2011

The Specific Role of the MLL CXXC Domain in MLL Fusion Protein Function

Laurie Ellen Risner
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Part of the Molecular Biology Commons

Recommended Citation
Risner, Laurie Ellen, "The Specific Role of the MLL CXXC Domain in MLL Fusion Protein Function" (2011). Dissertations. 182.
https://ecommons.luc.edu/luc_diss/182

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 2011 Laurie Ellen Risner
LOYOLA UNIVERSITY CHICAGO

THE SPECIFIC ROLE OF THE MLL CXXC DOMAIN
IN MLL FUSION PROTEIN FUNCTION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

PROGRAM IN MOLECULAR BIOLOGY

BY

LAURIE ELLEN RISNER

CHICAGO, ILLINOIS
AUGUST 2011
ACKNOWLEDGEMENTS

I would like to thank everyone who helped to support me through my graduate school career. First I would like to acknowledge the Loyola University Chicago Biomedical Sciences Graduate School and especially the Molecular Biology Program for allowing me the opportunity to earn my doctorate degree. I appreciate all of the help that Peggy Richied and Donna Buczek in the Molecular Biology office, and Lorelei Hacholski in the Cancer Center have always been so willing to offer. I would like to thank all of the members of my committee, Jiwang Zhang, M.D., Ph.D., Manuel Diaz, M.D., Lucy Godley, M.D., Ph.D., and Charles Hemenway, M.D., Ph.D., as well as all the members of the Leukemia group in the Cancer Center for all of their help and support with my dissertation project.

I would like to acknowledge John Bushweller, Ph.D. and his laboratory at the University of Virginia for their invaluable collaboration with our laboratory. My dissertation work was nicely complemented by the structural biology experiments performed in the Bushweller laboratory by Tomasz Cierpicki, Ph.D., and Jolanta Grembecka, Ph.D. I also appreciate all of the help from Kelly Schoenfelt, in the Department of Physiology at Loyola, and Aravinda Kuntimaddi, in the Bushweller laboratory, in collaborating with me to perform the protein purifications and DNA-protein binding experiments.
Thank you to Pat Simms in the FACS core for all her guidance and assistance with my flow cytometry experiments, and MaryKay Olsen for providing me with excellent tissue histology samples.

I am very grateful to my mentor, Nancy Zeleznik-Le, Ph.D., for taking me into her laboratory and teaching me how to be a good scientist. Nancy has been an exceptional role model for me. Her consistent encouragement and optimism have helped me to succeed in graduate school, and I know I will take all of what I have learned from her with me into the rest of my career.

I would also like to thank all of the members of the Zeleznik-Le laboratory that I have had the privilege to work with the past few years. When I joined the lab, Cassie, Donna, Relja, Xia and Frank were very welcoming and gracious with teaching me many things about the laboratory and experiments, and with helping me to start my project. The current members of the lab including Nick, Noah, Alyson, Yousaf, Sonia and Shubin have also been a pleasure to work with. I appreciate how all of our lab members have always been so willing to help each other in lab and support each other as friends.

Finally, thank you to my family. My parents have been a constant source of support and encouragement for me. They have always encouraged me to aim high and pursue my dreams, and I know I would not have achieved so much without them. My husband Artie has also been a wonderful source of moral support for me throughout graduate school and I have greatly appreciated his encouragement every time I needed it.
To my family
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xii</td>
</tr>
<tr>
<td>CHAPTER 1: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2: Literature Review</td>
<td>5</td>
</tr>
<tr>
<td>MLL Leukemia</td>
<td>5</td>
</tr>
<tr>
<td>MLL Protein Domains</td>
<td>7</td>
</tr>
<tr>
<td>MLL Function</td>
<td>9</td>
</tr>
<tr>
<td>MLL Fusion Proteins</td>
<td>12</td>
</tr>
<tr>
<td>MLL Gene Targets</td>
<td>14</td>
</tr>
<tr>
<td>HOX Genes</td>
<td>15</td>
</tr>
<tr>
<td>HOX9 Locus</td>
<td>17</td>
</tr>
<tr>
<td>Chromatin and Epigenetics</td>
<td>17</td>
</tr>
<tr>
<td>DNA Methylation</td>
<td>19</td>
</tr>
<tr>
<td>Histone Tail Modifications</td>
<td>20</td>
</tr>
<tr>
<td>CpG DNA Binding Domains</td>
<td>22</td>
</tr>
<tr>
<td>CpG DNA Binding Domain Proteins used in my research</td>
<td>24</td>
</tr>
<tr>
<td>MLL CXXC Domain</td>
<td>27</td>
</tr>
<tr>
<td>CHAPTER 3: Methods</td>
<td>28</td>
</tr>
<tr>
<td>Cloning of MSCVneo-MLL-AF9-Flag CXXC Domain Swap and Point Mutant Constructs</td>
<td>28</td>
</tr>
<tr>
<td>Cloning of GST-tagged CXXC and MBD Domain Proteins</td>
<td>33</td>
</tr>
<tr>
<td>Bacterial Expression and Column Purification of GST-tagged CXXC and MBD Domain Proteins</td>
<td>34</td>
</tr>
<tr>
<td>DNA binding experiments using Fluorescence Polarization (FP)</td>
<td>35</td>
</tr>
<tr>
<td>Immunoprecipitation and Western Blot Analysis</td>
<td>35</td>
</tr>
<tr>
<td>Retrovirus Production</td>
<td>36</td>
</tr>
<tr>
<td>Retroviral Titering</td>
<td>37</td>
</tr>
<tr>
<td>Primary Mouse Bone Marrow Isolation</td>
<td>37</td>
</tr>
<tr>
<td>C-Kit Positive Selection of Mouse Bone Marrow Progenitor Cells</td>
<td>38</td>
</tr>
<tr>
<td>Retroviral Transduction</td>
<td>39</td>
</tr>
<tr>
<td>Colony Assays</td>
<td>39</td>
</tr>
<tr>
<td>Cytospins</td>
<td>40</td>
</tr>
<tr>
<td>Leukemia Assays</td>
<td>40</td>
</tr>
</tbody>
</table>
Flow Cytometric Analysis of Leukemic Cells 41
RNA Isolation and cDNA Preparation 43
Quantitative RT-PCR 43
MicroRNA Detection Protocol 44
Chromatin Immunoprecipitation Assay 45
Quantitative PCR for Chromatin Immunoprecipitation Assay 47
Nucleofection of Mll-/- MEFs 47
Bisulfite Sequencing to Measure DNA Methylation 48

