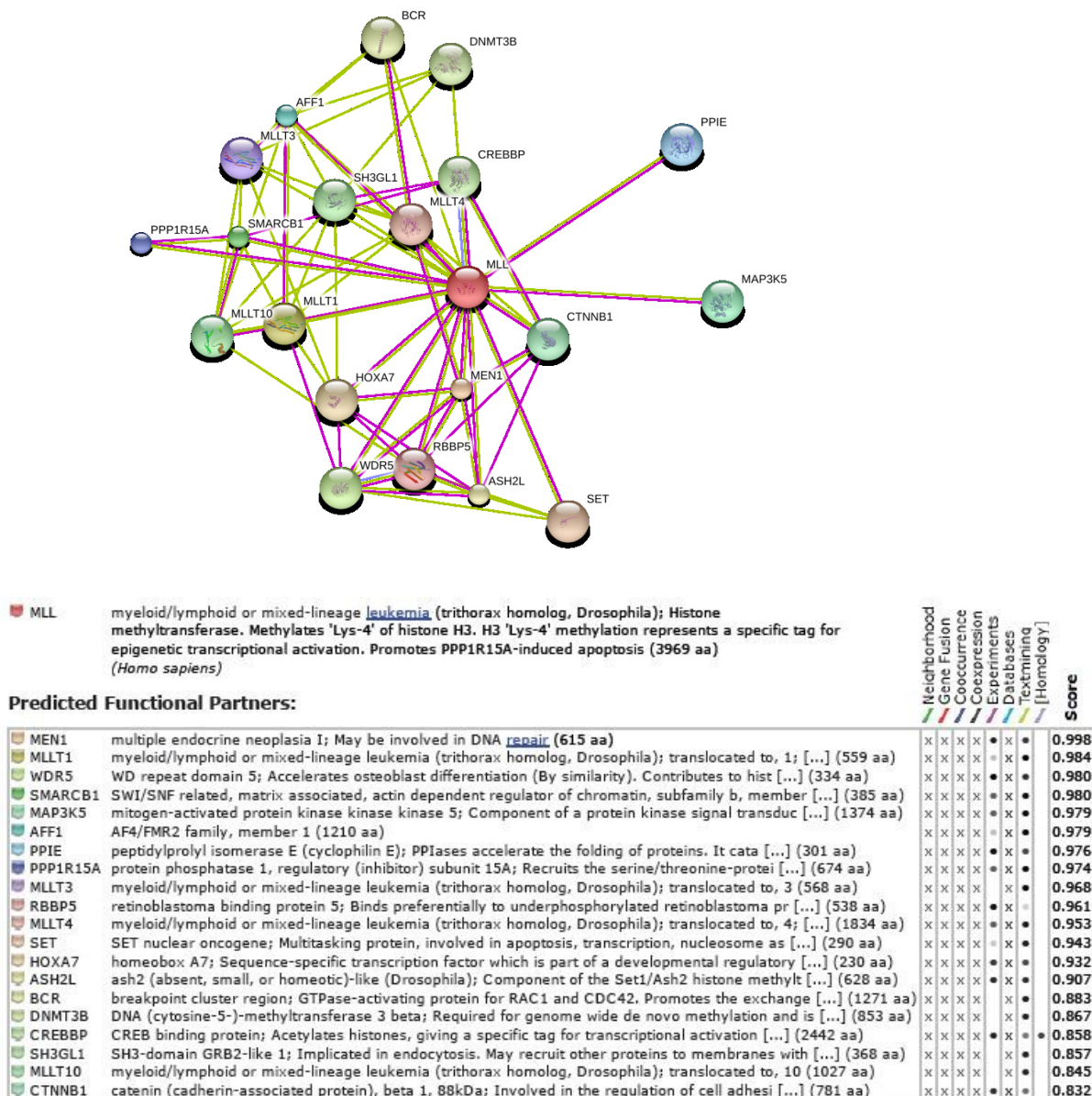


### Appendix Figure 1f. *Win* motif in SETd1b is highly conserved in metazoan orthologs

Amino acid sequence conservation within the *Win* motif of SETd1b orthologs is shown. *Top*, ClustalW (70) multiple sequence alignment for all the SETd1b orthologs found in NCBI. The conserved residues of the *Win* motif are highlighted in red. *bottom*, a bit-score representation of the amino acid conservation within the sequences used in ClustalW alignment. Weblogo server (weblogo.berkeley.edu) (335) was used to generate the consensus motif with respect to residue positions in the alignment. Degree of conservation is indicated by the size of the letters.

## Appendix 2

### Appendix 2a. STRING Interaction network for human MLL1

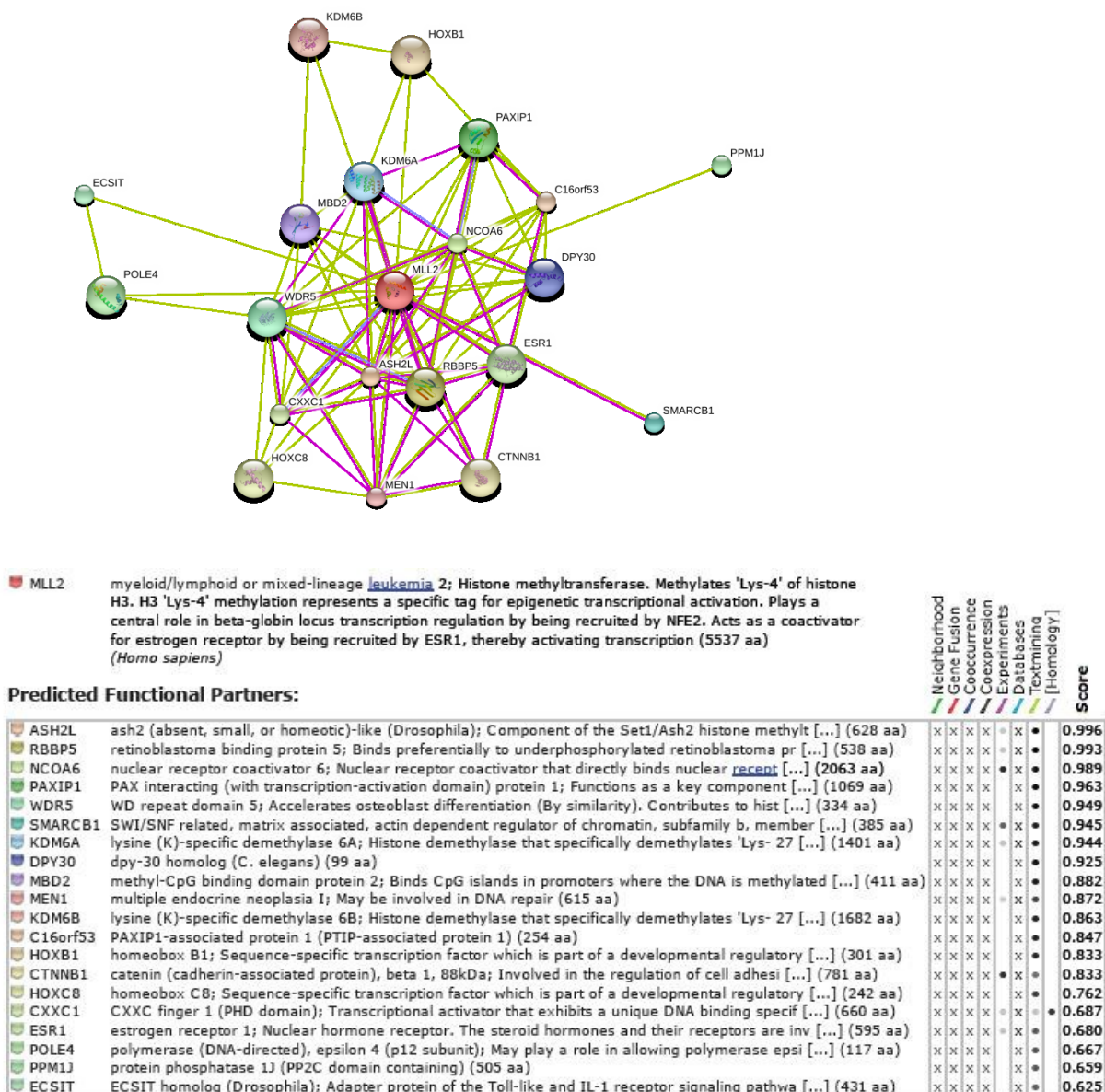


**Figure 2a STRING interaction network for human MLL1**

Protein-protein interactions associated with the full length MLL1 protein are shown. The network was created using the STRING 9.0 database (336, 337) and the evidence view was selected to display the interactions. The query sequence and the list of interaction partners are listed below the network. Only the top 20 partners for MLL1 identified through experimental evidence and text mining resulting from a STRING search set at confidence level 0.5 are shown. Nodes are either colored (if they are directly linked to the input - as in the table) or white (nodes of a higher iteration/depth). Edges, i.e. predicted functional links, consist of up to eight lines: one color for each type of evidence.



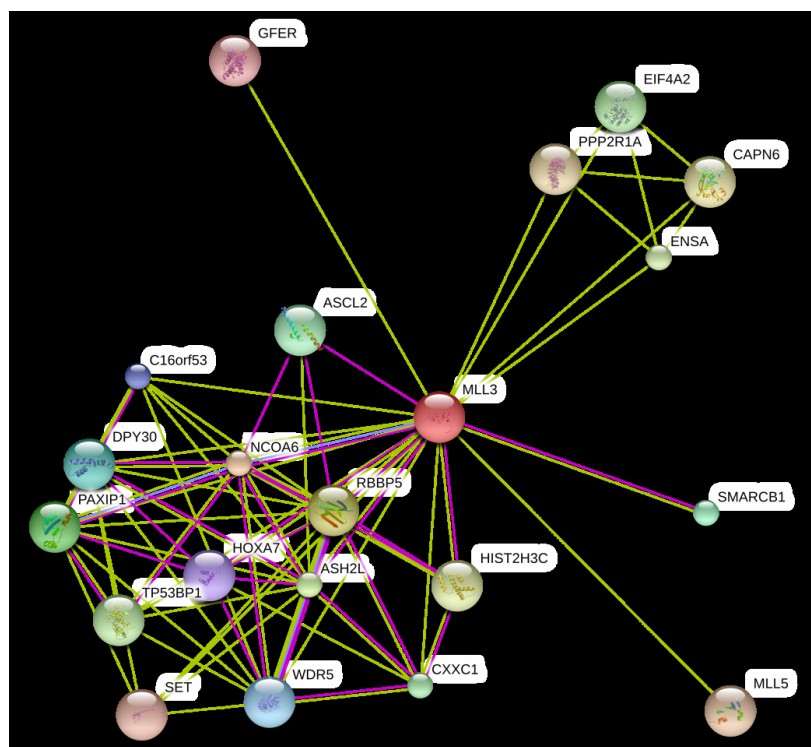
## Appendix 2b. STRING interaction network for human MLL2



**Figure 2b. STRING interaction network for human MLL2**

Protein-protein interactions associated with the full length MLL2 protein are shown. The network was created using the STRING 9.0 database (336, 337) as described in appendix Figure 2a.

## Appendix 2c. STRING Interaction network for human MLL3



### Your Input:

- **MLL3** myeloid/lymphoid or mixed-lineage leukemia 3; Histone methyltransferase. Methylates 'Lys-4' of histone H3. H3 'Lys-4' methylation represents a specific tag for epigenetic transcriptional activation. Central component of the MLL2/MLL3 complex, a coactivator complex of nuclear receptors, involved in transcriptional coactivation. MLL3 may be a catalytic subunit of this complex. May be involved in leukemogenesis and developmental disorder (4911 aa) (*Homo sapiens*)

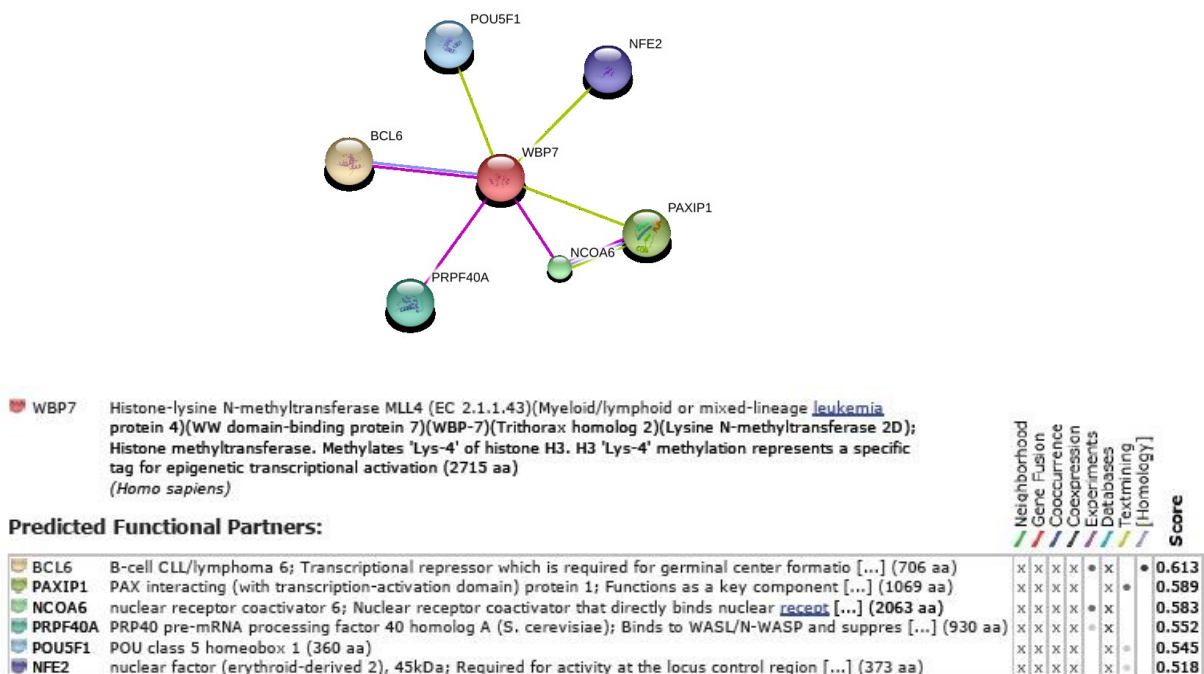
### Predicted Functional Partners:

		Neighborhood	Gene Fusion	Coincidence	Coincidence	Experiments	Databases	Textmining	Homology	Score
NCOA6	nuclear receptor coactivator 6; Nuclear receptor coactivator that directly binds nuclear <a href="#">receptor</a> [...] (2063 aa)	x	x	x	x	x	x	x	x	0.973
RBBP5	retinoblastoma binding protein 5; Binds preferentially to underphosphorylated retinoblastoma pr [...] (538 aa)	x	x	x	x	x	x	x	x	0.972
ASH2L	ash2 (absent, small, or homeotic)-like (Drosophila); Component of the Set1/Ash2 histone methyl [...] (628 aa)	x	x	x	x	x	x	x	x	0.970
PAXIP1	PAX interacting (with transcription-activation domain) protein 1; Functions as a key component [...] (1069 aa)	x	x	x	x	x	x	x	x	0.965
SMARCB1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member [...] (385 aa)	x	x	x	x	x	x	x	x	0.949
DPY30	dpy-30 homolog (C. elegans) (99 aa)	x	x	x	x	x	x	x	x	0.936
WDR5	WD repeat domain 5; Accelerates osteoblast differentiation (By similarity). Contributes to hist [...] (334 aa)	x	x	x	x	x	x	x	x	0.904
C16orf53	PAXIP1-associated protein 1 (PTIP-associated protein 1) (254 aa)	x	x	x	x	x	x	x	x	0.880
HOXA7	homeobox A7; Sequence-specific transcription factor which is part of a developmental regulatory [...] (230 aa)	x	x	x	x	x	x	x	x	0.792
GFER	growth factor, augments of liver regeneration; FAD-dependent sulfhydryl oxidase that catalyzes [...] (205 aa)	x	x	x	x	x	x	x	x	0.752
SET	SET nuclear oncogene; Multitasking protein, involved in apoptosis, transcription, nucleosome as [...] (290 aa)	x	x	x	x	x	x	x	x	0.670
MLL5	myeloid/lymphoid or mixed-lineage leukemia 5 (trithorax homolog, Drosophila); Histone methyltra [...] (1858 aa)	x	x	x	x	x	x	x	x	0.631
PPP2R1A	protein phosphatase 2 (formerly 2A), regulatory subunit A, alpha isoform; The PR65 subunit of p [...] (589 aa)	x	x	x	x	x	x	x	x	0.625
CAPN6	calpain 6; Not known; does not seem to have protease activity as it has lost the <a href="#">active site re</a> [...] (641 aa)	x	x	x	x	x	x	x	x	0.625
HIST2H3C	histone cluster 2, H3c (136 aa)	x	x	x	x	x	x	x	x	0.614
ENSA	endosulfine alpha; Endogenous ligand for sulfonylurea receptor. By inhibiting sulfonylurea from [...] (137 aa)	x	x	x	x	x	x	x	x	0.601
TP53BP1	tumor protein p53 binding protein 1; May have a role in checkpoint signaling during mitosis (By [...] (1977 aa)	x	x	x	x	x	x	x	x	0.596
EIF4A2	microRNA 1248; ATP-dependent RNA helicase which is a subunit of the eIF4F complex involved in c [...] (407 aa)	x	x	x	x	x	x	x	x	0.595
CXXC1	CXXC finger 1 (PHD domain); Transcriptional activator that exhibits a unique DNA binding specif [...] (660 aa)	x	x	x	x	x	x	x	x	0.587
ASCL2	achaete-scute complex homolog 2 (Drosophila); AS-C proteins are involved in the determination o [...] (193 aa)	x	x	x	x	x	x	x	x	0.577

**Figure 2c. STRING interaction network for human MLL3**

Protein-protein interactions associated with the full length MLL3 protein are shown. The network was created using the STRING 9.0 database (336, 337) as described in appendix Figure 2a.

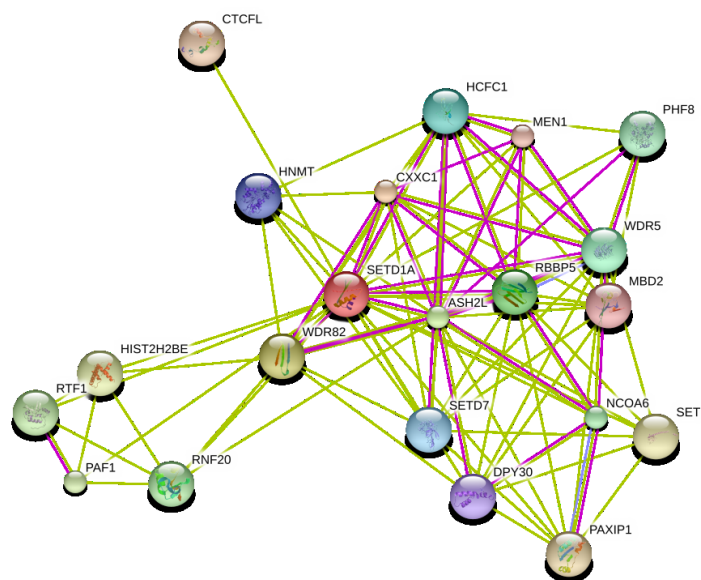
## Appendix 2d. STRING Interaction network for human MLL4



**Figure 2d. STRING interaction network for human MLL4**

Protein-protein interactions associated with the full length MLL4 protein are shown. The network was created using the STRING 9.0 database (336, 337) as described in appendix Figure 2a. STRING search for human MLL4 identified only 6 interaction partners and did not predict the components of WRAD sub-complex as interaction partners for MLL4.