CHAPTER 4: Results 50
Aim 1: Determine if the DNA binding ability of the MLL CXXC domain is essential to MLL fusion protein function 50
Identification of MLL CXXC Point Mutations critical to DNA Binding 50
Cloning of MLL CXXC Point Mutations into MSC-MLL-AF9-Flag 53
Retrovirus Production and Titers 54
Confirmation of Stable Expression of MLL-AF9-Flag CXXC Mutant Proteins 54
MLL CXXC point mutations that disrupt DNA binding reduce the transformation potential of MLL-AF9 in bone marrow colony assays 56
MLL CXXC point mutations that disrupt DNA binding abolish the ability of MLL-AF9 to promote leukemia in vivo 63
MLL CXXC DNA binding activity is essential for MLL fusion proteins to protect target Hoxa9 CpG DNA from methylation in Mll-/- mouse embryonic fibroblasts 66
MLL CXXC DNA binding activity is essential for MLL fusion proteins to promote overexpression of Hoxa9 and miR196b in bone marrow progenitor cells 67
MLL CXXC DNA binding activity is not critical for MLL fusion proteins to localize to the Hoxa9 Locus 69
Aim 2: Determine whether CpG DNA binding domains from other proteins can substitute for the MLL CXXC domain within the context of MLL-AF9 72
Cloning, Bacterial Expression and Protein Purification of CpG-Binding Domains in GST-Fusion Constructs 72
MLL CXXC domain binds with a higher affinity to unmethylated CpG DNA than other CXXC and MBD domains tested 74
Cloning of alternate CXXC or MBD domains in the place of the MLL CXXC domain into MSCV-MLL-AF9-Flag Constructs 75
Retrovirus Production and Titers and Confirmation of Stable Expression of MLL-AF9-Flag CXXC Domain Swap Proteins 77
DNMT1 CXXC domain can functionally replace the MLL CXXC domain in an MLL fusion protein in an in vitro immortalization assay 79
DNMT1 CXXC domain can functionally replace the MLL CXXC domain in an MLL fusion protein in an in vivo leukemia assay 82
DNMT1 CXXC domain can functionally replace the MLL CXXC domain in an MLL fusion protein to promote overexpression of Hoxa9 in bone marrow progenitor cells.

MLL fusion proteins can localize to the Hoxa9 locus regardless of which CpG binding domain replaces the MLL CXXC domain.

CHAPTER 5: Discussion

APPENDIX A: Cloning Strategy for creating an MLL construct with altered restriction enzyme sites to be used for swapping CXXC domains into MSCV-MLL-AF9

REFERENCES

VITA
LIST OF TABLES

Table                                                                                                                               Page

1. Primers used to generate MSCV-MLL AF9-Flag construct with alternate BamHI restriction site to use for swapping the CXXC domain 30

2. Primers used to amplify CXXC or CpG-Binding Domains for substitution into MLL-AF9                                           31

3. Primers for site-directed mutagenesis of MBD1 MBD                                                                                   32

4. Primers used to amplify CpG-Binding Domains for creation of GST fusion proteins                                                  34

5. Fluorochrome-conjugated Antibodies used for Flow Cytometry Analysis of Mouse Bone Marrow and Peripheral Blood Cells           42

6. Primers used for Chromatin Immunoprecipitation Analysis                                                                          47

7. Primers used for Bisulfite Sequencing                                                                                               49

8. MLL leukemia development in mice reconstituted with MLL-AF9 or MLL(DNMT CXXC)-AF9                                               83
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Protein Domains of MLL and MLL fusion proteins</td>
<td>8</td>
</tr>
<tr>
<td>2. Mammalian HOX gene clusters</td>
<td>16</td>
</tr>
<tr>
<td>3. Structures of the MBD1 MBD domain in complex with methylated CpG DNA and the DNMT1 CXXC domain in complex with unmethylated CpG DNA and alignment of CXXC domains</td>
<td>23</td>
</tr>
<tr>
<td>4. CXXC-Domain containing Proteins</td>
<td>25</td>
</tr>
<tr>
<td>5. Structure of the MLL CXXC domain in complex with DNA and CXXC point mutations that disrupt DNA binding</td>
<td>51</td>
</tr>
<tr>
<td>6. Retroviral Titering and Expression of MLL-AF9 CXXC Point Mutation Proteins</td>
<td>55</td>
</tr>
<tr>
<td>7. MLL-AF9 CXXC Point Mutation Bone Marrow Colony Assay</td>
<td>59</td>
</tr>
<tr>
<td>8. FACS Profiles of MLL-AF9 CXXC Point Mutation Bone Marrow Colony Assay Cells</td>
<td>61</td>
</tr>
<tr>
<td>9. <em>In vivo</em> MLL-AF9 CXXC Point Mutation Leukemia Assay</td>
<td>65</td>
</tr>
<tr>
<td>10. <em>Hoxa9</em> DNA Methylation Levels in MLL(C1188A)-AF9 and MLL(C1188D)-AF9-expressing <em>Mll-/-</em> Mouse Embryonic Fibroblasts</td>
<td>68</td>
</tr>
<tr>
<td>11. <em>Hoxa9</em> expression and fusion protein chromatin localization in MLL(C1188A)-AF9 and MLL(C1188D)-AF9-transduced bone marrow progenitors</td>
<td>70</td>
</tr>
<tr>
<td>12. Alignment of CXXC domains and Expression of CXXC-GST or MBD-GST fusion proteins</td>
<td>73</td>
</tr>
<tr>
<td>13. Relative binding affinities of isolated CXXC domains to unmethylated CpG-containing DNA</td>
<td>76</td>
</tr>
<tr>
<td>14. Schematic and Expression of MLL-AF9 CXXC Domain Swap Proteins</td>
<td>78</td>
</tr>
</tbody>
</table>
15. MLL-AF9 CXXC Domain Swap Bone Marrow Colony Assay

16. In vivo MLL-AF9 CXXC Domain Swap Leukemia Assay

17. FACS profiles of bone marrow cells from leukemic or healthy mice with MLL-AF9 or CXXC domain swapped MLL-AF9

18. Hoxa9 expression and fusion protein binding to the Hoxa9 locus in bone marrow cells expressing MLL-AF9 or MLL-AF9 with substituted CXXC domains

19. Model of the regulation of Hoxa9 locus transcription by the CXXC domain of MLL-AF9

20. CXXC amino acids highlighted by conservation to explain functional specificity of the domain in the context of an MLL-AF9 fusion
ABSTRACT

The MLL gene encodes a multi-domain protein that is involved in the maintenance of Hox gene expression during development and hematopoiesis, and was first identified through its involvement in chromosome translocations that cause leukemia. The CXXC domain of MLL, which is retained in leukemic MLL fusion proteins, is a cysteine rich DNA binding domain, with specificity for binding nonmethylated CpG-containing DNA, and is essential for MLL fusion proteins’ oncogenic properties.

Point mutations within MLL’s CXXC domain, at amino acids identified in structural and binding studies to be important for binding to target DNA, were introduced into the oncogenic MLL-AF9 fusion protein. We performed domain swap experiments in which CpG binding domains from other proteins were swapped in to replace MLL’s CXXC domain in the context of an oncogenic MLL fusion. CXXC domains from DNA methyltransferase 1 (DNMT1), CpG binding protein (CGBP), and methyl-CpG binding domain protein 1 (MBD1), as well as a methyl binding domain (MBD) from MBD1 were swapped into the MLL-AF9 fusion. These particular domains were chosen because their described CpG DNA binding capacity is either similar or different from that described for MLL. In vitro colony assays on isolated murine bone marrow progenitor cells infected with mutant or wild type MLL-AF9 fusion proteins were performed in order to determine whether CXXC point mutations or the CXXC domain swaps would affect the ability of
MLL-AF9 to give an enhanced proliferative capacity to bone marrow progenitor cells. *In vivo* murine studies determined whether the CXXC mutations alter the ability of MLL fusion proteins to cause leukemia. We predicted that the point mutations and different CpG binding domains would change the strength or specificity of MLL binding to DNA, which would affect the ability of MLL-AF9 to cause leukemia.

*In vitro* DNA binding experiments were performed to directly measure and compare the DNA binding affinity of the isolated CXXC domain from MLL to the other domain swap proteins. The results indicate that MLL CXXC domain has the highest binding affinity of the domains tested to the unmethylated DNA target. CGBP CXXC domain has the second highest affinity, at about 7-fold lower than MLL CXXC. DNMT CXXC could bind to this unmethylated DNA, but with a much lower affinity. As expected, neither MBD1 CXXC nor MBD1 MBD was able to bind to the unmethylated DNA.

The results of both *in vitro* replating assays and *in vivo* leukemogenesis experiments have shown significant differences between the ability of various CpG DNA binding domains to function in the context of an MLL-AF9 fusion protein. The point mutations that showed reduced DNA binding also show reduced colony formation and a loss of ability to cause leukemia in mice when present in MLL-AF9. MLL-AF9 containing the DNMT1 CXXC domain shows robust *in vitro* colony forming activity and *in vivo* leukemogenesis activity, similar to the oncogenic MLL-AF9 fusion. However, MLL-AF9 containing either the CXXC domain from CGBP or MBD1, or the MBD domain of MBD1 all show reduced colony forming ability and leukemogenicity *in vivo*. 
This suggests that CXXC domain properties in addition to DNA binding affinity, perhaps including protein recruitment, also contribute to an MLL fusion protein’s leukemogenic properties.
CHAPTER 1
INTRODUCTION