## Appendix 2e. STRING Interaction network for SETd1a



### Your Input:

- SETD1A** SET domain containing 1A; Histone methyltransferase that specifically methylates 'Lys-4' of histone H3, when part of the SET1 histone methyltransferase (HMT) complex, but not if the neighboring 'Lys- 9' residue is already methylated. H3 'Lys-4' methylation represents a specific tag for epigenetic transcriptional activation. The non-overlapping localization with SETD1B suggests that SETD1A and SETD1B make non-redundant contributions to the epigenetic control of chromatin structure and gene expression (1707 aa)  
(*Homo sapiens*)

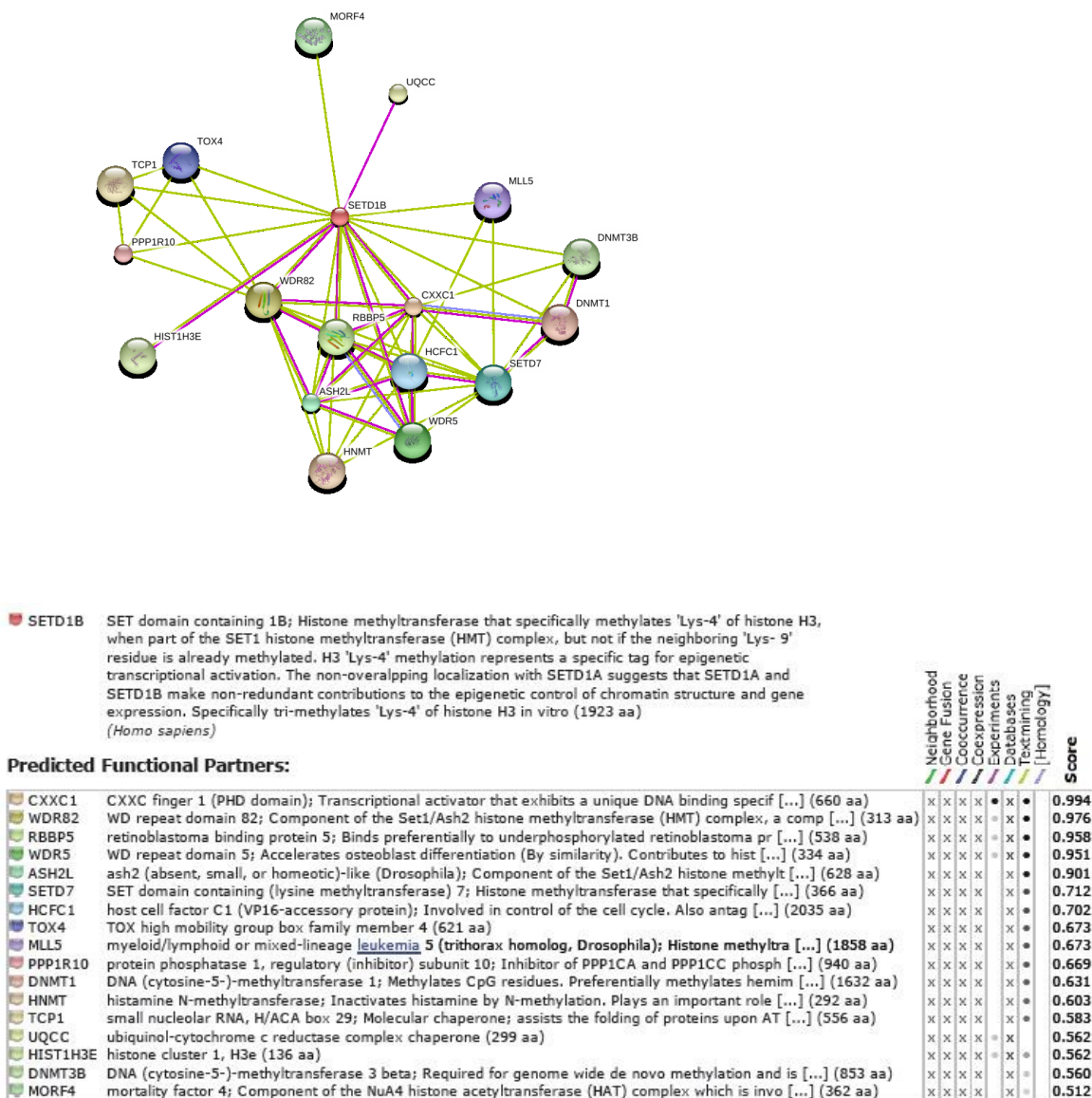
### Predicted Functional Partners:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	Homology	Score
CXXC1	CXXC finger 1 (PHD domain); Transcriptional activator that exhibits a unique DNA binding specif [...] (660 aa)	x	x	x	x	●	x	●	●	0.999
WDR82	WD repeat domain 82; Component of the Set1/Ash2 histone methyltransferase (HMT) complex, a comp [...] (313 aa)	x	x	x	x	●	x	●	●	0.997
ASH2L	ash2 (absent, small, or homeotic)-like ( <i>Drosophila</i> ); Component of the Set1/Ash2 histone methylt [...] (628 aa)	x	x	x	x	●	x	●	●	0.995
RBBP5	retinoblastoma binding protein 5; Binds preferentially to underphosphorylated retinoblastoma pr [...] (538 aa)	x	x	x	x	●	x	●	●	0.992
WDR5	WD repeat domain 5; Accelerates osteoblast differentiation (By similarity). Contributes to hist [...] (334 aa)	x	x	x	x	●	x	●	●	0.992
HCFC1	host cell factor C1 (VP16-accessory protein); Involved in control of the cell cycle. Also antag [...] (2035 aa)	x	x	x	x	●	x	●	●	0.984
SETD7	SET domain containing (lysine methyltransferase) 7; Histone methyltransferase that specifically [...] (366 aa)	x	x	x	x	x	x	●	●	0.962
HNMT	histamine N-methyltransferase; Inactivates histamine by N-methylation. Plays an important role [...] (292 aa)	x	x	x	x	x	x	●	●	0.928
DPY30	dpy-30 homolog ( <i>C. elegans</i> ) (99 aa)	x	x	x	x	x	x	●	●	0.919
MBD2	methyl-CpG binding domain protein 2; Binds CpG islands in promoters where the DNA is methylated [...] (411 aa)	x	x	x	x	x	x	●	●	0.900
MEN1	multiple endocrine neoplasia I; May be involved in DNA <a href="#">repair</a> (615 aa)	x	x	x	x	x	x	●	●	0.871
CTCF	CCCTC-binding factor (zinc finger protein)-like; Testis-specific DNA binding protein responsibl [...] (663 aa)	x	x	x	x	x	x	●	●	0.866
PAXIP1	PAX interacting (with transcription-activation domain) protein 1; Functions as a key component [...] (1069 aa)	x	x	x	x	x	x	●	●	0.857
SET	SET nuclear oncogene; Multitasking protein, involved in apoptosis, transcription, nucleosome as [...] (290 aa)	x	x	x	x	x	x	●	●	0.817
HIST2H2BE	histone cluster 2, H2be; Core component of nucleosome. Nucleosomes wrap and compact DNA into ch [...] (126 aa)	x	x	x	x	x	x	●	●	0.783
PAF1	Paf1, RNA polymerase II associated factor, homolog ( <i>S. cerevisiae</i> ); The mammalian PAF1 complex [...] (531 aa)	x	x	x	x	x	x	●	●	0.768
RTF1	Rtf1, Paf1/RNA polymerase II complex component, homolog ( <i>S. cerevisiae</i> ); Plays a role in transc [...] (710 aa)	x	x	x	x	x	x	●	●	0.762
RNF20	ring finger protein 20; E3 ubiquitin-protein ligase that mediates monoubiquitination of 'Lys-12 [...] (975 aa)	x	x	x	x	x	x	●	●	0.758
NCOA6	nuclear receptor coactivator 6; Nuclear receptor coactivator that directly binds nuclear recept [...] (2063 aa)	x	x	x	x	x	x	●	●	0.758
PHF8	PHD finger protein 8; Functions as an Fe(2+) and 2-oxoglutarate dependent histone lysine demeth [...] (1024 aa)	x	x	x	x	x	x	●	●	0.742

**Figure 2e. STRING interaction network for human SETd1a**

Protein-protein interactions associated with the full length SETd1a protein are shown. The network was created using the STRING 9.0 database (336, 337) as described in appendix Figure 2a.

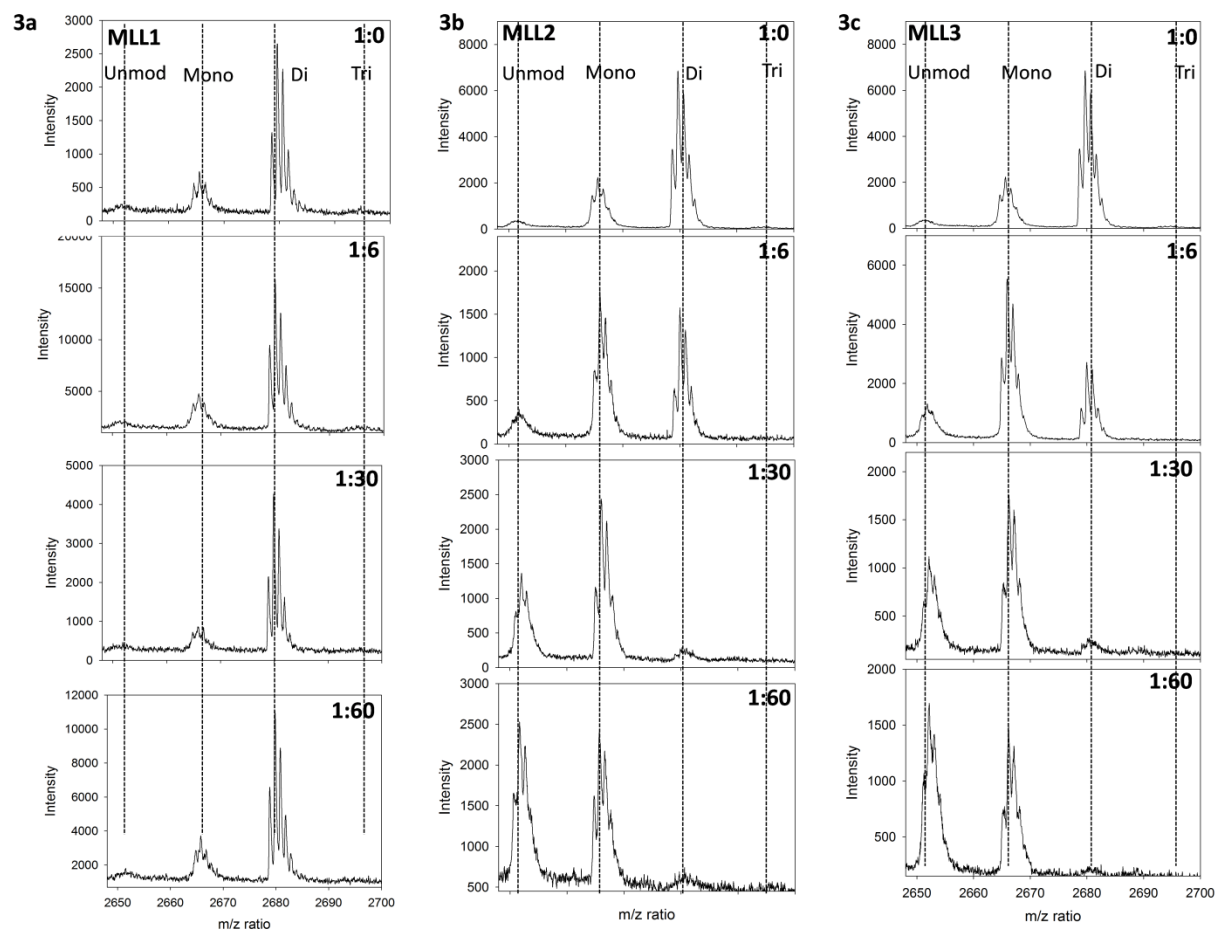
## Appendix 2f. STRING Interaction network for hSETd1b

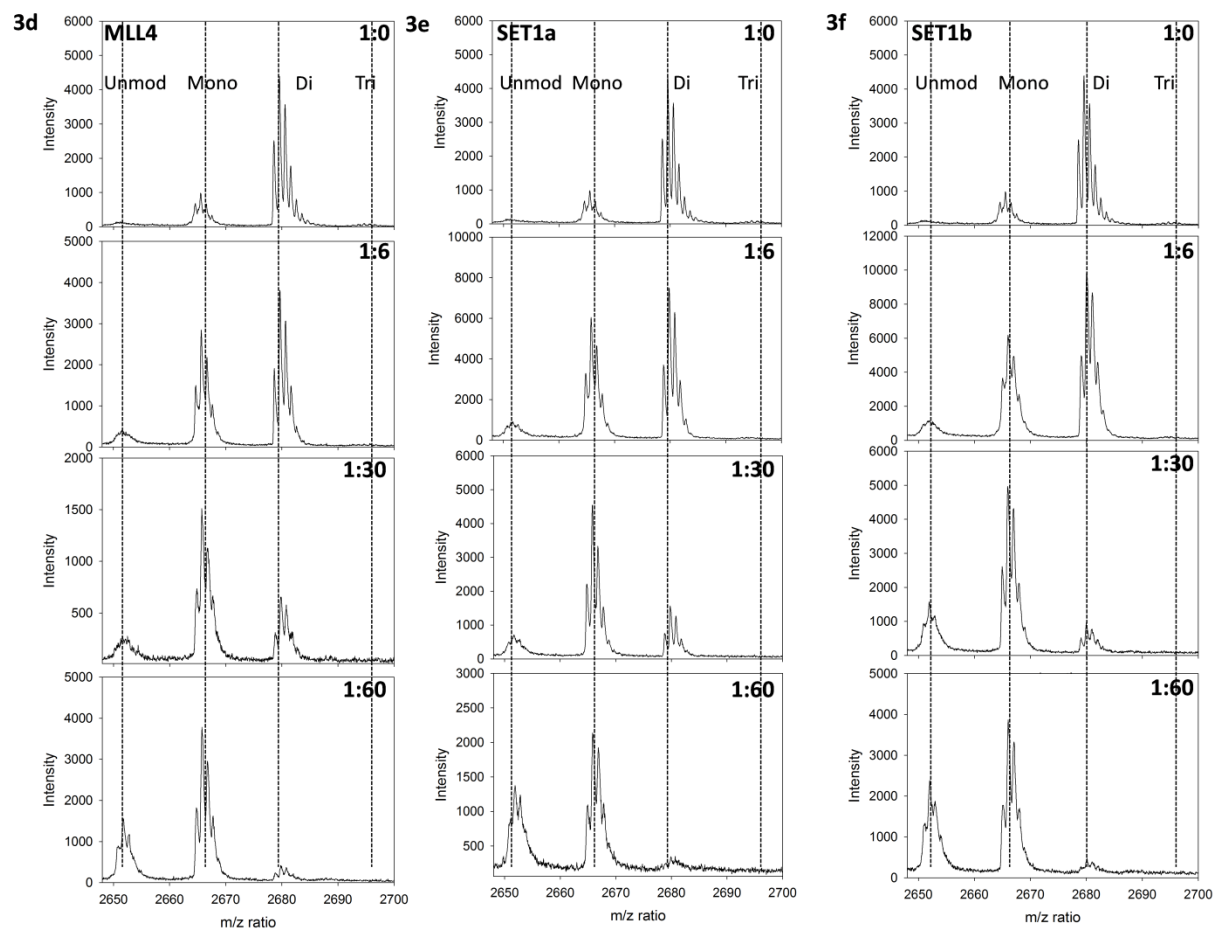


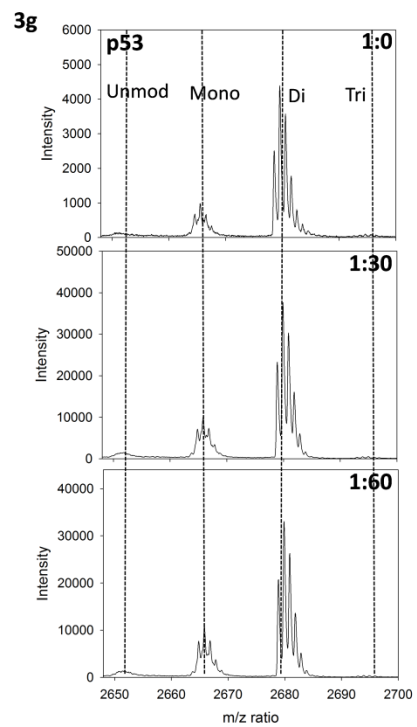
**Figure 2f. STRING interaction network for human SETd1b**

Protein-protein interactions associated with the full length SETd1b protein are shown. The network was created using the STRING 9.0 database (336, 337) as described in appendix Figure 2a.

### Appendix 3a-g: MALDI-TOF mass spectrometry data for human SET1 family Win motif peptides under no pre-incubation conditions

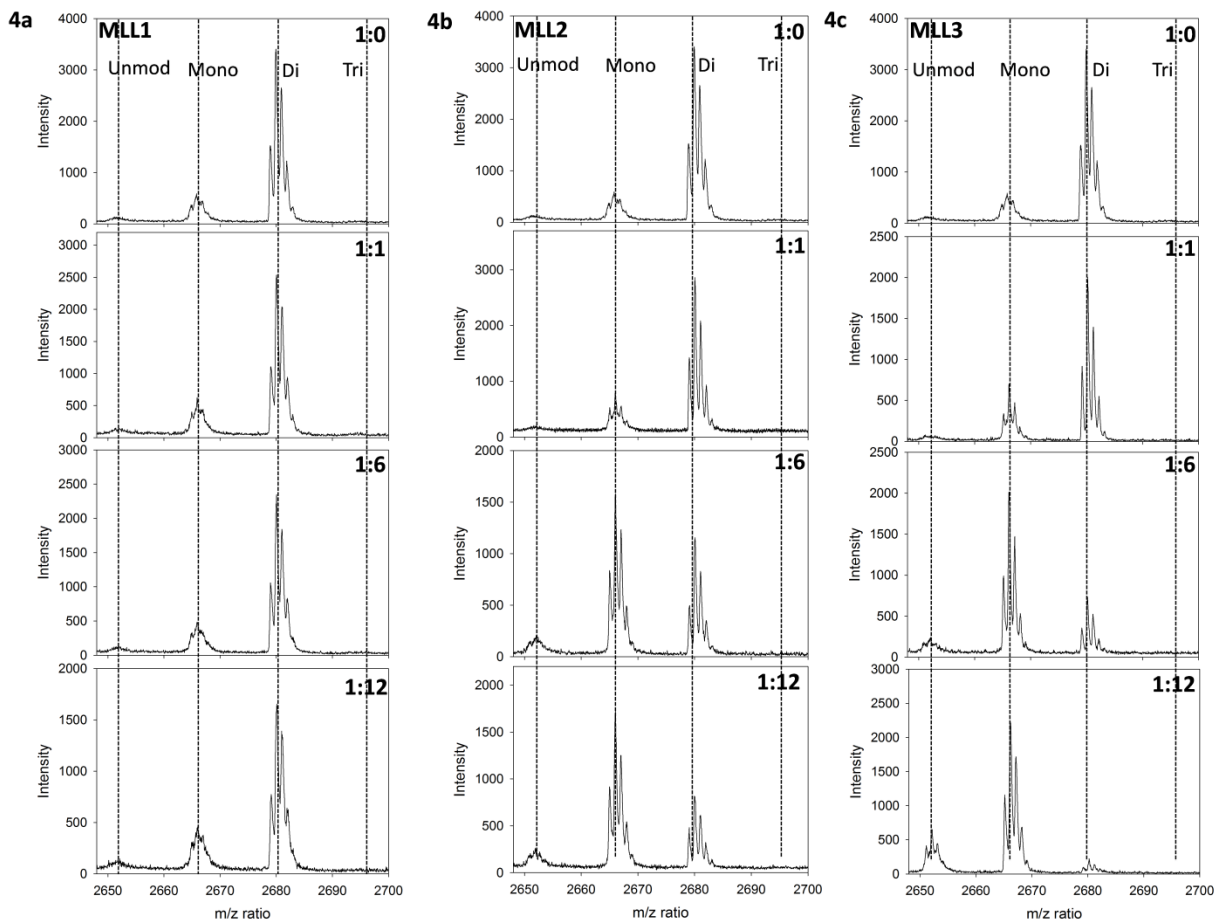


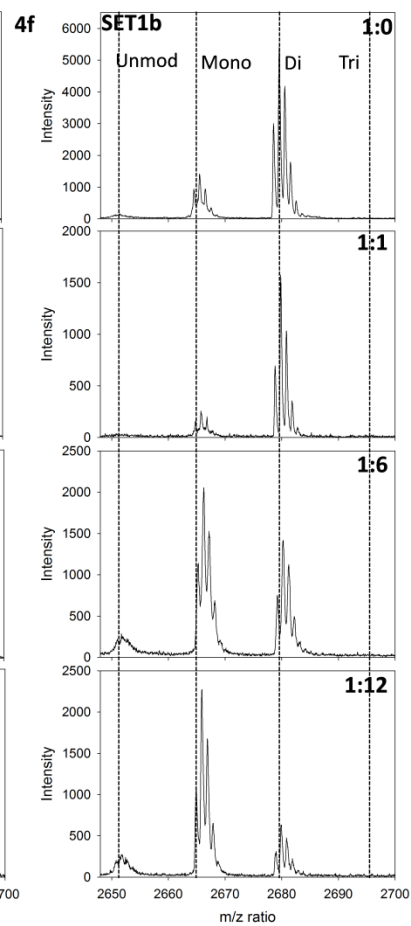
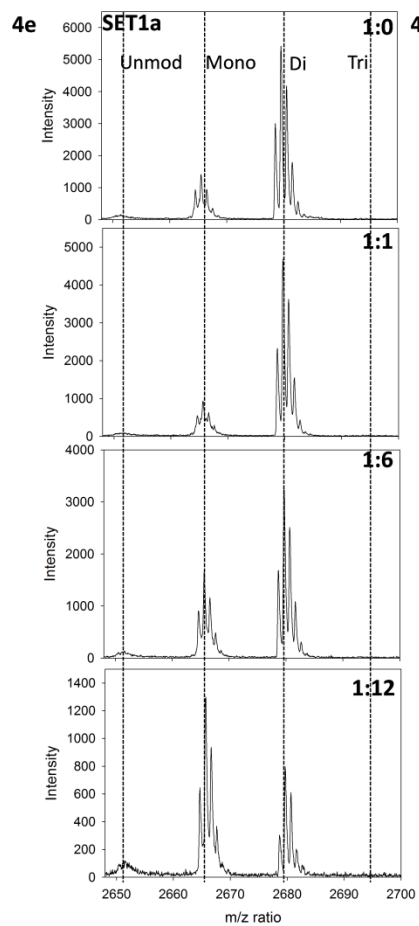
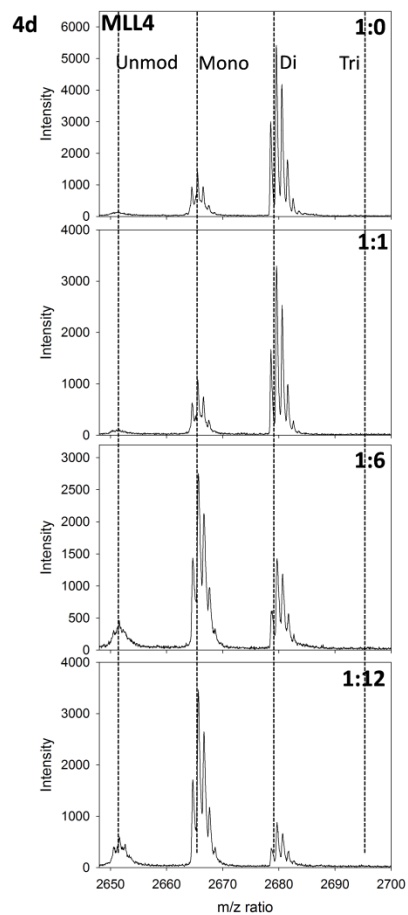


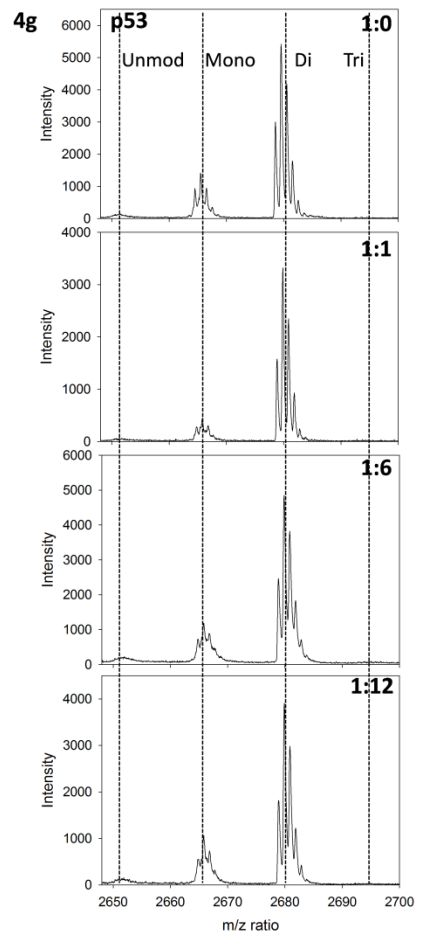


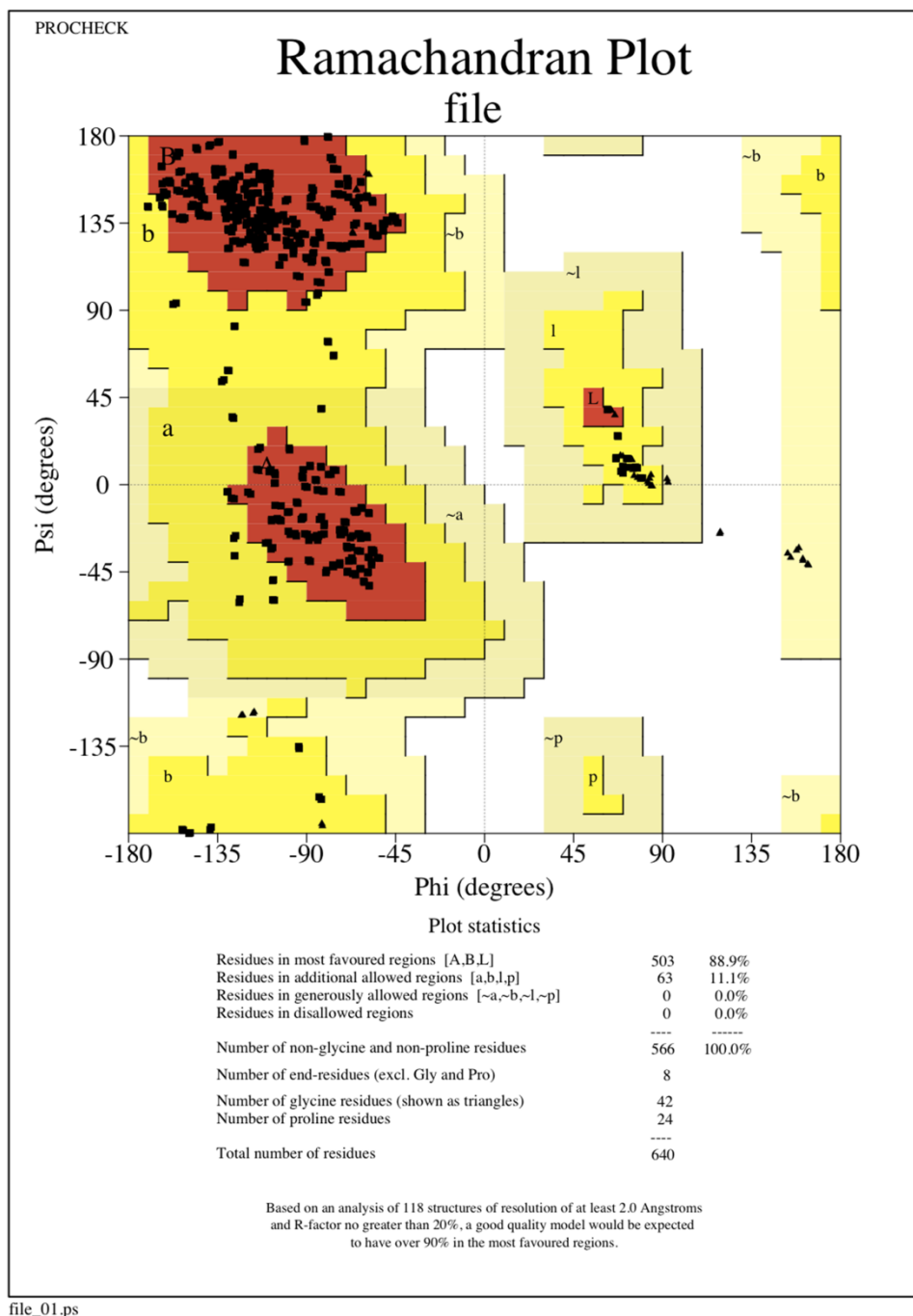


### Appendix 4a-g: MALDI-TOF mass spectrometry data for human SET1 family Win motif peptides under pre-incubation conditions