Leukemia involving the *MLL (Mixed Lineage Leukemia)* gene is a particularly aggressive disease that is correlated with a poor patient prognosis.\(^1\) MLL leukemia accounts for greater than eighty percent of infant leukemia, as well as up to 10% of adult leukemia.\(^2\) MLL leukemia develops when DNA damage occurs in a blood, or hematopoietic, stem or progenitor cell, so that the chromosome on which the *MLL* gene resides becomes broken and fused to a different chromosome.\(^3\) This translocation results in the fusion of two genes that encode a chimeric protein that is comprised of part of the MLL protein and part of another partner protein. There have been more than sixty MLL fusion partners identified that result in either an acute myeloid, acute lymphoid, or a mixed lineage type of leukemia for which the *MLL* gene is named.\(^4\) My project utilizes the MLL-AF9 fusion protein, which is one of the more common MLL fusions observed in MLL leukemia patients.\(^5\) Understanding the molecular basis of how and why these MLL fusion proteins cause leukemia to develop is essential to developing an effective targeted therapy for the disease.

Proteins are divided into specific domains, based on their molecular structure and function. This project focuses on a protein domain that is present in MLL and is retained in MLL fusion proteins, called the CXXC domain, which binds to DNA. The MLL protein is known to interact with DNA at the promoters of specific genes and help to
regulate the activity of these genes.\textsuperscript{6} When a leukemic MLL fusion protein is present in a cell, the normal target genes of MLL become misregulated by the MLL fusion protein.\textsuperscript{7} When certain target genes of MLL, including \textit{HOXA9}, become overactive, the normal process of blood cell development is disrupted, and leukemia develops. The presence of the MLL CXXC domain has been shown to be essential for MLL fusion proteins to promote overexpression of target gene \textit{Hoxa9} and to transform bone marrow cells \textit{in vitro}.\textsuperscript{8}

The MLL CXXC domain is known to interact specifically with unmethylated DNA.\textsuperscript{8,9} DNA methylation is an epigenetic modification that appears on certain CpG DNA motifs throughout the genome, and is associated with repression of gene transcription.\textsuperscript{11} Unmethylated CpG DNA, on the other hand, is associated with active gene transcription when it is found in the promoters of certain genes.\textsuperscript{12} Our lab has previously shown that MLL and MLL fusion proteins bind to the \textit{Hoxa9} gene locus, and protect specific CpG motifs in this locus from becoming DNA methylated. The MLL dependent DNA methylation status of these CpGs also correlates with \textit{Hoxa9} gene expression.\textsuperscript{6} We hypothesized that this DNA methylation protection function of MLL is facilitated by the MLL CXXC domain, as it is known to bind specifically to unmethylated DNA.

This project explores the functional role of the MLL CXXC domain in contributing to MLL leukemogenesis. Our collaborators at John Bushweller’s laboratory at the University of Virginia recently solved the structure of the MLL CXXC domain in complex with unmethylated CpG DNA. From this solved structure, they were able to identify the specific amino acids in the MLL CXXC domain that are involved in
contacting target CpG DNA. The Bushweller laboratory designed amino acid point mutations that reduce or abolish the DNA binding affinity of the MLL CXXC domain, while keeping the structure of the domain intact. In the first aim of my dissertation, I introduced the amino acid point mutations that disrupt the DNA binding function of the MLL CXXC domain into an MLL fusion protein. I assessed the contribution of the MLL CXXC DNA binding function to MLL fusion protein function by assaying CXXC point mutant MLL-AF9 proteins in their ability to promote the immortalization of bone marrow cells \textit{in vitro}, to promote leukemia development \textit{in vivo}, to promote overexpression of \textit{Hoxa9}, to protect target \textit{Hoxa9} CpGs from becoming DNA methylated, and to localize to \textit{Hoxa9}. I hypothesized that the DNA binding function of the MLL CXXC domain is critical to MLL fusion protein function.

In the second aim of my dissertation, domain swap experiments were designed in which the MLL CXXC domain was replaced with similar or different CpG DNA binding domains from other proteins in the context of an MLL-AF9 fusion protein. These experiments help determine whether the MLL CXXC domain is unique in its ability to contribute to MLL fusion protein function, or if other protein domains can replace the MLL CXXC domain functionally. We hypothesized that only those CpG binding domains that have the most similar known properties to the MLL CXXC domain, which include a conserved amino acid sequence and known DNA binding activity to unmethylated CpG DNA, would be able to function in an MLL fusion protein.

The results of this project find that the DNA binding function of the MLL CXXC domain is essential to MLL fusion protein function. Because disruption of MLL CXXC DNA binding causes MLL fusions proteins to be unable to promote leukemia
development, this suggests that targeting the MLL CXXC domain with a small molecule inhibitor that inhibits DNA binding may serve as a potential novel therapy for MLL leukemia patients. The CXXC domain swap experiments shed additional light onto the function of the MLL CXXC domain. Two of the alternate CXXC protein domains studied here show binding to unmethylated CpG DNA, similar to the MLL CXXC domain. However, only one of these alternate CXXC domains was able to function in an MLL fusion protein. This suggests that the DNA binding function of the MLL CXXC domain is not the only essential function of this domain. The MLL CXXC domain must contribute other functions to MLL fusion proteins, which may include recruitment of specific co-activator proteins, and this function is not retained in all CXXC domains.
CHAPTER 2