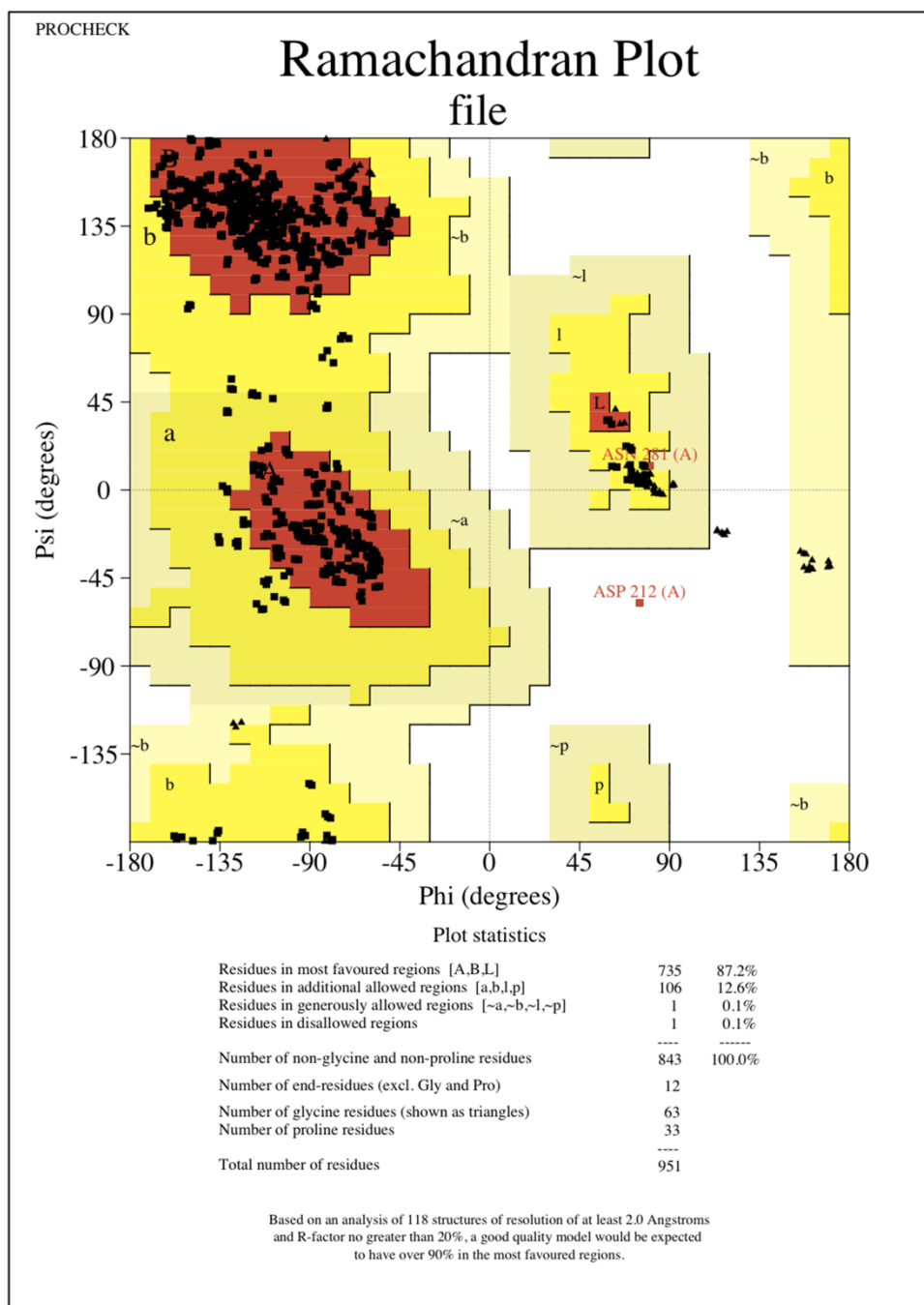




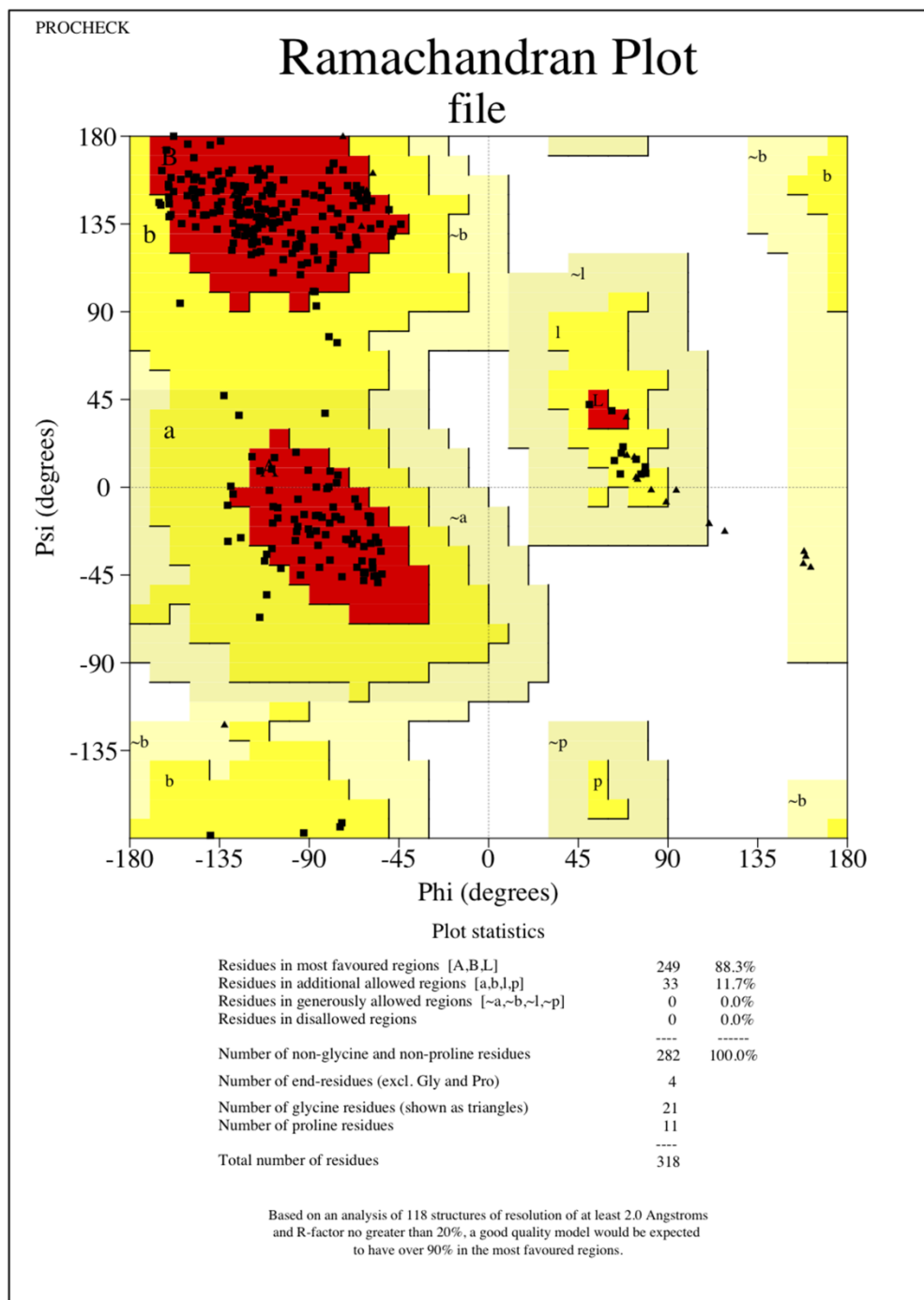




**Appendix Figure 5a: Ramachandran plot for MLL1 *Win* motif peptide**

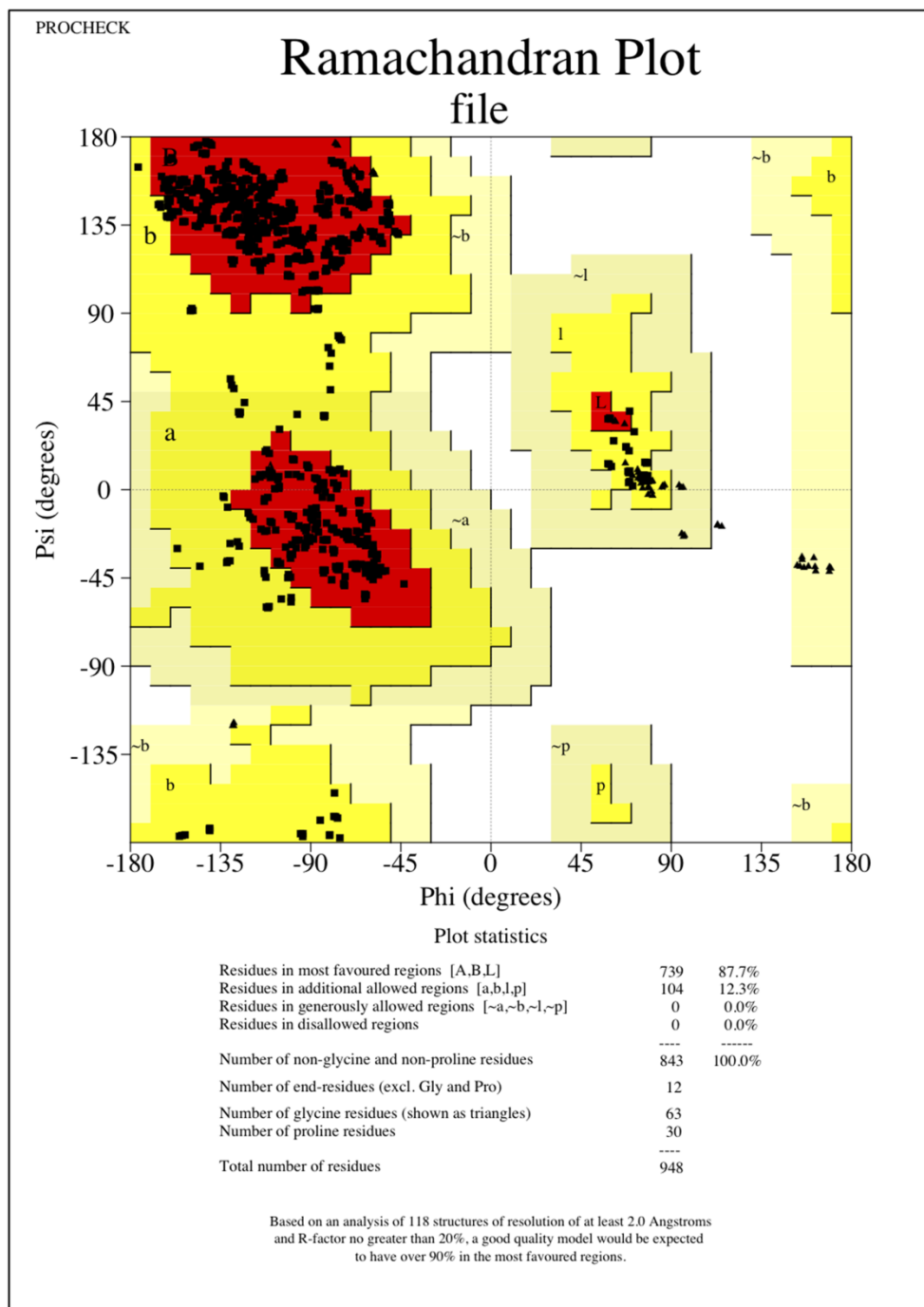


**Appendix Figure 5b: Ramachandran plot for MLL2 *Win* motif peptide**

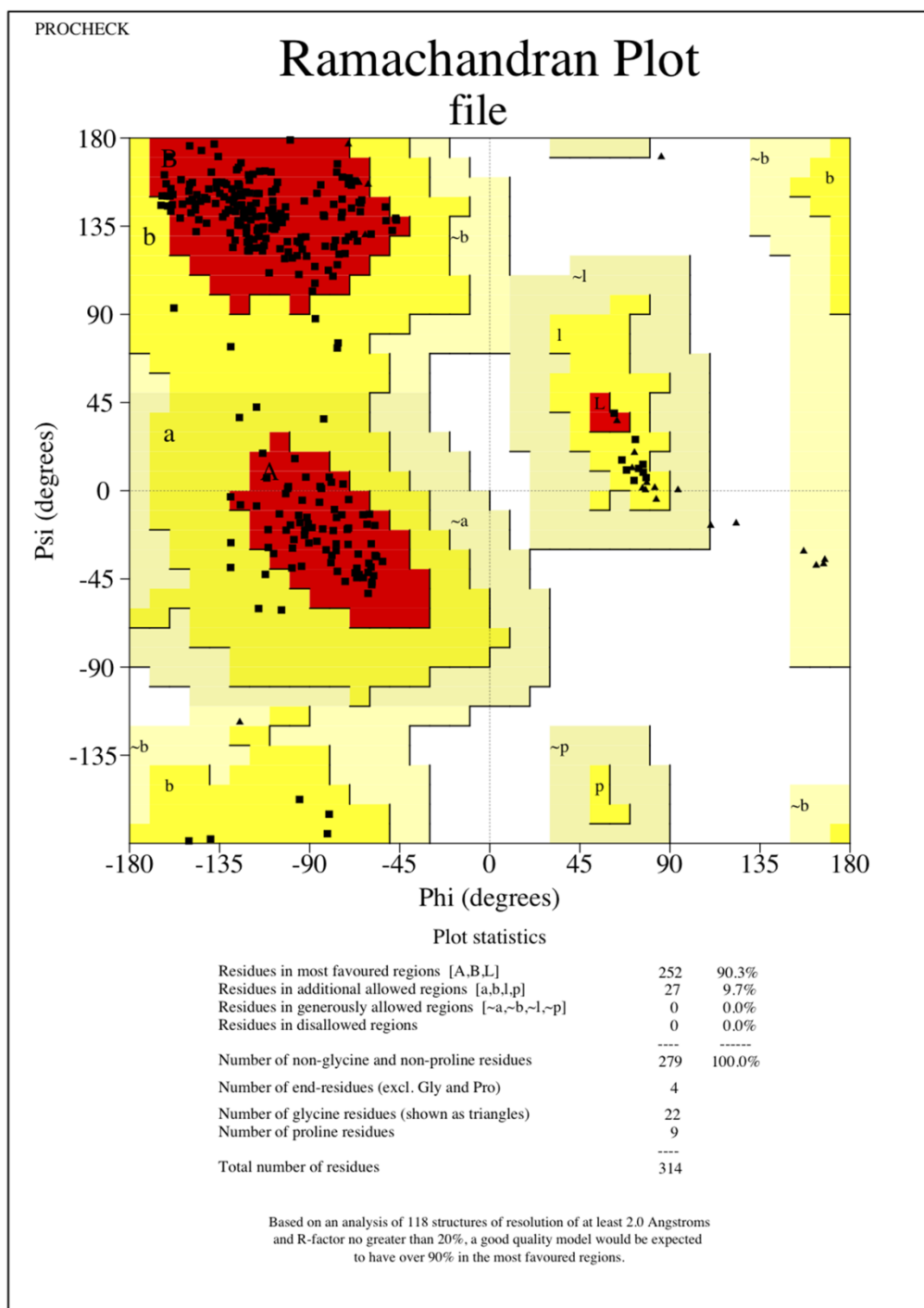


file\_01.ps

**Appendix Figure 5c: Ramachandran plot for MLL3 *Win* motif peptide**

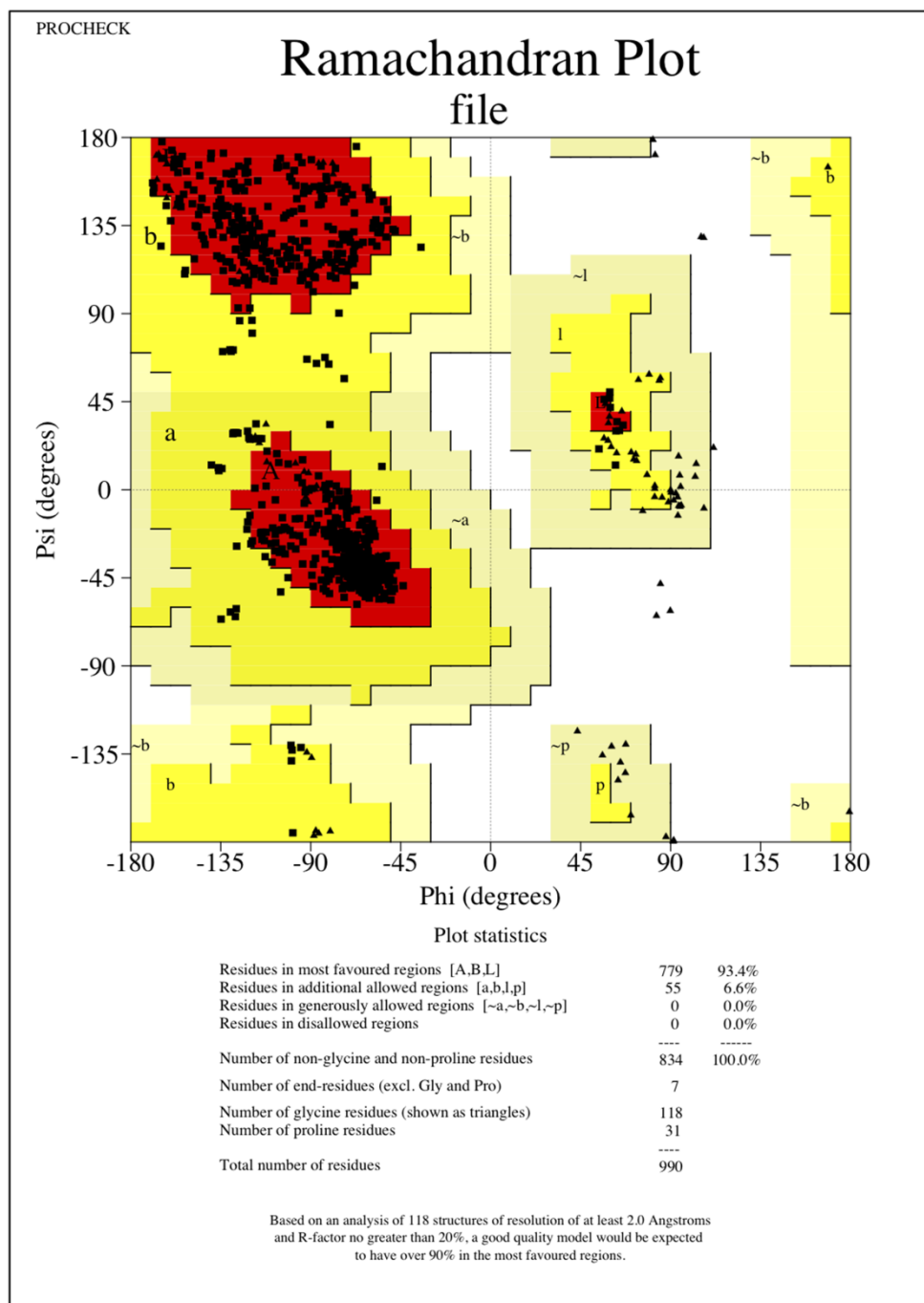


**Appendix Figure 5d: Ramachandran plot for MLL4 *Win* motif peptide**



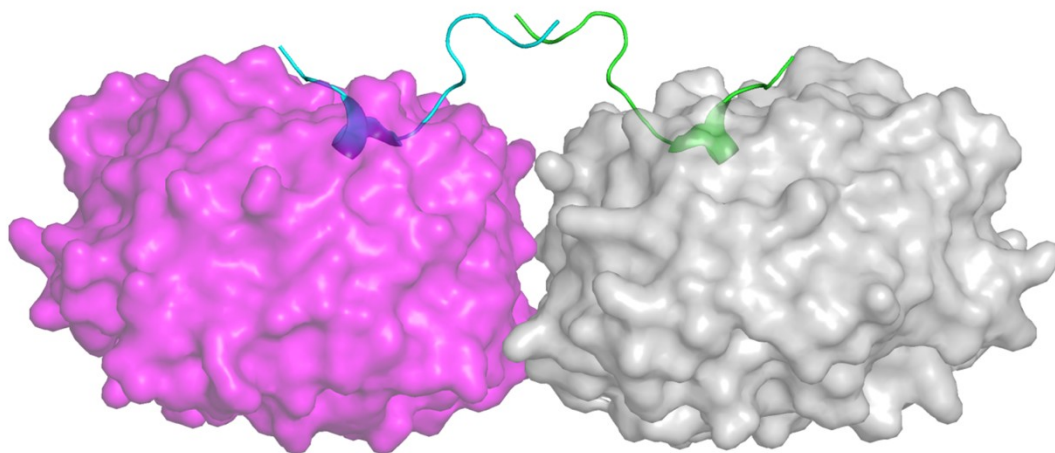
**Appendix Figure 5e: Ramachandran plot for SETd1a *Win* motif peptide**





file\_01.ps

**Appendix Figure 5f: Ramachandran plot for SETd1b *Win* motif peptide**



**Appendix 6: View of the Asymmetric Unit for MLL1 *Win* motif-WDR5 binary complex.**

The WDR5 molecules are shown as a surface representation (Magenta and Grey) and the two MLL1 *Win* motif peptide chains are shown as ribbons (Blue and Green). Structure figure was generated in PyMol.

## REFERENCES

- (1) Daser, A., and Rabbitts, T. H. (2005) The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis. *Semin Cancer Biol* 15, 175-88.
- (2) Ziemer-van der Poel, S., McCabe, N. R., Gill, H. J., Espinosa, R., 3rd, Patel, Y., Harden, A., Rubinelli, P., Smith, S. D., LeBeau, M. M., Rowley, J. D., and et al. (1991) Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. *Proc Natl Acad Sci U S A* 88, 10735-9.
- (3) Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B. D., and Evans, G. A. (1992) A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. *Nat Genet* 2, 113-8.
- (4) Tkachuk, D. C., Kohler, S., and Cleary, M. L. (1992) Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias. *Cell* 71, 691-700.
- (5) Liu, H., Cheng, E. H., and Hsieh, J. J. (2009) MLL fusions: pathways to leukemia. *Cancer Biol Ther* 8, 1204-11.
- (6) Slany, R. K. (2009) The molecular biology of mixed lineage leukemia. *Haematologica* 94, 984-93.
- (7) Dou, Y., and Hess, J. L. (2008) Mechanisms of transcriptional regulation by MLL and its disruption in acute leukemia. *Int J Hematol* 87, 10-8.
- (8) Krivtsov, A. V., and Armstrong, S. A. (2007) MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer* 7, 823-33.
- (9) Bigoni, R., Cuneo, A., Roberti, M. G., Moretti, S., De Angeli, C., Dabusti, M., Campioni, D., del Senno, L., Biondi, A., Chaplin, T., Young, B. D., and Castoldi, G. (1999) Therapy-related adult acute lymphoblastic leukemia with t(4;11)(q21; q23): MLL rearrangement, p53 mutation and multilineage involvement. *Leukemia* 13, 704-7.
- (10) Felix, C. A., Lange, B. J., Hosler, M. R., Fertala, J., and Bjornsti, M. A. (1995) Chromosome band 11q23 translocation breakpoints are DNA topoisomerase II cleavage sites. *Cancer Res* 55, 4287-92.
- (11) Caligiuri, M. A., Schichman, S. A., Strout, M. P., Mrozek, K., Baer, M. R., Frankel, S. R., Barcos, M., Herzig, G. P., Croce, C. M., and Bloomfield, C. D. (1994) Molecular rearrangement of the ALL-1 gene in acute myeloid leukemia without cytogenetic evidence of 11q23 chromosomal translocations. *Cancer Res* 54, 370-3.
- (12) Caligiuri, M. A., Strout, M. P., Lawrence, D., Arthur, D. C., Baer, M. R., Yu, F., Knuutila, S., Mrozek, K., Oberkircher, A. R., Marcucci, G., de la Chapelle, A., Elonen, E., Block, A. W., Rao, P. N., Herzig, G. P., Powell, B. L., Ruutu, T., Schiffer, C. A., and Bloomfield, C. D. (1998) Rearrangement of ALL1 (MLL) in acute myeloid leukemia with normal cytogenetics. *Cancer Res* 58, 55-9.
- (13) Schichman, S. A., Caligiuri, M. A., Strout, M. P., Carter, S. L., Gu, Y., Canaani, E., Bloomfield, C. D., and Croce, C. M. (1994) ALL-1 tandem duplication in acute myeloid leukemia with a normal karyotype involves homologous recombination between Alu elements. *Cancer Res* 54, 4277-80.
- (14) Schichman, S. A., Canaani, E., and Croce, C. M. (1995) Self-fusion of the ALL1 gene. A new genetic mechanism for acute leukemia. *Jama* 273, 571-6.

- (15) Dohner, K., Tobis, K., Ulrich, R., Frohling, S., Benner, A., Schlenk, R. F., and Dohner, H. (2002) Prognostic significance of partial tandem duplications of the MLL gene in adult patients 16 to 60 years old with acute myeloid leukemia and normal cytogenetics: a study of the Acute Myeloid Leukemia Study Group Ulm. *J Clin Oncol* 20, 3254-61.
- (16) Herry, A., Douet-Guilbert, N., Gueganic, N., Morel, F., Le Bris, M. J., Berthou, C., and De Braekeleer, M. (2006) Del(5q) and MLL amplification in homogeneously staining region in acute myeloblastic leukemia: a recurrent cytogenetic association. *Ann Hematol* 85, 244-9.
- (17) Streubel, B., Valent, P., Jager, U., Edelhauser, M., Wandt, H., Wagner, T., Buchner, T., Lechner, K., and Fonatsch, C. (2000) Amplification of the MLL gene on double minutes, a homogeneously staining region, and ring chromosomes in five patients with acute myeloid leukemia or myelodysplastic syndrome. *Genes Chromosomes Cancer* 27, 380-6.
- (18) Avet-Loiseau, H., Godon, C., Li, J. Y., Daviet, A., Mellerin, M. P., Talmant, P., Harousseau, J. L., and Bataille, R. (1999) Amplification of the 11q23 region in acute myeloid leukemia. *Genes Chromosomes Cancer* 26, 166-70.
- (19) Allen, R. J., Smith, S. D., Moldwin, R. L., Lu, M. M., Giordano, L., Vignon, C., Suto, Y., Harden, A., Tomek, R., Veldman, T., Ried, T., Larson, R. A., Le Beau, M. M., Rowley, J. D., and Zeleznik-Le, N. (1998) Establishment and characterization of a megakaryoblast cell line with amplification of MLL. *Leukemia* 12, 1119-27.
- (20) Ariyama, Y., Fukuda, Y., Okuno, Y., Seto, M., Date, K., Abe, T., Nakamura, Y., and Inazawa, J. (1998) Amplification on double-minute chromosomes and partial-tandem duplication of the MLL gene in leukemic cells of a patient with acute myelogenous leukemia. *Genes Chromosomes Cancer* 23, 267-72.
- (21) Cuthbert, G., McCullough, S., Finney, R., Breese, G., and Bown, N. (1999) Jumping translocation at 11q23 with MLL gene rearrangement and interstitial telomeric sequences. *Genes Chromosomes Cancer* 24, 295-8.
- (22) Poppe, B., Vandesompele, J., Schoch, C., Lindvall, C., Mrozek, K., Bloomfield, C. D., Beverloo, H. B., Michaux, L., Dastugue, N., Herens, C., Yigit, N., De Paepe, A., Hagemeijer, A., and Speleman, F. (2004) Expression analyses identify MLL as a prominent target of 11q23 amplification and support an etiologic role for MLL gain of function in myeloid malignancies. *Blood* 103, 229-35.
- (23) Liedtke, M., and Cleary, M. L. (2009) Therapeutic targeting of MLL. *Blood* 113, 6061-8.
- (24) Altucci, L., Clarke, N., Nebbioso, A., Scognamiglio, A., and Gronemeyer, H. (2005) Acute myeloid leukemia: therapeutic impact of epigenetic drugs. *Int J Biochem Cell Biol* 37, 1752-62.
- (25) Downing, J. R. (2008) Targeted therapy in leukemia. *Mod Pathol* 21 Suppl 2, S2-7.
- (26) Dillon, S. C., Zhang, X., Trievel, R. C., and Cheng, X. (2005) The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol* 6, 227.
- (27) Cosgrove, M. S. (2006) PHinDing a new histone "effector" domain. *Structure* 14, 1096-8.
- (28) Flanagan, J. F., Mi, L. Z., Chruszcz, M., Cymborowski, M., Clines, K. L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* 438, 1181-5.
- (29) Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R., 3rd, and Grant, P. A. (2005) Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* 433, 434-8.

- (30) Santos-Rosa, H., Schneider, R., Bernstein, B. E., Karabetsou, N., Morillon, A., Weise, C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2003) Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Mol Cell* 12, 1325-32.
- (31) Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M., Tackett, A. J., Chait, B. T., Badenhorst, P., Wu, C., and Allis, C. D. (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 442, 86-90.
- (32) Patel, A., Dharmarajan, V., Vought, V. E., and Cosgrove, M. S. (2009) On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem* 284, 24242-56.
- (33) Patel, A., Vought, V. E., Dharmarajan, V., and Cosgrove, M. S. (2011) A novel non-SET domain multi-subunit methyltransferase required for sequential nucleosomal histone H3 methylation by the mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem* 286, 3359-69.
- (34) Briggs, S. D., Bryk, M., Strahl, B. D., Cheung, W. L., Davie, J. K., Dent, S. Y., Winston, F., and Allis, C. D. (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev* 15, 3286-95.
- (35) Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39, 311-8.
- (36) Nislow, C., Ray, E., and Pillus, L. (1997) SET1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes. *Mol Biol Cell* 8, 2421-36.
- (37) Schneider, J., Wood, A., Lee, J. S., Schuster, R., Dueker, J., Maguire, C., Swanson, S. K., Florens, L., Washburn, M. P., and Shilatifard, A. (2005) Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Mol Cell* 19, 849-56.
- (38) Liu, C. L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S. L., Friedman, N., and Rando, O. J. (2005) Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol* 3, e328.
- (39) Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D. K., and Young, R. A. (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122, 517-27.
- (40) Dou, Y., Milne, T. A., Tackett, A. J., Smith, E. R., Fukuda, A., Wysocka, J., Allis, C. D., Chait, B. T., Hess, J. L., and Roeder, R. G. (2005) Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* 121, 873-85.
- (41) Lee, J. H., and Skalnik, D. G. (2005) CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J Biol Chem* 280, 41725-31.
- (42) Lee, J. H., Tate, C. M., You, J. S., and Skalnik, D. G. (2007) Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *J Biol Chem* 282, 13419-28.

- (43) Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J. F., and Shilatifard, A. (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A* 98, 12902-7.
- (44) Nagy, P. L., Griesenbeck, J., Kornberg, R. D., and Cleary, M. L. (2002) A trithorax-group complex purified from *Saccharomyces cerevisiae* is required for methylation of histone H3. *Proc Natl Acad Sci U S A* 99, 90-4.
- (45) Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C. M., and Canaani, E. (2002) ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol Cell* 10, 1119-28.
- (46) Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W. W., Wilm, M., Aasland, R., and Stewart, A. F. (2001) The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *Embo J* 20, 7137-48.
- (47) Steward, M. M., Lee, J. S., O'Donovan, A., Wyatt, M., Bernstein, B. E., and Shilatifard, A. (2006) Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat Struct Mol Biol* 13, 852-4.
- (48) Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D. J., Kitabayashi, I., Herr, W., and Cleary, M. L. (2004) Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol Cell Biol* 24, 5639-49.
- (49) Cao, F., Chen, Y., Cierpicki, T., Liu, Y., Basrur, V., Lei, M., and Dou, Y. An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain. *PLoS One* 5, e14102.
- (50) Takahashi, Y. H., Westfield, G. H., Oleskie, A. N., Trievel, R. C., Shilatifard, A., and Skiniotis, G. (2012) Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. *Proc Natl Acad Sci U S A* 108, 20526-31.
- (51) Dou, Y., Milne, T. A., Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., Allis, C. D., and Roeder, R. G. (2006) Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol* 13, 713-9.
- (52) Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K. R., Aasland, R., Anastassiadis, K., Ang, S. L., and Stewart, A. F. (2006) Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* 133, 1423-32.
- (53) Hughes, C. M., Rozenblatt-Rosen, O., Milne, T. A., Copeland, T. D., Levine, S. S., Lee, J. C., Hayes, D. N., Shanmugam, K. S., Bhattacharjee, A., Biondi, C. A., Kay, G. F., Hayward, N. K., Hess, J. L., and Meyerson, M. (2004) Menin associates with a trithorax family histone methyltransferase complex and with the *hoxc8* locus. *Mol Cell* 13, 587-97.
- (54) Milne, T. A., Briggs, S. D., Brock, H. W., Martin, M. E., Gibbs, D., Allis, C. D., and Hess, J. L. (2002) MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol Cell* 10, 1107-17.
- (55) Prasad, R., Zhadanov, A. B., Sedkov, Y., Bullrich, F., Druck, T., Rallapalli, R., Yano, T., Alder, H., Croce, C. M., Huebner, K., Mazo, A., and Canaani, E. (1997) Structure and expression pattern of human ALR, a novel gene with strong homology to ALL-1 involved in acute leukemia and to *Drosophila* trithorax. *Oncogene* 15, 549-60.
- (56) Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N., and Herr, W. (2003) Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are

- tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev* 17, 896-911.
- (57) Cho, Y. W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G. R., Copeland, T. D., Kalkum, M., and Ge, K. (2007) PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem* 282, 20395-406.
- (58) Wang, P., Lin, C., Smith, E. R., Guo, H., Sanderson, B. W., Wu, M., Gogol, M., Alexander, T., Seidel, C., Wiedemann, L. M., Ge, K., Krumlauf, R., and Shilatifard, A. (2009) Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol Cell Biol* 29, 6074-85.
- (59) Wysocka, J., Swigut, T., Milne, T. A., Dou, Y., Zhang, X., Burlingame, A. L., Roeder, R. G., Brivanlou, A. H., and Allis, C. D. (2005) WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* 121, 859-72.
- (60) Patel, A., Dharmarajan, V., and Cosgrove, M. S. (2008) Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. *J Biol Chem* 283, 32158-61.
- (61) Patel, A., Vought, V. E., Dharmarajan, V., and Cosgrove, M. S. (2008) A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. *J Biol Chem* 283, 32162-75.
- (62) Terranova, R., Agherbi, H., Boned, A., Meresse, S., and Djabali, M. (2006) Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc Natl Acad Sci U S A* 103, 6629-34.
- (63) Trievel, R. C., and Shilatifard, A. (2009) WDR5, a complexed protein. *Nat Struct Mol Biol* 16, 678-80.
- (64) Schuetz, A., Allali-Hassani, A., Martin, F., Loppnau, P., Vedadi, M., Bochkarev, A., Plotnikov, A. N., Arrowsmith, C. H., and Min, J. (2006) Structural basis for molecular recognition and presentation of histone H3 by WDR5. *Embo J* 25, 4245-52.
- (65) Han, Z., Guo, L., Wang, H., Shen, Y., Deng, X. W., and Chai, J. (2006) Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. *Mol Cell* 22, 137-44.
- (66) Couture, J. F., Collazo, E., and Trievel, R. C. (2006) Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat Struct Mol Biol* 13, 698-703.
- (67) Ruthenburg, A. J., Wang, W., Graybosch, D. M., Li, H., Allis, C. D., Patel, D. J., and Verdine, G. L. (2006) Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nat Struct Mol Biol* 13, 704-12.
- (68) Song, J. J., and Kingston, R. E. (2008) WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *J Biol Chem* 283, 35258-64.
- (69) Lee, J. H., and Skalnik, D. G. (2008) Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol Cell Biol* 28, 609-18.
- (70) Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-8.

- (71) Henikoff, S., and Henikoff, J. G. (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* 89, 10915-9.
- (72) Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., Beck, A. E., Tabor, H. K., Cooper, G. M., Mefford, H. C., Lee, C., Turner, E. H., Smith, J. D., Rieder, M. J., Yoshiura, K., Matsumoto, N., Ohta, T., Niikawa, N., Nickerson, D. A., Bamshad, M. J., and Shendure, J. (2011) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 42, 790-3.
- (73) Paulussen, A. D., Stegmann, A. P., Blok, M. J., Tserpelis, D., Posma-Velter, C., Detisch, Y., Smeets, E. E., Wagemans, A., Schrandt, J. J., van den Boogaard, M. J., van der Smagt, J., van Haeringen, A., Stolte-Dijkstra, I., Kerstjens-Frederikse, W. S., Mancini, G. M., Wessels, M. W., Hennekam, R. C., Vreeburg, M., Geraedts, J., de Ravel, T., Fryns, J. P., Smeets, H. J., Devriendt, K., and Schrandt-Stumpel, C. T. (2011) MLL2 mutation spectrum in 45 patients with Kabuki syndrome. *Hum Mutat* 32, E2018-25.
- (74) Huntsman, D. G., Chin, S. F., Muleris, M., Batley, S. J., Collins, V. P., Wiedemann, L. M., Aparicio, S., and Caldas, C. (1999) MLL2, the second human homolog of the *Drosophila trithorax* gene, maps to 19q13.1 and is amplified in solid tumor cell lines. *Oncogene* 18, 7975-84.
- (75) FitzGerald, K. T., and Diaz, M. O. (1999) MLL2: A new mammalian member of the *trx/MLL* family of genes. *Genomics* 59, 187-92.
- (76) Quentmeier, H., Reinhardt, J., Zaborski, M., and Drexler, H. G. (2003) MLL partial tandem duplications in acute leukemia cell lines. *Leukemia* 17, 980-1.
- (77) Schichman, S. A., Caligiuri, M. A., Gu, Y., Strout, M. P., Canaani, E., Bloomfield, C. D., and Croce, C. M. (1994) ALL-1 partial duplication in acute leukemia. *Proc Natl Acad Sci U S A* 91, 6236-9.
- (78) Dorrance, A. M., Liu, S., Chong, A., Pulley, B., Nemer, D., Guimond, M., Yuan, W., Chang, D., Whitman, S. P., Marcucci, G., and Caligiuri, M. A. (2008) The Mll partial tandem duplication: differential, tissue-specific activity in the presence or absence of the wild-type allele. *Blood* 112, 2508-11.
- (79) Dorrance, A. M., Liu, S., Yuan, W., Becknell, B., Arnoczky, K. J., Guimond, M., Strout, M. P., Feng, L., Nakamura, T., Yu, L., Rush, L. J., Weinstein, M., Leone, G., Wu, L., Ferketich, A., Whitman, S. P., Marcucci, G., and Caligiuri, M. A. (2006) Mll partial tandem duplication induces aberrant Hox expression in vivo via specific epigenetic alterations. *J Clin Invest* 116, 2707-16.
- (80) Whitman, S. P., Liu, S., Vukosavljevic, T., Rush, L. J., Yu, L., Liu, C., Klisovic, M. I., Maharry, K., Guimond, M., Strout, M. P., Becknell, B., Dorrance, A., Klisovic, R. B., Plass, C., Bloomfield, C. D., Marcucci, G., and Caligiuri, M. A. (2005) The MLL partial tandem duplication: evidence for recessive gain-of-function in acute myeloid leukemia identifies a novel patient subgroup for molecular-targeted therapy. *Blood* 106, 345-52.
- (81) Dharmarajan, V., and Cosgrove, M. S. (2011) *Biochemistry of the Mixed Lineage Leukemia 1 (MLL1) Protein and Targeted Therapies for Associated Leukemia*, Intech, Rijeka.
- (82) Prasad, R., Yano, T., Sorio, C., Nakamura, T., Rallapalli, R., Gu, Y., Leshkowitz, D., Croce, C. M., and Canaani, E. (1995) Domains with transcriptional regulatory activity within the ALL1 and AF4 proteins involved in acute leukemia. *Proc Natl Acad Sci U S A* 92, 12160-4.



- (83) Ernst, P., Wang, J., and Korsmeyer, S. J. (2002) The role of MLL in hematopoiesis and leukemia. *Curr Opin Hematol* 9, 282-7.
- (84) Eguchi, M., Eguchi-Ishimae, M., and Greaves, M. (2004) The small oligomerization domain of gephyrin converts MLL to an oncogene. *Blood* 103, 3876-82.
- (85) Lavau, C., Du, C., Thirman, M., and Zeleznik-Le, N. (2000) Chromatin-related properties of CBP fused to MLL generate a myelodysplastic-like syndrome that evolves into myeloid leukemia. *Embo J* 19, 4655-64.
- (86) Luo, R. T., Lavau, C., Du, C., Simone, F., Polak, P. E., Kawamata, S., and Thirman, M. J. (2001) The elongation domain of ELL is dispensable but its ELL-associated factor 1 interaction domain is essential for MLL-ELL-induced leukemogenesis. *Mol Cell Biol* 21, 5678-87.
- (87) Mitterbauer-Hohendanner, G., and Mannhalter, C. (2004) The biological and clinical significance of MLL abnormalities in haematological malignancies. *Eur J Clin Invest* 34 Suppl 2, 12-24.
- (88) Mueller, D., Garcia-Cuellar, M. P., Bach, C., Buhl, S., Maethner, E., and Slany, R. K. (2009) Misguided transcriptional elongation causes mixed lineage leukemia. *PLoS Biol* 7, e1000249.
- (89) Debernardi, S., Bassini, A., Jones, L. K., Chaplin, T., Linder, B., de Bruijn, D. R., Meese, E., and Young, B. D. (2002) The MLL fusion partner AF10 binds GAS41, a protein that interacts with the human SWI/SNF complex. *Blood* 99, 275-81.
- (90) Yagi, H., Deguchi, K., Aono, A., Tani, Y., Kishimoto, T., and Komori, T. (1998) Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice. *Blood* 92, 108-17.
- (91) Hess, J. L., Yu, B. D., Li, B., Hanson, R., and Korsmeyer, S. J. (1997) Defects in yolk sac hematopoiesis in Mll-null embryos. *Blood* 90, 1799-806.
- (92) Abramovich, C., and Humphries, R. K. (2005) Hox regulation of normal and leukemic hematopoietic stem cells. *Curr Opin Hematol* 12, 210-6.
- (93) Ernst, P., Mabon, M., Davidson, A. J., Zon, L. I., and Korsmeyer, S. J. (2004) An Mll-dependent Hox program drives hematopoietic progenitor expansion. *Curr Biol* 14, 2063-9.
- (94) Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A., and Korsmeyer, S. J. (1995) Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* 378, 505-8.
- (95) Ernst, P., Fisher, J. K., Avery, W., Wade, S., Foy, D., and Korsmeyer, S. J. (2004) Definitive hematopoiesis requires the mixed-lineage leukemia gene. *Dev Cell* 6, 437-43.
- (96) Hanson, R. D., Hess, J. L., Yu, B. D., Ernst, P., van Lohuizen, M., Berns, A., van der Lugt, N. M., Shashikant, C. S., Ruddle, F. H., Seto, M., and Korsmeyer, S. J. (1999) Mammalian Trithorax and polycomb-group homologues are antagonistic regulators of homeotic development. *Proc Natl Acad Sci U S A* 96, 14372-7.
- (97) McMahan, K. A., Hiew, S. Y., Hadjur, S., Veiga-Fernandes, H., Menzel, U., Price, A. J., Kioussis, D., Williams, O., and Brady, H. J. (2007) Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal. *Cell Stem Cell* 1, 338-45.
- (98) Gan, T., Jude, C. D., Zaffuto, K., and Ernst, P. Developmentally induced Mll1 loss reveals defects in postnatal haematopoiesis. *Leukemia* 24, 1732-41.
- (99) Jude, C. D., Climer, L., Xu, D., Artinger, E., Fisher, J. K., and Ernst, P. (2007) Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. *Cell Stem Cell* 1, 324-37.

- (100) Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., and Korsmeyer, S. J. (2002) MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* 30, 41-7.
- (101) Ayton, P. M., and Cleary, M. L. (2003) Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9. *Genes Dev* 17, 2298-307.
- (102) Ferrando, A. A., Armstrong, S. A., Neuberg, D. S., Sallan, S. E., Silverman, L. B., Korsmeyer, S. J., and Look, A. T. (2003) Gene expression signatures in MLL-rearranged T-lineage and B-precursor acute leukemias: dominance of HOX dysregulation. *Blood* 102, 262-8.
- (103) Takeda, S., Chen, D. Y., Westergard, T. D., Fisher, J. K., Rubens, J. A., Sasagawa, S., Kan, J. T., Korsmeyer, S. J., Cheng, E. H., and Hsieh, J. J. (2006) Proteolysis of MLL family proteins is essential for taspase1-orchestrated cell cycle progression. *Genes Dev* 20, 2397-409.
- (104) Milne, T. A., Hughes, C. M., Lloyd, R., Yang, Z., Rozenblatt-Rosen, O., Dou, Y., Schnepf, R. W., Krankel, C., Livolsi, V. A., Gibbs, D., Hua, X., Roeder, R. G., Meyerson, M., and Hess, J. L. (2005) Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proc Natl Acad Sci U S A* 102, 749-54.
- (105) Yamashita, M., Hirahara, K., Shinnakasu, R., Hosokawa, H., Norikane, S., Kimura, M. Y., Hasegawa, A., and Nakayama, T. (2006) Crucial role of MLL for the maintenance of memory T helper type 2 cell responses. *Immunity* 24, 611-22.
- (106) Ansari, K. I., and Mandal, S. S. (2010) Mixed lineage leukemia: roles in gene expression, hormone signaling and mRNA processing. *Febs J* 277, 1790-804.
- (107) Scharf, S., Zech, J., Bursen, A., Schraets, D., Oliver, P. L., Kliem, S., Pfitzner, E., Gillert, E., Dingermann, T., and Marschalek, R. (2007) Transcription linked to recombination: a gene-internal promoter coincides with the recombination hot spot II of the human MLL gene. *Oncogene* 26, 1361-71.
- (108) Luger, K., and Hansen, J. C. (2005) Nucleosome and chromatin fiber dynamics. *Curr Opin Struct Biol* 15, 188-96.
- (109) Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251-60.
- (110) Li, G., and Reinberg, D. (2011) Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev* 21, 175-86.
- (111) Robinson, P. J., Fairall, L., Huynh, V. A., and Rhodes, D. (2006) EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc Natl Acad Sci U S A* 103, 6506-11.
- (112) Cosgrove, M. S., and Wolberger, C. (2005) How does the histone code work? *Biochem Cell Biol* 83, 468-76.
- (113) Cosgrove, M. S., Boeke, J. D., and Wolberger, C. (2004) Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol* 11, 1037-43.
- (114) Kornberg, R. D., and Lorch, Y. (1999) Chromatin-modifying and -remodeling complexes. *Curr Opin Genet Dev* 9, 148-51.
- (115) Strahl, B. D., and Allis, C. D. (2000) The language of covalent histone modifications. *Nature* 403, 41-5.

- (116) Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S., Kouzarides, T., and Schreiber, S. L. (2002) Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc Natl Acad Sci U S A* 99, 8695-700.
- (117) Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., 3rd, Gingeras, T. R., Schreiber, S. L., and Lander, E. S. (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120, 169-81.
- (118) Ng, H. H., Robert, F., Young, R. A., and Struhl, K. (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* 11, 709-19.
- (119) Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002) Active genes are tri-methylated at K4 of histone H3. *Nature* 419, 407-11.
- (120) Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2004) Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol* 6, 73-7.
- (121) Schubeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D. E., O'Neill, L. P., Turner, B. M., Delrow, J., Bell, S. P., and Groudine, M. (2004) The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev* 18, 1263-71.
- (122) Volkel, P., and Angrand, P. O. (2007) The control of histone lysine methylation in epigenetic regulation. *Biochimie* 89, 1-20.
- (123) Jenuwein, T., and Allis, C. D. (2001) Translating the histone code. *Science* 293, 1074-80.
- (124) Noma, K., and Grewal, S. I. (2002) Histone H3 lysine 4 methylation is mediated by Set1 and promotes maintenance of active chromatin states in fission yeast. *Proc Natl Acad Sci U S A* 99 Suppl 4, 16438-45.
- (125) Strahl, B. D., Ohba, R., Cook, R. G., and Allis, C. D. (1999) Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. *Proc Natl Acad Sci U S A* 96, 14967-72.
- (126) Litt, M. D., Simpson, M., Gaszner, M., Allis, C. D., and Felsenfeld, G. (2001) Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science* 293, 2453-5.
- (127) Boggs, B. A., Cheung, P., Heard, E., Spector, D. L., Chinault, A. C., and Allis, C. D. (2002) Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. *Nat Genet* 30, 73-6.
- (128) van Dijk, K., Marley, K. E., Jeong, B. R., Xu, J., Hesson, J., Cerny, R. L., Waterborg, J. H., and Cerutti, H. (2005) Monomethyl histone H3 lysine 4 as an epigenetic mark for silenced euchromatin in Chlamydomonas. *Plant Cell* 17, 2439-53.
- (129) Krogan, N. J., Dover, J., Khorrami, S., Greenblatt, J. F., Schneider, J., Johnston, M., and Shilatifard, A. (2002) COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem* 277, 10753-5.
- (130) Goo, Y. H., Sohn, Y. C., Kim, D. H., Kim, S. W., Kang, M. J., Jung, D. J., Kwak, E., Barlev, N. A., Berger, S. L., Chow, V. T., Roeder, R. G., Azorsa, D. O., Meltzer, P. S., Suh, P. G., Song, E. J., Lee, K. J., Lee, Y. C., and Lee, J. W. (2003) Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. *Mol Cell Biol* 23, 140-9.