LITERATURE REVIEW

MLL Leukemia

The *MLL* (*Mixed Lineage Leukemia*) gene was first discovered due to its involvement in chromosome translocations that cause leukemia. Upon DNA damage, chromosome 11 at band q23 can undergo a double-stranded DNA break in the breakpoint cluster region within the *MLL* gene. If a simultaneous double-stranded DNA break occurs, often in another chromosome, at one of many possible locations, the two different chromosome regions can fuse. If the break occurs in-frame on the second gene, then the resulting translocation will encode an MLL fusion protein encompassing approximately 1400 amino-terminal amino acids of the MLL protein fused to a C-terminal portion of a partner protein. Over 60 different partner genes have been identified so far that can fuse to *MLL*. In addition to *MLL* translocations, the *MLL* gene can undergo mutations such as partial tandem duplications in which part of the gene is repeated within itself, resulting in an abnormal MLL protein. When *MLL* translocations or rearrangements occur in a bone marrow cell at a specific early stage of maturation, they promote the development of leukemia. *MLL* translocations can cause acute lymphoid leukemia, acute myeloid leukemia or, rarely, a bi-phenotypic mixed lineage, or myeloid-lymphoid leukemia for which the gene is named.
MLL translocations are responsible for 5-10% of all human leukemias.\(^2\) While MLL translocations occur infrequently in adult de novo leukemias, they predominate in infant and in therapy-related leukemias which can develop in patients previously treated with DNA topoisomerase II inhibitors for another cancer. As reviewed by Mohan et al, 69-79% of infants with acute lymphoid leukemia (ALL) have MLL translocations, and this number increases to 90% among infants with congenital ALL. At 35-50%, a smaller but still significant proportion of infants with acute myeloid leukemia (AML) have MLL translocations. Among ALL patients, MLL rearrangements are a marker of poor prognosis, and infants with this disease have a significantly lower survival rate than those with non-MLL rearranged ALL. Among patients with AML, which can be harder to treat than ALL, MLL involvement confers an intermediate risk.\(^{15}\) 5-10% of MLL translocations are found in patients with therapy-related leukemias.\(^2\) MLL rearrangements account for the majority of therapy related leukemias that arise in patients who had been previously treated with DNA topoisomerase II inhibitors,\(^{16}\) but MLL translocations may also be caused by radiotherapy and other cancer treatments that induce DNA damage.\(^2\) The MLL therapy related leukemias, which have a short onset compared to other therapy related leukemias, most often present as AML, but can also appear as ALL, myelodysplastic syndrome, or chronic myelomonocytic leukemia.\(^{17}\) The fusion partner of MLL is often instructive as to whether ALL or AML develops in MLL patients. The MLL-AF9 fusion is one of the more common MLL translocations. It usually causes the development of acute myeloid leukemia, although a small percentage of patients with the MLL-AF9 fusion develop acute lymphoid leukemia.\(^{18}\) Overall, MLL
leukemia patients have a relatively poor prognosis,\(^1,19\) so the development of effective targeted therapies would be beneficial for treating MLL leukemias.

**MLL Protein Domains**

The *MLL* gene encodes a large, multi-domain protein (Figure 1A) that is involved in the maintenance of *HOX* gene expression during development and hematopoiesis.\(^20\) The SET domain resides at the carboxy-terminus of MLL, and possesses histone methyltransferase activity, specific for lysine 4 on histone H3.\(^21\) The MLL transcriptional activation domain (TAD) also appears in the carboxy-terminal half of MLL, where it binds to the co-activator protein CBP,\(^22\) and has been shown to positively regulate gene transcription in reporter assays.\(^23\) The MLL protein undergoes post-translational cleavage into two fragments by the enzyme Taspase1.\(^24\) After cleavage, the 320 kiloDalton (kDa) N-terminal fragment and 180 kDa C-terminal fragment of MLL associate noncovalently via their FYRN and FYRC domains.\(^25\) Several more characterized domains are present in the amino-terminal fragment of MLL. Near the amino terminus of MLL, a specific interacting motif mediates binding to Menin and LEDGF;\(^26\) this interaction appears critical to MLL function and localization.\(^27,28\) Also near the amino terminus of MLL are three AT hook motifs, which bind to A-T rich, bent DNA.\(^23\) The AT hook region of MLL also mediates binding to several proteins including SET, protein phosphatase 2A (PP2A), and GADD34.\(^29,30\) Interestingly, MLL fusion proteins were seen to suppress the pro-apoptotic function of GADD34.\(^30\) Nuclear localization motifs target MLL to the nuclei of cells.\(^31\) The repression domain (RD) of MLL has been shown to downregulate expression of target genes in reporter-gene assays.\(^23\) The repression domain recruits histone deacetylases HDAC1/2 and other co-
Figure 1. Protein Domains of MLL and MLL fusion proteins