- (131) Issaeva, I., Zonis, Y., Rozovskaia, T., Orlovsky, K., Croce, C. M., Nakamura, T., Mazo, A., Eisenbach, L., and Canaani, E. (2007) Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol Cell Biol* 27, 1889-903.
- (132) Sedkov, Y., Cho, E., Petruk, S., Cherbas, L., Smith, S. T., Jones, R. S., Cherbas, P., Canaani, E., Jaynes, J. B., and Mazo, A. (2003) Methylation at lysine 4 of histone H3 in ecdysone-dependent development of *Drosophila*. *Nature* 426, 78-83.
- (133) Daniel, J. A., Santos, M. A., Wang, Z., Zang, C., Schwab, K. R., Jankovic, M., Filsuf, D., Chen, H. T., Gazumyan, A., Yamane, A., Cho, Y. W., Sun, H. W., Ge, K., Peng, W., Nussenzweig, M. C., Casellas, R., Dressler, G. R., Zhao, K., and Nussenzweig, A. (2010) PTIP promotes chromatin changes critical for immunoglobulin class switch recombination. *Science* 329, 917-23.
- (134) Ardehali, M. B., Mei, A., Zobeck, K. L., Caron, M., Lis, J. T., and Kusch, T. (2011) *Drosophila* Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. *Embo J* 30, 2817-28.
- (135) Wu, M., Wang, P. F., Lee, J. S., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008) Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol Cell Biol* 28, 7337-44.
- (136) Qian, C., and Zhou, M. M. (2006) SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell Mol Life Sci* 63, 2755-63.
- (137) Nottke, A., Colaiacovo, M. P., and Shi, Y. (2009) Developmental roles of the histone lysine demethylases. *Development* 136, 879-89.
- (138) Sigrist, C. J., Cerutti, L., de Castro, E., Langendijk-Genevaux, P. S., Bulliard, V., Bairoch, A., and Hulo, N. (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* 38, D161-6.
- (139) Cosgrove, M. S., and Patel, A. (2010) Mixed lineage leukemia: a structure-function perspective of the MLL1 protein. *Febs Journal* 277, 1832-1842.
- (140) Collins, R. E., Tachibana, M., Tamaru, H., Smith, K. M., Jia, D., Zhang, X., Selker, E. U., Shinkai, Y., and Cheng, X. (2005) In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *J Biol Chem* 280, 5563-70.
- (141) Takahashi, Y. H., Lee, J. S., Swanson, S. K., Saraf, A., Florens, L., Washburn, M. P., Trievel, R. C., and Shilatifard, A. (2009) Regulation of H3K4 trimethylation via Cps40 (Spp1) of COMPASS is monoubiquitination independent: implication for a Phe/Tyr switch by the catalytic domain of Set1. *Mol Cell Biol* 29, 3478-86.
- (142) Southall, S. M., Wong, P. S., Odho, Z., Roe, S. M., and Wilson, J. R. (2009) Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol Cell* 33, 181-91.
- (143) Cheng, X., Collins, R. E., and Zhang, X. (2005) Structural and sequence motifs of protein (histone) methylation enzymes. *Annu Rev Biophys Biomol Struct* 34, 267-94.
- (144) Gori, F., and Demay, M. B. (2005) The effects of BIG-3 on osteoblast differentiation are not dependent upon endogenously produced BMPs. *Exp Cell Res* 304, 287-92.
- (145) Gori, F., Friedman, L., and Demay, M. B. (2005) Wdr5, a novel WD repeat protein, regulates osteoblast and chondrocyte differentiation in vivo. *J Musculoskelet Neuronal Interact* 5, 338-9.

- (146) Zhu, E. D., Demay, M. B., and Gori, F. (2008) Wdr5 is essential for osteoblast differentiation. *J Biol Chem* 283, 7361-7.
- (147) Adamson, A. L., and Shearn, A. (1996) Molecular genetic analysis of *Drosophila ash2*, a member of the trithorax group required for imaginal disc pattern formation. *Genetics* 144, 621-33.
- (148) Hsu, D. R., Chuang, P. T., and Meyer, B. J. (1995) DPY-30, a nuclear protein essential early in embryogenesis for *Caenorhabditis elegans* dosage compensation. *Development* 121, 3323-34.
- (149) Hsu, D. R., and Meyer, B. J. (1994) The dpy-30 gene encodes an essential component of the *Caenorhabditis elegans* dosage compensation machinery. *Genetics* 137, 999-1018.
- (150) Tan, C. C., Sindhu, K. V., Li, S., Nishio, H., Stoller, J. Z., Oishi, K., Puttreddy, S., Lee, T. J., Epstein, J. A., Walsh, M. J., and Gelb, B. D. (2008) Transcription factor Ap2delta associates with Ash2l and ALR, a trithorax family histone methyltransferase, to activate Hoxc8 transcription. *Proc Natl Acad Sci U S A* 105, 7472-7.
- (151) Rampalli, S., Li, L., Mak, E., Ge, K., Brand, M., Tapscott, S. J., and Dilworth, F. J. (2007) p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation. *Nat Struct Mol Biol* 14, 1150-6.
- (152) Chen, Y., Wan, B., Wang, K. C., Cao, F., Yang, Y., Protacio, A., Dou, Y., Chang, H. Y., and Lei, M. (2011) Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding. *EMBO Rep* 12, 797-803.
- (153) Sarvan, S., Avdic, V., Tremblay, V., Chaturvedi, C. P., Zhang, P., Lanouette, S., Blais, A., Brunzelle, J. S., Brand, M., and Couture, J. F. (2011) Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain. *Nat Struct Mol Biol* 18, 857-9.
- (154) Ang, Y. S., Tsai, S. Y., Lee, D. F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., Wang, J., Rendl, M., Bernstein, E., Schaniel, C., and Lemischka, I. R. (2011) Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* 145, 183-97.
- (155) Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B. R., Protacio, A., Flynn, R. A., Gupta, R. A., Wysocka, J., Lei, M., Dekker, J., Helms, J. A., and Chang, H. Y. (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472, 120-4.
- (156) Luscher-Firzlaff, J., Gawlista, I., Vervoorts, J., Kapelle, K., Braunschweig, T., Walsemann, G., Rodgarkia-Schamberger, C., Schuchlantz, H., Dreschers, S., Kremmer, E., Lilischkis, R., Cerni, C., Wellmann, A., and Luscher, B. (2008) The human trithorax protein hASH2 functions as an oncoprotein. *Cancer Res* 68, 749-58.
- (157) Bralten, L. B., Kloosterhof, N. K., Gravendeel, L. A., Sacchetti, A., Duijm, E. J., Kros, J. M., van den Bent, M. J., Hoogenraad, C. C., Sillevius Smitt, P. A., and French, P. J. (2010) Integrated genomic profiling identifies candidate genes implicated in glioma-genesis and a novel LEO1-SLC12A1 fusion gene. *Genes Chromosomes Cancer* 49, 509-17.
- (158) Yokoyama, A., Kitabayashi, I., Ayton, P. M., Cleary, M. L., and Ohki, M. (2002) Leukemia proto-oncoprotein MLL is proteolytically processed into 2 fragments with opposite transcriptional properties. *Blood* 100, 3710-8.
- (159) Hsieh, J. J., Cheng, E. H., and Korsmeyer, S. J. (2003) Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell* 115, 293-303.

- (160) Hsieh, J. J., Ernst, P., Erdjument-Bromage, H., Tempst, P., and Korsmeyer, S. J. (2003) Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. *Mol Cell Biol* 23, 186-94.
- (161) Tyagi, S., Chabes, A. L., Wysocka, J., and Herr, W. (2007) E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Mol Cell* 27, 107-19.
- (162) Xia, Z. B., Anderson, M., Diaz, M. O., and Zeleznik-Le, N. J. (2003) MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. *Proc Natl Acad Sci U S A* 100, 8342-7.
- (163) Ernst, P., Wang, J., Huang, M., Goodman, R. H., and Korsmeyer, S. J. (2001) MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. *Mol Cell Biol* 21, 2249-58.
- (164) Rozenblatt-Rosen, O., Rozovskaia, T., Burakov, D., Sedkov, Y., Tillib, S., Blechman, J., Nakamura, T., Croce, C. M., Mazo, A., and Canaani, E. (1998) The C-terminal SET domains of ALL-1 and TRITHORAX interact with the INI1 and SNR1 proteins, components of the SWI/SNF complex. *Proc Natl Acad Sci U S A* 95, 4152-7.
- (165) Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C. M., and Canaani, E. (1992) The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to *Drosophila trithorax*, to the AF-4 gene. *Cell* 71, 701-8.
- (166) Huret, J. L., Dessen, P., and Bernheim, A. (2001) An atlas of chromosomes in hematological malignancies. Example: 11q23 and MLL partners. *Leukemia* 15, 987-9.
- (167) Schoch, C., Schnittger, S., Klaus, M., Kern, W., Hiddemann, W., and Haferlach, T. (2003) AML with 11q23/MLL abnormalities as defined by the WHO classification: incidence, partner chromosomes, FAB subtype, age distribution, and prognostic impact in an unselected series of 1897 cytogenetically analyzed AML cases. *Blood* 102, 2395-402.
- (168) Burmeister, T., Meyer, C., Schwartz, S., Hofmann, J., Molkentin, M., Kowarz, E., Schneider, B., Raff, T., Reinhardt, R., Gokbuget, N., Hoelzer, D., Thiel, E., and Marschalek, R. (2009) The MLL recombinome of adult CD10-negative B-cell precursor acute lymphoblastic leukemia: results from the GMALL study group. *Blood* 113, 4011-5.
- (169) Meyer, C., Kowarz, E., Hofmann, J., Renneville, A., Zuna, J., Trka, J., Ben Abdelali, R., Macintyre, E., De Braekeleer, E., De Braekeleer, M., Delabesse, E., de Oliveira, M. P., Cave, H., Clappier, E., van Dongen, J. J., Balgobind, B. V., van den Heuvel-Eibrink, M. M., Beverloo, H. B., Panzer-Grumayer, R., Teigler-Schlegel, A., Harbott, J., Kjeldsen, E., Schnittger, S., Koehl, U., Gruhn, B., Heidenreich, O., Chan, L. C., Yip, S. F., Krzywinski, M., Eckert, C., Moricke, A., Schrappe, M., Alonso, C. N., Schafer, B. W., Krauter, J., Lee, D. A., Zur Stadt, U., Te Kronnie, G., Sutton, R., Izraeli, S., Trakhtenbrot, L., Lo Nigro, L., Tsauro, G., Fechina, L., Szczepanski, T., Strehl, S., Ilencikova, D., Molkentin, M., Burmeister, T., Dingermann, T., Klingebiel, T., and Marschalek, R. (2009) New insights to the MLL recombinome of acute leukemias. *Leukemia* 23, 1490-9.
- (170) Meyer, C., Schneider, B., Jakob, S., Strehl, S., Attarbaschi, A., Schnittger, S., Schoch, C., Jansen, M. W., van Dongen, J. J., den Boer, M. L., Pieters, R., Ennas, M. G., Angelucci, E., Koehl, U., Greil, J., Griesinger, F., Zur Stadt, U., Eckert, C., Szczepanski, T., Niggli, F. K., Schafer, B. W., Kempinski, H., Brady, H. J., Zuna, J., Trka, J., Nigro, L. L., Biondi, A., Delabesse, E., Macintyre, E., Stanulla, M., Schrappe, M., Haas, O. A., Burmeister, T.,

- Dingermann, T., Klingebiel, T., and Marschalek, R. (2006) The MLL recombinome of acute leukemias. *Leukemia* 20, 777-84.
- (171) Wang, J., Iwasaki, H., Krivtsov, A., Febbo, P. G., Thorner, A. R., Ernst, P., Anastasiadou, E., Kutok, J. L., Kogan, S. C., Zinkel, S. S., Fisher, J. K., Hess, J. L., Golub, T. R., Armstrong, S. A., Akashi, K., and Korsmeyer, S. J. (2005) Conditional MLL-CBP targets GMP and models therapy-related myeloproliferative disease. *Embo J* 24, 368-81.
- (172) So, C. W., Karsunky, H., Passegue, E., Cozzio, A., Weissman, I. L., and Cleary, M. L. (2003) MLL-GAS7 transforms multipotent hematopoietic progenitors and induces mixed lineage leukemias in mice. *Cancer Cell* 3, 161-71.
- (173) Ida, K., Kitabayashi, I., Taki, T., Taniwaki, M., Noro, K., Yamamoto, M., Ohki, M., and Hayashi, Y. (1997) Adenoviral E1A-associated protein p300 is involved in acute myeloid leukemia with t(11;22)(q23;q13). *Blood* 90, 4699-704.
- (174) Bernard, O. A., Mauchauffe, M., Mecucci, C., Van den Berghe, H., and Berger, R. (1994) A novel gene, AF-1p, fused to HRX in t(1;11)(p32;q23), is not related to AF-4, AF-9 nor ENL. *Oncogene* 9, 1039-45.
- (175) Taki, T., Sako, M., Tsuchida, M., and Hayashi, Y. (1997) The t(11;16)(q23;p13) translocation in myelodysplastic syndrome fuses the MLL gene to the CBP gene. *Blood* 89, 3945-50.
- (176) Dobson, C. L., Warren, A. J., Pannell, R., Forster, A., and Rabbitts, T. H. (2000) Tumorigenesis in mice with a fusion of the leukaemia oncogene Mll and the bacterial lacZ gene. *Embo J* 19, 843-51.
- (177) Hall, P. A., and Russell, S. E. (2004) The pathobiology of the septin gene family. *J Pathol* 204, 489-505.
- (178) Ayton, P. M., Chen, E. H., and Cleary, M. L. (2004) Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. *Mol Cell Biol* 24, 10470-8.
- (179) Macrini, C. M., Pombo-de-Oliveira, M. S., Ford, A. M., and Alves, G. (2003) MLL AT-hook sequence is strongly conserved in infant acute leukemia with or without MLL gene rearrangement. *Leukemia* 17, 1432-3.
- (180) Dobson, C. L., Warren, A. J., Pannell, R., Forster, A., Lavenir, I., Corral, J., Smith, A. J., and Rabbitts, T. H. (1999) The mll-AF9 gene fusion in mice controls myeloproliferation and specifies acute myeloid leukaemogenesis. *Embo J* 18, 3564-74.
- (181) Chen, W., Li, Q., Hudson, W. A., Kumar, A., Kirchhof, N., and Kersey, J. H. (2006) A murine Mll-AF4 knock-in model results in lymphoid and myeloid deregulation and hematologic malignancy. *Blood* 108, 669-77.
- (182) Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V. M., Su, L., Xu, G., and Zhang, Y. (2005) hDOT1L links histone methylation to leukemogenesis. *Cell* 121, 167-78.
- (183) Simone, F., Polak, P. E., Kaberlein, J. J., Luo, R. T., Levitan, D. A., and Thirman, M. J. (2001) EAF1, a novel ELL-associated factor that is delocalized by expression of the MLL-ELL fusion protein. *Blood* 98, 201-9.
- (184) Bitoun, E., Oliver, P. L., and Davies, K. E. (2007) The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. *Hum Mol Genet* 16, 92-106.

- (185) Mueller, D., Bach, C., Zeisig, D., Garcia-Cuellar, M. P., Monroe, S., Sreekumar, A., Zhou, R., Nesvizhskii, A., Chinnaiyan, A., Hess, J. L., and Slany, R. K. (2007) A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. *Blood* 110, 4445-54.
- (186) Sobulo, O. M., Borrow, J., Tomek, R., Reshmi, S., Harden, A., Schlegelberger, B., Housman, D., Doggett, N. A., Rowley, J. D., and Zeleznik-Le, N. J. (1997) 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* 94, 8732-7.
- (187) Cheung, N., Chan, L. C., Thompson, A., Cleary, M. L., and So, C. W. (2007) Protein arginine-methyltransferase-dependent oncogenesis. *Nat Cell Biol* 9, 1208-15.
- (188) Nie, Z., Yan, Z., Chen, E. H., Sechi, S., Ling, C., Zhou, S., Xue, Y., Yang, D., Murray, D., Kanakubo, E., Cleary, M. L., and Wang, W. (2003) Novel SWI/SNF chromatin-remodeling complexes contain a mixed-lineage leukemia chromosomal translocation partner. *Mol Cell Biol* 23, 2942-52.
- (189) Schreiner, S. A., Garcia-Cuellar, M. P., Fey, G. H., and Slany, R. K. (1999) The leukemogenic fusion of MLL with ENL creates a novel transcriptional transactivator. *Leukemia* 13, 1525-33.
- (190) Martin, M. E., Milne, T. A., Bloyer, S., Galoian, K., Shen, W., Gibbs, D., Brock, H. W., Slany, R., and Hess, J. L. (2003) Dimerization of MLL fusion proteins immortalizes hematopoietic cells. *Cancer Cell* 4, 197-207.
- (191) Rozovskaia, T., Feinstein, E., Mor, O., Foa, R., Blechman, J., Nakamura, T., Croce, C. M., Cimino, G., and Canaani, E. (2001) Upregulation of Meis1 and HoxA9 in acute lymphocytic leukemias with the t(4 : 11) abnormality. *Oncogene* 20, 874-8.
- (192) Yeoh, E. J., Ross, M. E., Shurtleff, S. A., Williams, W. K., Patel, D., Mahfouz, R., Behm, F. G., Raimondi, S. C., Relling, M. V., Patel, A., Cheng, C., Campana, D., Wilkins, D., Zhou, X., Li, J., Liu, H., Pui, C. H., Evans, W. E., Naeve, C., Wong, L., and Downing, J. R. (2002) Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1, 133-43.
- (193) Zeisig, B. B., Milne, T., Garcia-Cuellar, M. P., Schreiner, S., Martin, M. E., Fuchs, U., Borkhardt, A., Chanda, S. K., Walker, J., Soden, R., Hess, J. L., and Slany, R. K. (2004) Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization. *Mol Cell Biol* 24, 617-28.
- (194) Hess, J. L., Bittner, C. B., Zeisig, D. T., Bach, C., Fuchs, U., Borkhardt, A., Frampton, J., and Slany, R. K. (2006) c-Myb is an essential downstream target for homeobox-mediated transformation of hematopoietic cells. *Blood* 108, 297-304.
- (195) Kroon, E., Kros, J., Thorsteinsdottir, U., Baban, S., Buchberg, A. M., and Sauvageau, G. (1998) Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *Embo J* 17, 3714-25.
- (196) Kumar, A. R., Hudson, W. A., Chen, W., Nishiuchi, R., Yao, Q., and Kersey, J. H. (2004) Hoxa9 influences the phenotype but not the incidence of Mll-AF9 fusion gene leukemia. *Blood* 103, 1823-8.
- (197) So, C. W., Karsunky, H., Wong, P., Weissman, I. L., and Cleary, M. L. (2004) Leukemic transformation of hematopoietic progenitors by MLL-GAS7 in the absence of Hoxa7 or Hoxa9. *Blood* 103, 3192-9.



- (198) Yocum, A. K., Busch, C. M., Felix, C. A., and Blair, I. A. (2006) Proteomics-based strategy to identify biomarkers and pharmacological targets in leukemias with t(4;11) translocations. *J Proteome Res* 5, 2743-53.
- (199) Brown, P., Levis, M., Shurtleff, S., Campana, D., Downing, J., and Small, D. (2005) FLT3 inhibition selectively kills childhood acute lymphoblastic leukemia cells with high levels of FLT3 expression. *Blood* 105, 812-20.
- (200) Armstrong, S. A., Kung, A. L., Mabon, M. E., Silverman, L. B., Stam, R. W., Den Boer, M. L., Pieters, R., Kersey, J. H., Sallan, S. E., Fletcher, J. A., Golub, T. R., Griffin, J. D., and Korsmeyer, S. J. (2003) Inhibition of FLT3 in MLL. Validation of a therapeutic target identified by gene expression based classification. *Cancer Cell* 3, 173-83.
- (201) Carnicer, M. J., Nomdedeu, J. F., Lasa, A., Estivill, C., Brunet, S., Aventin, A., and Sierra, J. (2004) FLT3 mutations are associated with other molecular lesions in AML. *Leuk Res* 28, 19-23.
- (202) Yao, Q., Nishiuchi, R., Kitamura, T., and Kersey, J. H. (2005) Human leukemias with mutated FLT3 kinase are synergistically sensitive to FLT3 and Hsp90 inhibitors: the key role of the STAT5 signal transduction pathway. *Leukemia* 19, 1605-12.
- (203) Yasui, H., Hideshima, T., Hamasaki, M., Roccaro, A. M., Shiraishi, N., Kumar, S., Tassone, P., Ishitsuka, K., Raje, N., Tai, Y. T., Podar, K., Chauhan, D., Leoni, L. M., Kanekal, S., Elliott, G., Munshi, N. C., and Anderson, K. C. (2005) SDX-101, the R-enantiomer of etodolac, induces cytotoxicity, overcomes drug resistance, and enhances the activity of dexamethasone in multiple myeloma. *Blood* 106, 706-12.
- (204) Liang, D. C., Shih, L. Y., Fu, J. F., Li, H. Y., Wang, H. I., Hung, I. J., Yang, C. P., Jaing, T. H., Chen, S. H., and Liu, H. C. (2006) K-Ras mutations and N-Ras mutations in childhood acute leukemias with or without mixed-lineage leukemia gene rearrangements. *Cancer* 106, 950-6.
- (205) Stubbs, M. C., Kim, Y. M., Krivtsov, A. V., Wright, R. D., Feng, Z., Agarwal, J., Kung, A. L., and Armstrong, S. A. (2008) MLL-AF9 and FLT3 cooperation in acute myelogenous leukemia: development of a model for rapid therapeutic assessment. *Leukemia* 22, 66-77.
- (206) Wang, Z., Smith, K. S., Murphy, M., Piloto, O., Somerville, T. C., and Cleary, M. L. (2008) Glycogen synthase kinase 3 in MLL leukaemia maintenance and targeted therapy. *Nature* 455, 1205-9.
- (207) Oguchi, K., Takagi, M., Tsuchida, R., Taya, Y., Ito, E., Isoyama, K., Ishii, E., Zannini, L., Delia, D., and Mizutani, S. (2003) Missense mutation and defective function of ATM in a childhood acute leukemia patient with MLL gene rearrangement. *Blood* 101, 3622-7.
- (208) Felix, C. A., Megonigal, M. D., Chervinsky, D. S., Leonard, D. G., Tsuchida, N., Kakati, S., Block, A. M., Fisher, J., Grossi, M., Salhany, K. I., Jani-Sait, S. N., and Aplan, P. D. (1998) Association of germline p53 mutation with MLL segmental jumping translocation in treatment-related leukemia. *Blood* 91, 4451-6.
- (209) Mahgoub, N., Parker, R. I., Hosler, M. R., Close, P., Winick, N. J., Masterson, M., Shannon, K. M., and Felix, C. A. (1998) RAS mutations in pediatric leukemias with MLL gene rearrangements. *Genes Chromosomes Cancer* 21, 270-5.
- (210) Taketani, T., Taki, T., Sugita, K., Furuichi, Y., Ishii, E., Hanada, R., Tsuchida, M., Sugita, K., Ida, K., and Hayashi, Y. (2004) FLT3 mutations in the activation loop of tyrosine kinase domain are frequently found in infant ALL with MLL rearrangements and pediatric ALL with hyperdiploidy. *Blood* 103, 1085-8.

- (211) Stam, R. W., den Boer, M. L., Passier, M. M., Janka-Schaub, G. E., Sallan, S. E., Armstrong, S. A., and Pieters, R. (2006) Silencing of the tumor suppressor gene FHIT is highly characteristic for MLL gene rearranged infant acute lymphoblastic leukemia. *Leukemia* 20, 264-71.
- (212) Ross, M. E., Mahfouz, R., Onciu, M., Liu, H. C., Zhou, X., Song, G., Shurtleff, S. A., Pounds, S., Cheng, C., Ma, J., Ribeiro, R. C., Rubnitz, J. E., Girtman, K., Williams, W. K., Raimondi, S. C., Liang, D. C., Shih, L. Y., Pui, C. H., and Downing, J. R. (2004) Gene expression profiling of pediatric acute myelogenous leukemia. *Blood* 104, 3679-87.
- (213) Whitman, S. P., Hackanson, B., Liyanarachchi, S., Liu, S., Rush, L. J., Maharry, K., Margeson, D., Davuluri, R., Wen, J., Witte, T., Yu, L., Liu, C., Bloomfield, C. D., Marcucci, G., Plass, C., and Caligiuri, M. A. (2008) DNA hypermethylation and epigenetic silencing of the tumor suppressor gene, SLC5A8, in acute myeloid leukemia with the MLL partial tandem duplication. *Blood* 112, 2013-6.
- (214) Ganapathy, V., Gopal, E., Miyauchi, S., and Prasad, P. D. (2005) Biological functions of SLC5A8, a candidate tumour suppressor. *Biochem Soc Trans* 33, 237-40.
- (215) Gupta, N., Martin, P. M., Prasad, P. D., and Ganapathy, V. (2006) SLC5A8 (SMCT1)-mediated transport of butyrate forms the basis for the tumor suppressive function of the transporter. *Life Sci* 78, 2419-25.
- (216) Hess, J. L. (2004) MLL: a histone methyltransferase disrupted in leukemia. *Trends Mol Med* 10, 500-7.
- (217) Chen, Y. X., Yan, J., Keeshan, K., Tubbs, A. T., Wang, H., Silva, A., Brown, E. J., Hess, J. L., Pear, W. S., and Hua, X. (2006) The tumor suppressor menin regulates hematopoiesis and myeloid transformation by influencing Hox gene expression. *Proc Natl Acad Sci U S A* 103, 1018-23.
- (218) Yokoyama, A., Somerville, T. C., Smith, K. S., Rozenblatt-Rosen, O., Meyerson, M., and Cleary, M. L. (2005) The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell* 123, 207-18.
- (219) Larsson, C., Skogseid, B., Oberg, K., Nakamura, Y., and Nordenskjold, M. (1988) Multiple endocrine neoplasia type 1 gene maps to chromosome 11 and is lost in insulinoma. *Nature* 332, 85-7.
- (220) Chandrasekharappa, S. C., Guru, S. C., Manickam, P., Olufemi, S. E., Collins, F. S., Emmert-Buck, M. R., Debelenko, L. V., Zhuang, Z., Lubensky, I. A., Liotta, L. A., Crabtree, J. S., Wang, Y., Roe, B. A., Weisemann, J., Boguski, M. S., Agarwal, S. K., Kester, M. B., Kim, Y. S., Heppner, C., Dong, Q., Spiegel, A. M., Burns, A. L., and Marx, S. J. (1997) Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science* 276, 404-7.
- (221) Chandrasekharappa, S. C., and Teh, B. T. (2001) Clinical and molecular aspects of multiple endocrine neoplasia type 1. *Front Horm Res* 28, 50-80.
- (222) Thakker, R. V. (2001) Multiple endocrine neoplasia. *Horm Res* 56 Suppl 1, 67-72.
- (223) Lemmens, I., Van de Ven, W. J., Kas, K., Zhang, C. X., Giraud, S., Wautot, V., Buisson, N., De Witte, K., Salandre, J., Lenoir, G., Pugeat, M., Calender, A., Parente, F., Quincey, D., Gaudray, P., De Wit, M. J., Lips, C. J., Hoppener, J. W., Khodaei, S., Grant, A. L., Weber, G., Kytola, S., Teh, B. T., Farnebo, F., Thakker, R. V., and et al. (1997) Identification of the multiple endocrine neoplasia type 1 (MEN1) gene. The European Consortium on MEN1. *Hum Mol Genet* 6, 1177-83.

- (224) Dong, Q., Debelenko, L. V., Chandrasekharappa, S. C., Emmert-Buck, M. R., Zhuang, Z., Guru, S. C., Manickam, P., Skarulis, M., Lubensky, I. A., Liotta, L. A., Collins, F. S., Marx, S. J., and Spiegel, A. M. (1997) Loss of heterozygosity at 11q13: analysis of pituitary tumors, lung carcinoids, lipomas, and other uncommon tumors in subjects with familial multiple endocrine neoplasia type 1. *J Clin Endocrinol Metab* 82, 1416-20.
- (225) Crabtree, J. S., Scacheri, P. C., Ward, J. M., Garrett-Beal, L., Emmert-Buck, M. R., Edgemon, K. A., Lorang, D., Libutti, S. K., Chandrasekharappa, S. C., Marx, S. J., Spiegel, A. M., and Collins, F. S. (2001) A mouse model of multiple endocrine neoplasia, type 1, develops multiple endocrine tumors. *Proc Natl Acad Sci U S A* 98, 1118-23.
- (226) Stewart, C., Parente, F., Piehl, F., Farnebo, F., Quincey, D., Silins, G., Bergman, L., Carle, G. F., Lemmens, I., Grimmond, S., Xian, C. Z., Khodei, S., Teh, B. T., Lagercrantz, J., Siggers, P., Calender, A., Van de Vem, V., Kas, K., Weber, G., Hayward, N., Gaudray, P., and Larsson, C. (1998) Characterization of the mouse Men1 gene and its expression during development. *Oncogene* 17, 2485-93.
- (227) Bertolino, P., Radovanovic, I., Casse, H., Aguzzi, A., Wang, Z. Q., and Zhang, C. X. (2003) Genetic ablation of the tumor suppressor menin causes lethality at mid-gestation with defects in multiple organs. *Mech Dev* 120, 549-60.
- (228) Bertolino, P., Tong, W. M., Galendo, D., Wang, Z. Q., and Zhang, C. X. (2003) Heterozygous Men1 mutant mice develop a range of endocrine tumors mimicking multiple endocrine neoplasia type 1. *Mol Endocrinol* 17, 1880-92.
- (229) Chandrasekharappa, S. C., and Teh, B. T. (2003) Functional studies of the MEN1 gene. *J Intern Med* 253, 606-15.
- (230) Khodaei, S., O'Brien, K. P., Dumanski, J., Wong, F. K., and Weber, G. (1999) Characterization of the MEN1 ortholog in zebrafish. *Biochem Biophys Res Commun* 264, 404-8.
- (231) Guru, S. C., Crabtree, J. S., Brown, K. D., Dunn, K. J., Manickam, P., Prasad, N. B., Wangsa, D., Burns, A. L., Spiegel, A. M., Marx, S. J., Pavan, W. J., Collins, F. S., and Chandrasekharappa, S. C. (1999) Isolation, genomic organization, and expression analysis of Men1, the murine homolog of the MEN1 gene. *Mamm Genome* 10, 592-6.
- (232) Guru, S. C., Prasad, N. B., Shin, E. J., Hemavathy, K., Lu, J., Ip, Y. T., Agarwal, S. K., Marx, S. J., Spiegel, A. M., Collins, F. S., Oliver, B., and Chandrasekharappa, S. C. (2001) Characterization of a MEN1 ortholog from *Drosophila melanogaster*. *Gene* 263, 31-8.
- (233) Manickam, P., Vogel, A. M., Agarwal, S. K., Oda, T., Spiegel, A. M., Marx, S. J., Collins, F. S., Weinstein, B. M., and Chandrasekharappa, S. C. (2000) Isolation, characterization, expression and functional analysis of the zebrafish ortholog of MEN1. *Mamm Genome* 11, 448-54.
- (234) Poisson, A., Zablewska, B., and Gaudray, P. (2003) Menin interacting proteins as clues toward the understanding of multiple endocrine neoplasia type 1. *Cancer Lett* 189, 1-10.
- (235) Guru, S. C., Goldsmith, P. K., Burns, A. L., Marx, S. J., Spiegel, A. M., Collins, F. S., and Chandrasekharappa, S. C. (1998) Menin, the product of the MEN1 gene, is a nuclear protein. *Proc Natl Acad Sci U S A* 95, 1630-4.
- (236) Balogh, K., Racz, K., Patocs, A., and Hunyady, L. (2006) Menin and its interacting proteins: elucidation of menin function. *Trends Endocrinol Metab* 17, 357-64.
- (237) Yokoyama, A., and Cleary, M. L. (2008) Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell* 14, 36-46.

- (238) Jin, S., Zhao, H., Yi, Y., Nakata, Y., Kalota, A., and Gewirtz, A. M. (2010) c-Myb binds MLL through menin in human leukemia cells and is an important driver of MLL-associated leukemogenesis. *J Clin Invest* 120, 593-606.
- (239) Wu, X., and Hua, X. (2008) Menin, histone h3 methyltransferases, and regulation of cell proliferation: current knowledge and perspective. *Curr Mol Med* 8, 805-15.
- (240) Heppner, C., Bilimoria, K. Y., Agarwal, S. K., Kester, M., Whitty, L. J., Guru, S. C., Chandrasekharappa, S. C., Collins, F. S., Spiegel, A. M., Marx, S. J., and Burns, A. L. (2001) The tumor suppressor protein menin interacts with NF-kappaB proteins and inhibits NF-kappaB-mediated transactivation. *Oncogene* 20, 4917-25.
- (241) Agarwal, S. K., Guru, S. C., Heppner, C., Erdos, M. R., Collins, R. M., Park, S. Y., Saggari, S., Chandrasekharappa, S. C., Collins, F. S., Spiegel, A. M., Marx, S. J., and Burns, A. L. (1999) Menin interacts with the AP1 transcription factor JunD and represses JunD-activated transcription. *Cell* 96, 143-52.
- (242) Gobl, A. E., Berg, M., Lopez-Egido, J. R., Oberg, K., Skogseid, B., and Westin, G. (1999) Menin represses JunD-activated transcription by a histone deacetylase-dependent mechanism. *Biochim Biophys Acta* 1447, 51-6.
- (243) Kaji, H., Canaff, L., Lebrun, J. J., Goltzman, D., and Hendy, G. N. (2001) Inactivation of menin, a Smad3-interacting protein, blocks transforming growth factor type beta signaling. *Proc Natl Acad Sci U S A* 98, 3837-42.
- (244) Sowa, H., Kaji, H., Hendy, G. N., Canaff, L., Komori, T., Sugimoto, T., and Chihara, K. (2004) Menin is required for bone morphogenetic protein 2- and transforming growth factor beta-regulated osteoblastic differentiation through interaction with Smads and Runx2. *J Biol Chem* 279, 40267-75.
- (245) La, P., Schnepf, R. W., C, D. P., A, C. S., and Hua, X. (2004) Tumor suppressor menin regulates expression of insulin-like growth factor binding protein 2. *Endocrinology* 145, 3443-50.
- (246) Jin, S., Mao, H., Schnepf, R. W., Sykes, S. M., Silva, A. C., D'Andrea, A. D., and Hua, X. (2003) Menin associates with FANCD2, a protein involved in repair of DNA damage. *Cancer Res* 63, 4204-10.
- (247) Lemmens, I. H., Forsberg, L., Pannett, A. A., Meyen, E., Piehl, F., Turner, J. J., Van de Ven, W. J., Thakker, R. V., Larsson, C., and Kas, K. (2001) Menin interacts directly with the homeobox-containing protein Pem. *Biochem Biophys Res Commun* 286, 426-31.
- (248) Sukhodolets, K. E., Hickman, A. B., Agarwal, S. K., Sukhodolets, M. V., Obungu, V. H., Novotny, E. A., Crabtree, J. S., Chandrasekharappa, S. C., Collins, F. S., Spiegel, A. M., Burns, A. L., and Marx, S. J. (2003) The 32-kilodalton subunit of replication protein A interacts with menin, the product of the MEN1 tumor suppressor gene. *Mol Cell Biol* 23, 493-509.
- (249) Schnepf, R. W., Hou, Z., Wang, H., Petersen, C., Silva, A., Masai, H., and Hua, X. (2004) Functional interaction between tumor suppressor menin and activator of S-phase kinase. *Cancer Res* 64, 6791-6.
- (250) Lopez-Egido, J., Cunningham, J., Berg, M., Oberg, K., Bongcam-Rudloff, E., and Gobl, A. (2002) Menin's interaction with glial fibrillary acidic protein and vimentin suggests a role for the intermediate filament network in regulating menin activity. *Exp Cell Res* 278, 175-83.
- (251) Agarwal, S. K., Novotny, E. A., Crabtree, J. S., Weitzman, J. B., Yaniv, M., Burns, A. L., Chandrasekharappa, S. C., Collins, F. S., Spiegel, A. M., and Marx, S. J. (2003)

- Transcription factor JunD, deprived of menin, switches from growth suppressor to growth promoter. *Proc Natl Acad Sci U S A* 100, 10770-5.
- (252) Ohkura, N., Kishi, M., Tsukada, T., and Yamaguchi, K. (2001) Menin, a gene product responsible for multiple endocrine neoplasia type 1, interacts with the putative tumor metastasis suppressor nm23. *Biochem Biophys Res Commun* 282, 1206-10.
- (253) Dreijerink, K. M., Mulder, K. W., Winkler, G. S., Hoppener, J. W., Lips, C. J., and Timmers, H. T. (2006) Menin links estrogen receptor activation to histone H3K4 trimethylation. *Cancer Res* 66, 4929-35.
- (254) Maillard, I., Chen, Y. X., Friedman, A., Yang, Y., Tubbs, A. T., Shestova, O., Pear, W. S., and Hua, X. (2009) Menin regulates the function of hematopoietic stem cells and lymphoid progenitors. *Blood* 113, 1661-9.
- (255) Maillard, I., and Hess, J. L. (2009) The role of menin in hematopoiesis. *Adv Exp Med Biol* 668, 51-7.
- (256) Sims, R. J., 3rd, Chen, C. F., Santos-Rosa, H., Kouzarides, T., Patel, S. S., and Reinberg, D. (2005) Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J Biol Chem* 280, 41789-92.
- (257) Caslini, C., Yang, Z., El-Osta, M., Milne, T. A., Slany, R. K., and Hess, J. L. (2007) Interaction of MLL amino terminal sequences with menin is required for transformation. *Cancer Res* 67, 7275-83.
- (258) Roudaia, L., and Speck, N. A. (2008) A MENAge a Trois in leukemia. *Cancer Cell* 14, 3-5.
- (259) Grembecka, J., Belcher, A. M., Hartley, T., and Cierpicki, T. (2010) Molecular basis of the mixed lineage leukemia-menin interaction: implications for targeting mixed lineage leukemias. *J Biol Chem* 285, 40690-8.
- (260) Nilson, I., Reichel, M., Ennas, M. G., Greim, R., Knorr, C., Siegler, G., Greil, J., Fey, G. H., and Marschalek, R. (1997) Exon/intron structure of the human AF-4 gene, a member of the AF-4/LAF-4/FMR-2 gene family coding for a nuclear protein with structural alterations in acute leukaemia. *Br J Haematol* 98, 157-69.
- (261) Heerema, N. A., Sather, H. N., Ge, J., Arthur, D. C., Hilden, J. M., Trigg, M. E., and Reaman, G. H. (1999) Cytogenetic studies of infant acute lymphoblastic leukemia: poor prognosis of infants with t(4;11) - a report of the Children's Cancer Group. *Leukemia* 13, 679-86.
- (262) Behm, F. G., Raimondi, S. C., Frestedt, J. L., Liu, Q., Crist, W. M., Downing, J. R., Rivera, G. K., Kersey, J. H., and Pui, C. H. (1996) Rearrangement of the MLL gene confers a poor prognosis in childhood acute lymphoblastic leukemia, regardless of presenting age. *Blood* 87, 2870-7.
- (263) Faderl, S., Kantarjian, H. M., Talpaz, M., and Estrov, Z. (1998) Clinical significance of cytogenetic abnormalities in adult acute lymphoblastic leukemia. *Blood* 91, 3995-4019.
- (264) Pui, C. H., Chessells, J. M., Camitta, B., Baruchel, A., Biondi, A., Boyett, J. M., Carroll, A., Eden, O. B., Evans, W. E., Gadner, H., Harbott, J., Harms, D. O., Harrison, C. J., Harrison, P. L., Heerema, N., Janka-Schaub, G., Kamps, W., Masera, G., Pullen, J., Raimondi, S. C., Richards, S., Riehm, H., Sallan, S., Sather, H., Shuster, J., Silverman, L. B., Valsecchi, M. G., Vilmer, E., Zhou, Y., Gaynon, P. S., and Schrappe, M. (2003) Clinical heterogeneity in childhood acute lymphoblastic leukemia with 11q23 rearrangements. *Leukemia* 17, 700-6.