(A) The full length MLL protein is cleaved by the enzyme Taspase1 into N-terminal and C-terminal fragments. The C-terminal MLL fragment contains a SET domain which possesses methyltransferase activity specific to histone H3 at lysine 4, and a transcriptional activation domain. The FYRC motif interacts with the FYRN motif of N-terminal MLL to keep the two protein fragments of MLL dimerized. The N-terminal MLL fragment contains a motif that interacts with DNA-binding proteins Menin and LEDGF. Three AT hooks interact with AT-rich DNA. SNL1 and 2 denote the speckle nuclear localization targeting motifs. The CXXC domain interacts with unmethylated CpG DNA. Four PHD domains and an atypical Bromodomain also appear in the N-terminal fragment of MLL, but are not retained in MLL fusion proteins as they appear C-terminal to the breakpoint cluster region at which the MLL gene undergoes chromosome translocations in MLL leukemia patients. These translocations to partner genes result in the production of MLL fusion proteins (B).
repressor proteins including CtBP and the Polycomb group proteins BMI-1 and HPC2.\textsuperscript{32} Recently, the MLL repression domain has also been shown to bind to the polymerase associated factor (PAF) complex, which functions in promoting transcription.\textsuperscript{33, 34} The DNA binding CXXC domain is within the repression domain and was identified by its homology to a region of DNA methyltransferase-1.\textsuperscript{35} The final domains in the amino-terminal fragment of MLL include an atypical bromodomain and four plant homeodomain (PHD) zinc fingers. These domains function in protein-protein interactions and are often identified in chromatin-associated proteins.\textsuperscript{36} The second PHD finger of MLL was shown to function in homodimerization and the third PHD finger mediates binding to trimethylated histone H3K4 and the cyclophilin CYP33.\textsuperscript{37, 38} MLL’s breakpoint cluster region of 8.3kb, just upstream of the region encoding the PHD fingers, is where the MLL gene undergoes translocations to partner genes in leukemia patients.\textsuperscript{39} These translocations result in the loss of MLL’s PHD fingers, bromodomain, activation domain and SET domain and the gain of fusion partner protein domains (Figure 1B).

**MLL Function**

*MLL* is an ortholog of the *Drosophila trithorax* gene, and so belongs to the Trithorax Group of proteins, which positively maintain target gene expression during development. These proteins act in opposition to the Polycomb Group proteins which serve to downregulate the transcription of an overlapping group of target genes, including the *HOX* genes.\textsuperscript{40} *Mll* hypomorphic mutant mice were created by insertion of a lacZ cassette in exon 3b of the *Mll* gene, 3’ of the AT hooks, which caused loss of *Mll* expression.\textsuperscript{41} The null *Mll-/-* and heterozygous *Mll+/-* mice were used to determine the specific functions of *MLL* during development and hematopoiesis. *Mll-/-* mice die during
embryonic development at embryonic day E10.5, indicating that *MLL* is an essential gene for life.\(^4^1\) *Mll* also shows haploinsufficiency, as *Mll+/-* mice were observed to be small at birth and show retarded growth. These heterozygous mice also showed body segmental defects typical of homeotic transformations.\(^4^1\) *In vitro* differentiation assays using *Mll-/-* and *Mll+/-* progenitor cells from the yolk sacs of mouse embryos indicated that *Mll* is required during early hematopoiesis to maintain proper numbers of progenitor cells and to promote differentiation, but is not required for terminal differentiation of hematopoietic cells.\(^4^2\) Another study determined that *Mll* is required for the maintenance of target gene expression during early embryogenesis. *Hoxa7* expression was activated but not sustained in *Mll-/-* embryos by day E9.\(^4^3\) The *Mll-/-* embryos showed phenotypic defects by day E10.5 including defects in the branchial arches and segmental boundaries. It was concluded that *Mll* is required for the development of multiple tissues during embryogenesis including skeletal, hematopoietic, neural, and craniofacial tissues.\(^4^3\)

It is thought that MLL regulates transcription of target genes on the level of chromatin organization.\(^4^4\) The SET domain of MLL possesses histone methyltransferase activity specific for trimethylation of the histone H3 tail at lysine #4.\(^2^1\) This trimethylated H3K4 mark on histone tails acts as a mark of transcriptional activation. It is up to some debate whether H3K4 trimethylation actively promotes transcription or is merely present as a marker for active genes. It has been suggested that for MLL target genes specifically, the trimethylation of histone tails on H3K4 at the promoters of genes by MLL is necessary to recruit the pre-initiation complex or other factors necessary for activation of transcription.\(^4^5\)
The MLL protein contains many functional domains, which play a variety of roles themselves, but MLL does not act alone. Besides the direct protein interactions mediated by the domains of MLL already mentioned, the MLL protein has been found in several large multi-protein complexes. Two of these complexes were isolated from the K562 cell line, a non-MLL rearranged leukemia cell line. The first complex isolated by Nakamura et al found MLL in a complex of at least 29 proteins. These proteins include components of the TFIID transcriptional activation complex, the histone deacetylating NuRD and SIN3 complexes, the SWI/SNF ATP-dependent chromatin remodeling complex, and the Set1 protein complex that has histone H3K4 methylation activity. All of these proteins were found to be present on the promoter of MLL target gene \textit{HOXA9} when MLL was also present, suggesting a functional role of the complex in regulating MLL target genes. The second protein complex isolated by Yokoyama et al was much smaller. The only common elements between the two MLL complexes were the Set1 histone methyltransferase protein complex members WDR5 and RBBP5, which interact with the SET domain of MLL. Additional Set1 proteins Ash2L and HCF were also identified in this second MLL complex. The final member identified in this MLL complex was menin, which binds to an N-terminal motif in MLL. Menin is a DNA-binding protein and a tumor suppressor for multiple endocrine tumors. A third group isolated an MLL complex using an anti-WDR5 antibody in HeLa cells. They found some of the same proteins in the MLL complexes mentioned above, but also identified a MYST family histone acetyltransferase protein called MOF. MOF specifically acetylates histone H4K16, and this activity was shown to be necessary for activation of the MLL target gene \textit{HOXA9}.51
MLL Fusion Proteins