- (265) Chen, C. S., Hilden, J. M., Frestedt, J., Domer, P. H., Moore, R., Korsmeyer, S. J., and Kersey, J. H. (1993) The chromosome 4q21 gene (AF-4/FEL) is widely expressed in normal tissues and shows breakpoint diversity in t(4;11)(q21;q23) acute leukemia. *Blood* 82, 1080-5.
- (266) Rubnitz, J. E., Link, M. P., Shuster, J. J., Carroll, A. J., Hakami, N., Frankel, L. S., Pullen, D. J., and Cleary, M. L. (1994) Frequency and prognostic significance of HRX rearrangements in infant acute lymphoblastic leukemia: a Pediatric Oncology Group study. *Blood* 84, 570-3.
- (267) Srinivasan, R. S., Nesbit, J. B., Marrero, L., Erfurth, F., LaRussa, V. F., and Hemenway, C. S. (2004) The synthetic peptide PFWT disrupts AF4-AF9 protein complexes and induces apoptosis in t(4;11) leukemia cells. *Leukemia* 18, 1364-72.
- (268) Palermo, C. M., Bennett, C. A., Winters, A. C., and Hemenway, C. S. (2008) The AF4-mimetic peptide, PFWT, induces necrotic cell death in MV4-11 leukemia cells. *Leuk Res* 32, 633-42.
- (269) Bennett, C. A., Winters, A. C., Barretto, N. N., and Hemenway, C. S. (2009) Molecular targeting of MLL-rearranged leukemia cell lines with the synthetic peptide PFWT synergistically enhances the cytotoxic effect of established chemotherapeutic agents. *Leuk Res* 33, 937-47.
- (270) Melko, M., Douguet, D., Bensaid, M., Zongaro, S., Verheggen, C., Gecz, J., and Bardoni, B. (2011) Functional characterization of the AFF (AF4/FMR2) family of RNA-binding proteins: insights into the molecular pathology of FRAXE intellectual disability. *Hum Mol Genet* 20, 1873-85.
- (271) Isnard, P., Core, N., Naquet, P., and Djabali, M. (2000) Altered lymphoid development in mice deficient for the mAF4 proto-oncogene. *Blood* 96, 705-10.
- (272) Morrissey, J., Tkachuk, D. C., Milatovich, A., Francke, U., Link, M., and Cleary, M. L. (1993) A serine/proline-rich protein is fused to HRX in t(4;11) acute leukemias. *Blood* 81, 1124-31.
- (273) Oliver, P. L., Bitoun, E., Clark, J., Jones, E. L., and Davies, K. E. (2004) Mediation of Af4 protein function in the cerebellum by Siah proteins. *Proc Natl Acad Sci U S A* 101, 14901-6.
- (274) Bensaid, M., Melko, M., Bechara, E. G., Davidovic, L., Berretta, A., Catania, M. V., Gecz, J., Lalli, E., and Bardoni, B. (2009) FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic Acids Res* 37, 1269-79.
- (275) Gu, Y., Shen, Y., Gibbs, R. A., and Nelson, D. L. (1996) Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island. *Nat Genet* 13, 109-13.
- (276) von Bergh, A., Gargallo, P., De Prijck, B., Vranckx, H., Marschalek, R., Larripa, I., Kluin, P., Schuurin, E., and Hagemeijer, A. (2001) Cryptic t(4;11) encoding MLL-AF4 due to insertion of 5' MLL sequences in chromosome 4. *Leukemia* 15, 595-600.
- (277) von Bergh, A. R., Beverloo, H. B., Rombout, P., van Wering, E. R., van Weel, M. H., Beverstock, G. C., Kluin, P. M., Slater, R. M., and Schuurin, E. (2002) LAF4, an AF4-related gene, is fused to MLL in infant acute lymphoblastic leukemia. *Genes Chromosomes Cancer* 35, 92-6.
- (278) Su, M. A., Wisotzkey, R. G., and Newfeld, S. J. (2001) A screen for modifiers of decapentaplegic mutant phenotypes identifies lilliputian, the only member of the Fragile-

- X/Burkitt's Lymphoma family of transcription factors in *Drosophila melanogaster*. *Genetics* 157, 717-25.
- (279) Taki, T., Kano, H., Taniwaki, M., Sako, M., Yanagisawa, M., and Hayashi, Y. (1999) AF5q31, a newly identified AF4-related gene, is fused to MLL in infant acute lymphoblastic leukemia with ins(5;11)(q31;q13q23). *Proc Natl Acad Sci U S A* 96, 14535-40.
- (280) Domer, P. H., Fakharzadeh, S. S., Chen, C. S., Jockel, J., Johansen, L., Silverman, G. A., Kersey, J. H., and Korsmeyer, S. J. (1993) Acute mixed-lineage leukemia t(4;11)(q21;q23) generates an MLL-AF4 fusion product. *Proc Natl Acad Sci U S A* 90, 7884-8.
- (281) Ma, C., and Staudt, L. M. (1996) LAF-4 encodes a lymphoid nuclear protein with transactivation potential that is homologous to AF-4, the gene fused to MLL in t(4;11) leukemias. *Blood* 87, 734-45.
- (282) Geetz, J., Gedeon, A. K., Sutherland, G. R., and Mulley, J. C. (1996) Identification of the gene FMR2, associated with FRA3E mental retardation. *Nat Genet* 13, 105-8.
- (283) Bursen, A., Schwabe, K., Ruster, B., Henschler, R., Ruthardt, M., Dinger, T., and Marschalek, R. (2010) The AF4.MLL fusion protein is capable of inducing ALL in mice without requirement of MLL-AF4. *Blood* 115, 3570-9.
- (284) Benedikt, A., Baltruschat, S., Scholz, B., Bursen, A., Arrey, T. N., Meyer, B., Varagnolo, L., Muller, A. M., Karas, M., Dinger, T., and Marschalek, R. (2011) The leukemogenic AF4-MLL fusion protein causes P-TEFb kinase activation and altered epigenetic signatures. *Leukemia* 25, 135-44.
- (285) Estable, M. C., Naghavi, M. H., Kato, H., Xiao, H., Qin, J., Vahlne, A., and Roeder, R. G. (2002) MCEF, the newest member of the AF4 family of transcription factors involved in leukemia, is a positive transcription elongation factor-b-associated protein. *J Biomed Sci* 9, 234-45.
- (286) Erfurth, F., Hemenway, C. S., de Erkenez, A. C., and Domer, P. H. (2004) MLL fusion partners AF4 and AF9 interact at subnuclear foci. *Leukemia* 18, 92-102.
- (287) Mitelman, F., and Heim, S. (1992) Quantitative acute leukemia cytogenetics. *Genes Chromosomes Cancer* 5, 57-66.
- (288) Hemenway, C. S., de Erkenez, A. C., and Gould, G. C. (2001) The polycomb protein MPc3 interacts with AF9, an MLL fusion partner in t(9;11)(p22;q23) acute leukemias. *Oncogene* 20, 3798-805.
- (289) Nakamura, T., Alder, H., Gu, Y., Prasad, R., Canaani, O., Kamada, N., Gale, R. P., Lange, B., Crist, W. M., Nowell, P. C., and et al. (1993) Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs. *Proc Natl Acad Sci U S A* 90, 4631-5.
- (290) Rubnitz, J. E., Morrissey, J., Savage, P. A., and Cleary, M. L. (1994) ENL, the gene fused with HRX in t(11;19) leukemias, encodes a nuclear protein with transcriptional activation potential in lymphoid and myeloid cells. *Blood* 84, 1747-52.
- (291) Welch, M. D., and Drubin, D. G. (1994) A nuclear protein with sequence similarity to proteins implicated in human acute leukemias is important for cellular morphogenesis and actin cytoskeletal function in *Saccharomyces cerevisiae*. *Mol Biol Cell* 5, 617-32.
- (292) Cairns, B. R., Henry, N. L., and Kornberg, R. D. (1996) TFG/TAF30/ANC1, a component of the yeast SWI/SNF complex that is similar to the leukemogenic proteins ENL and AF-9. *Mol Cell Biol* 16, 3308-16.

- (293) Carlson, M., and Laurent, B. C. (1994) The SNF/SWI family of global transcriptional activators. *Curr Opin Cell Biol* 6, 396-402.
- (294) Cote, J., Quinn, J., Workman, J. L., and Peterson, C. L. (1994) Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science* 265, 53-60.
- (295) Doty, R. T., Vanasse, G. J., Disteché, C. M., and Willerford, D. M. (2002) The leukemia-associated gene Mllt1/ENL: characterization of a murine homolog and demonstration of an essential role in embryonic development. *Blood Cells Mol Dis* 28, 407-17.
- (296) Collins, E. C., Appert, A., Ariza-McNaughton, L., Pannell, R., Yamada, Y., and Rabbitts, T. H. (2002) Mouse Af9 is a controller of embryo patterning, like Mll, whose human homologue fuses with Af9 after chromosomal translocation in leukemia. *Mol Cell Biol* 22, 7313-24.
- (297) Srinivasan, R. S., de Erkenez, A. C., and Hemenway, C. S. (2003) The mixed lineage leukemia fusion partner AF9 binds specific isoforms of the BCL-6 corepressor. *Oncogene* 22, 3395-406.
- (298) Stong, R. C., Korsmeyer, S. J., Parkin, J. L., Arthur, D. C., and Kersey, J. H. (1985) Human acute leukemia cell line with the t(4;11) chromosomal rearrangement exhibits B lineage and monocytic characteristics. *Blood* 65, 21-31.
- (299) Lange, B., Valtieri, M., Santoli, D., Caracciolo, D., Mavilio, F., Gemperlein, I., Griffin, C., Emanuel, B., Finan, J., Nowell, P., and et al. (1987) Growth factor requirements of childhood acute leukemia: establishment of GM-CSF-dependent cell lines. *Blood* 70, 192-9.
- (300) Cohen, A., Grunberger, T., Vanek, W., Dube, I. D., Doherty, P. J., Letarte, M., Roifman, C., and Freedman, M. H. (1991) Constitutive expression and role in growth regulation of interleukin-1 and multiple cytokine receptors in a biphenotypic leukemic cell line. *Blood* 78, 94-102.
- (301) Tsuchiya, S., Yamabe, M., Yamaguchi, Y., Kobayashi, Y., Konno, T., and Tada, K. (1980) Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *Int J Cancer* 26, 171-6.
- (302) Matsuo, Y., MacLeod, R. A., Uphoff, C. C., Drexler, H. G., Nishizaki, C., Katayama, Y., Kimura, G., Fujii, N., Omoto, E., Harada, M., and Orita, K. (1997) Two acute monocytic leukemia (AML-M5a) cell lines (MOLM-13 and MOLM-14) with interclonal phenotypic heterogeneity showing MLL-AF9 fusion resulting from an occult chromosome insertion, ins(11;9)(q23;p22p23). *Leukemia* 11, 1469-77.
- (303) Minowada, J., Onuma, T., and Moore, G. E. (1972) Rosette-forming human lymphoid cell lines. I. Establishment and evidence for origin of thymus-derived lymphocytes. *J Natl Cancer Inst* 49, 891-5.
- (304) Koziner, B., Stavnezer, J., and Al-Katib, A. (1985) Surface immunoglobulin light-chain expression by the "common" all cell line REH. *Clin Immunol Immunopathol* 37, 135-41.
- (305) Campo Dell'Orto, M., Banelli, B., Giarin, E., Accordi, B., Trentin, L., Romani, M., te Kronnie, G., and Basso, G. (2007) Down-regulation of DLX3 expression in MLL-AF4 childhood lymphoblastic leukemias is mediated by promoter region hypermethylation. *Oncol Rep* 18, 417-23.
- (306) Wiederschain, D., Kawai, H., Shilatifard, A., and Yuan, Z. M. (2005) Multiple mixed lineage leukemia (MLL) fusion proteins suppress p53-mediated response to DNA damage. *J Biol Chem* 280, 24315-21.



- (307) Nakanishi, H., Nakamura, T., Canaani, E., and Croce, C. M. (2007) ALL1 fusion proteins induce deregulation of EphA7 and ERK phosphorylation in human acute leukemias. *Proc Natl Acad Sci U S A* 104, 14442-7.
- (308) Basccke, J., Whelan, J. T., Griesinger, F., and Bertrand, F. E. (2006) The MLL partial tandem duplication in acute myeloid leukaemia. *Br J Haematol* 135, 438-49.
- (309) Dharmarajan, V., Lee, J. H., Patel, A., Skalnik, D. G., and Cosgrove, M. S. (2012) Structural basis for Win (WDR5 interaction) motif recognition in human SET1 family histone methyltransferases. *J Biol Chem*.
- (310) Karatas, H., Townsend, E. C., Bernard, D., Dou, Y., and Wang, S. (2010) Analysis of the binding of mixed lineage leukemia 1 (MLL1) and histone 3 peptides to WD repeat domain 5 (WDR5) for the design of inhibitors of the MLL1-WDR5 interaction. *J Med Chem* 53, 5179-85.
- (311) Marschalek, R. (2010) Mixed lineage leukemia: roles in human malignancies and potential therapy. *Febs J* 277, 1822-31.
- (312) Okada, Y., Jiang, Q., Lemieux, M., Jeannotte, L., Su, L., and Zhang, Y. (2006) Leukaemic transformation by CALM-AF10 involves upregulation of Hoxa5 by hDOT1L. *Nat Cell Biol* 8, 1017-24.
- (313) Allen, M. D., Grummitt, C. G., Hilcenko, C., Min, S. Y., Tonkin, L. M., Johnson, C. M., Freund, S. M., Bycroft, M., and Warren, A. J. (2006) Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *Embo J* 25, 4503-12.
- (314) Cierpicki, T., Risner, L. E., Grembecka, J., Lukasik, S. M., Popovic, R., Omonkowska, M., Shultis, D. D., Zeleznik-Le, N. J., and Bushweller, J. H. (2010) Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. *Nat Struct Mol Biol* 17, 62-8.
- (315) Shapiro, G. I. (2006) Cyclin-dependent kinase pathways as targets for cancer treatment. *J Clin Oncol* 24, 1770-83.
- (316) Chen, R., Wierda, W. G., Chubb, S., Hawtin, R. E., Fox, J. A., Keating, M. J., Gandhi, V., and Plunkett, W. (2009) Mechanism of action of SNS-032, a novel cyclin-dependent kinase inhibitor, in chronic lymphocytic leukemia. *Blood* 113, 4637-45.
- (317) Cheng, Y., and Prusoff, W. H. (1973) Relationship between the inhibition constant (K<sub>i</sub>) and the concentration of inhibitor which causes 50 per cent inhibition (I<sub>50</sub>) of an enzymatic reaction. *Biochem Pharmacol* 22, 3099-108.
- (318) Dyer, P. N., Edayathumangalam, R. S., White, C. L., Bao, Y., Chakravarthy, S., Muthurajan, U. M., and Luger, K. (2004) Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol* 375, 23-44.
- (319) Flaus, A., Rencurel, C., Ferreira, H., Wiechens, N., and Owen-Hughes, T. (2004) Sin mutations alter inherent nucleosome mobility. *Embo J* 23, 343-53.
- (320) Ohtaka, H., Muzammil, S., Schon, A., Velazquez-Campoy, A., Vega, S., and Freire, E. (2004) Thermodynamic rules for the design of high affinity HIV-1 protease inhibitors with adaptability to mutations and high selectivity towards unwanted targets. *Int J Biochem Cell Biol* 36, 1787-99.
- (321) Velazquez-Campoy, A., Luque, I., Todd, M. J., Milutinovich, M., Kiso, Y., and Freire, E. (2000) Thermodynamic dissection of the binding energetics of KNI-272, a potent HIV-1 protease inhibitor. *Protein Sci* 9, 1801-9.

- (322) Velazquez-Campoy, A., Todd, M. J., and Freire, E. (2000) HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity. *Biochemistry* 39, 2201-7.
- (323) Ward, W. H., and Holdgate, G. A. (2001) Isothermal titration calorimetry in drug discovery. *Prog Med Chem* 38, 309-76.
- (324) W.Minor, Z. O. a. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* 276, 307-326.
- (325) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67, 235-42.
- (326) Vagin , A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst* 30, 1022-1025.
- (327) Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54, 905-21.
- (328) Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47 ( Pt 2), 110-9.
- (329) Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66, 12-21.
- (330) Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35, W375-83.
- (331) Schrodinger, L. (2010).
- (332) Laskowski R A, M. M. W., Moss D S, Thornton J M (1993) PROCHECK - a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26, 283-291.
- (333) Zhang, P., Lee, H., Brunzelle, J. S., and Couture, J. F. (2012) The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Res* 40, 4237-46.
- (334) Pierce, M. M., Raman, C. S., and Nall, B. T. (1999) Isothermal titration calorimetry of protein-protein interactions. *Methods* 19, 213-21.
- (335) Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res* 14, 1188-90.
- (336) Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., and von Mering, C. (2009) STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 37, D412-6.

- (337) von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M. A., and Bork, P. (2005) STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res* 33, D433-7.

**BIOGRAPHICAL DATA**

NAME	Venkatasubramanian Dharmarajan
PLACE OF BIRTH	Nagapatinam, Tamil Nadu, INDIA
ELEMENTARY SCHOOL	Don Bosco Matriculation Higher Secondary School, Thanjavur, Tamil Nadu, INDIA
HIGH SCHOOL	Kamala Subramanian Matriculation Higher Secondary School, Thanjavur, Tamil Nadu, INDIA, Graduated 2000
UNDERGRADUATE (2004)	Anna University, Chennai, Tamil Nadu, INDIA B. Tech, Industrial Biotechnology, 2004
GRADUATE	Department of Biology, Syracuse University, Syracuse, NY Ph. D, Structural Biology, Biophysics, and Biochemistry Program, 2013