More than 60 MLL partner genes have been identified that fuse to MLL in MLL leukemia patients. Some MLL fusion partners are seen very rarely; several have only been documented in one patient. As reviewed by Krivstov and Armstrong, partner genes fall into categories based on their known or proposed functions. Fusions to cytoplasmic proteins account for about 10% of MLL rearranged leukemias. These cytoplasmic proteins include EPS15, GAS7, EEN, and AF6. All of these proteins have a coiled-coil oligomerization domain that is essential to confer leukemogenic activity when fused to MLL. More rare, at less than 1% of MLL leukemias, are MLL fusions to the septin family of proteins, including SEPT2, SEPT5, SEPT6, SEPT9/MSF, and SEPT11. The septin proteins are also cytoplasmic and interact with cytoskeletal filaments, playing a role in mitosis. In those normally cytoplasmic proteins which have been studied, fusion to MLL causes the MLL-partner fusion proteins to become nuclear. Another group of very rare MLL fusion partners are the nuclear histone acetyltransferase proteins including CBP and p300. The most common MLL partner genes, which account for greater than 80% of MLL rearranged leukemias, encode nuclear proteins. These include AF4, AF9, ENL, ELL, AF10, and AF5q31 and are thought to be involved in activation of the elongation step of transcription. The AF9 and ENL proteins are related and show high sequence homology to each other, while AF4 and AF5q31 belong to a family of proteins that also includes FMR2 and LAF-4, another rare MLL fusion partner.

Different fusion partners, sometimes from the same functional group, are more or less common in pediatric versus adult MLL leukemias, or in the myeloid versus lymphoid phenotype of MLL leukemia. For example, of MLL leukemias, MLL-AF4 accounts for
90% of adult ALLs, but only 34% of pediatric ALLs. MLL-ENL is the second most common fusion in adult AMLs at 14%, and is common (24%) in pediatric ALLs, but is not seen in pediatric AML. Among therapy related leukemias, ELL is the most common MLL fusion partner, followed by AF10, AF9 and AF4 and others. Among all MLL-related ALLs, AF4 accounts for more than half of the MLL fusions, followed by ENL, AF9, AF10, AF6, and EPS15. There are more major fusion partners found in AML, but AF4 is rarely seen. The major AML fusion partners in decreasing order of appearance include AF9, AF10, ELL, AF6, ENL, AF17, AF1Q, SEPT6, and EPS15, which together account for 80% of MLL AMLs.

Interestingly, while the fusion partners are often instructive as to the particular MLL leukemia phenotype, many of the most common fusion proteins interact with each other. Specifically, MLL-AF4 almost always promotes ALL, and MLL-AF9 is more common in AML, but AF9 and AF4 proteins directly bind to each other, and this interaction is maintained in MLL-AF9 and MLL-AF4 fusion proteins. The AF9-AF4 interaction proved essential for survival of MLL-AF4 and MLL-AF9 leukemia cells. It is unclear why AF9 and AF4 fusions would promote different types of leukemia when they colocalize to activate the same set of genes. In addition to the AF9-AF4 interaction, several other proteins have been found to bind in a complex. The super elongation complex, or SEC, identified by Lin et al, found AF5q31/AFF4 (a rare MLL fusion partner), AF4, ENL, ELL, AF9, and pTEFb proteins bound to MLL fusion proteins. Similarly, Yokoyama et al identified a similar group of AF4, AF5q31, ENL, and pTEFb as the AEP, or AF4 family/ENL family/P-TEFb complex. It is hypothesized these elongation complex proteins are constitutively recruited to target genes by MLL fusion...
proteins, which results in the inappropriate by-pass of a regulated step in transcriptional activation, over-expression of the target genes, and leukemia development. Several MLL fusion proteins including ENL, AF9, AF17, and AF10 have also been shown to interact in a complex with the histone methyltransferase enzyme Dot1L. Dot1L methylates histone tail residue H3K79, which is a mark of active transcription. It has been suggested that in the presence of MLL fusion proteins, the H3K79 methylation mark functionally replaces the H4K4 trimethylation mark normally deposited by the SET domain of MLL, which is lost in fusion proteins. The presence of Dot1L has been shown to be essential to the survival of MLL fusion protein-expressing cells.

**MLL Gene Targets**

MLL has been well established as a regulator of \textit{HOX} gene expression, but has more recently also been suggested to be a master regulator of transcription. A genome-wide ChIP-on-chip experiment was performed using MLL antibodies. MLL was found at the promoters of over 5,000 genes, and was present at 90% of RNA Polymerase II occupied genes. 92% of the MLL occupied regions also showed the MLL characteristic H3K4 trimethylation mark. These results suggest that MLL helps to regulate most genes that are transcribed by RNA Polymerase II. This study conflicts with other reports however. Milne et al performed conventional chromatin immunoprecipitation experiments using MLL and RNA Polymerase II antibodies on specific target genes. Promoters of genes such as \textit{Hoxa9}, \textit{Hoxa7} and \textit{Meis1} showed binding of MLL and RNA Polymerase II in cells types in which those genes are expressed. However non-MLL target genes including \textit{Gapdh}, \textit{Hoxa1}, \textit{Pbx1}, \textit{Pbx3}, and \textit{Mll} itself showed RNA Pol II but not MLL binding. The authors concluded from these experiments that MLL cannot be a
master regulator of transcription.\textsuperscript{65} In another study, Wang et al performed tiling arrays for H3K4 trimethylation, using \textit{Mll} wild type and null mouse embryonic fibroblast (MEF) cells. This study found that \textit{Mll} was responsible for the H3K4 trimethylation marks of less than 5\%, or a few hundred, of the genes that have this mark, including a subset of \textit{Hox} genes.\textsuperscript{66} This report also supports the idea that MLL is responsible for the regulation of specific target genes.

\textbf{HOX Genes}

\textit{HOX} genes, which encode DNA-binding homeodomain-containing proteins, are evolutionarily conserved transcription factors that regulate developmental processes such as body segment identity and cell fates in developing embryos.\textsuperscript{67} In humans and mice, the \textit{HOX} genes are present in four clusters: A, B, C, and D, which are located on four different chromosomes (Figure 2). There are thirteen paralogous \textit{HOX} genes, whose chromosomal order is evolutionarily maintained, though no one cluster contains all thirteen genes. \textit{HOX} genes are regulated spatially and temporally in developing embryos such that the more 3' genes on their chromosomes are expressed earlier and toward the anterior side of embryos. \textit{HOX} gene expression patterns spread to the 5' genes later in development and on the more posterior side of the embryo.\textsuperscript{67} \textit{HOX} gene expression patterns also play an essential role in hematopoiesis.\textsuperscript{68} Similar to the developing embryo, in general, more primitive bone marrow cells express the 3' \textit{HOX} genes while more mature bone marrow cells express 5' \textit{HOX} genes.\textsuperscript{70} MLL and Polycomb Group proteins act in opposition to each other on \textit{HOX} genes to either activate or repress their expression, respectively.\textsuperscript{20}
The four mammalian HOX gene clusters are found on four different chromosomes. Genes at the 3’ ends of the clusters are expressed earlier in development and hematopoiesis than the 5’ genes. Figure is adapted from Shah and Sukumar.\cite{69}
**HOXA9 Locus**

*HOXA9* is one of the most well-studied and important MLL target genes. Increased *HOXA9* expression has been identified as an important marker of poor prognosis in AML. In addition, overexpression of *HOXA9* by MLL fusion proteins has been shown to be critical to MLL leukemia development. Several transcripts arise from the *HOXA9* locus. The conical transcript arises from exon CD which splices to exon II. An alternative exon AB was identified 4.5 kilobases upstream of the canonical *HOXA9* promoter. Exon AB is located in a region of high sequence homology across species and encodes for a microRNA, mir196b. Several CpG islands appear throughout the *HOXA9* locus. Our laboratory has shown that MLL binds to a CpG island upstream of exon AB, and when MLL or an MLL fusion protein is present, a specific cluster of CpG residues are protected from DNA methylation which correlates with transcriptional activation of the locus. When MLL is absent, DNA methylation levels in this region increase, and transcription of *Hoxa9* and mir196b becomes repressed.

**Chromatin and Epigenetics**

The human genome consists of more than 3 billion base pairs of DNA, all of which must fit into the nuclei of every one of our cells in our bodies. In order to do this, DNA is highly organized and compacted in complex with proteins into a structure called chromatin. At its most compact, DNA forms the metaphase chromosomes typically seen in karyotypes. Histone proteins primarily facilitate chromatin formation. Two proteins each of H2A and H2B, and two each of H3 and H4 form tetramers that then associate with each other to form octamers of core histone proteins. 146 base pairs of DNA wrap around the core histone particles to form a nucleosome. Nucleosomes are spaced 20-60
base pairs apart, to form the 10 nm beads-on-a-string model of chromatin. The linker histone H1 can then facilitate further compaction of the DNA and nucleosomes by coiling the chromatin into a 30nm diameter fiber. Chromatin in these states is referred to as euchromatin, which includes regions on a chromosome where active genes are located. The 30 nm fiber chromatin can be quickly reverted to the 10 nm beads-on-a-string DNA. Genes in these regions are easily accessible to chromatin remodeling complexes that can slide nucleosomes to give transcription factors and polymerases access to the genes for transcriptional activation. Chromatin can be further compacted into heterochromatin to facilitate transcriptional repression. Heterochromatic DNA regions include centromeres, telomeres, repetitive satellite sequences and the inactivated X chromosomes in females. All DNA and chromatin becomes highly condensed during mitosis and meiosis.

Epigenetics refers to heritable changes to the genome that do not involve alterations of the genetic code. Epigenetic marks include DNA methylation and histone tail modifications such as acetylation, methylation, or ubiquitination. These modifications affect specific protein complex recruitment and DNA packaging into chromatin and results in a specific locus being more or less accessible to transcription factors and activation. The histone code hypothesis proposed by Jenuwein and Allis describes how different epigenetic marks are recognized or “read” by specific protein domains and that it is the combination of epigenetic marks and recruited proteins on genes that leads to specific transcriptional states. The authors propose that because epigenetic marks are heritable, they greatly extend the complexity and amount of information stored in our genomes, beyond that of our genetic DNA sequences.
this epigenetic level that MLL and MLL fusion proteins function to activate their gene targets.

**DNA Methylation**

When a cytosine nucleotide appears 5’ to a guanine nucleotide in a DNA sequence, this dinucleotide is termed CpG, for cytosine-phosphate-guanine. DNA methylation occurs on the number five carbon of the cytosine ring of CpG dinucleotides. Because cytosine and guanine nucleotides base pair with each other, DNA methylation appears symmetrically on both complementary strands of DNA at CpGs. This epigenetic mark appears on certain CpGs throughout the genome, resulting in a DNA methylation pattern that is associated with transcriptional repression. Species-, cell-, and tissue-specific DNA methylation patterns have been documented, and the mark has been found to be involved in the regulation of gene transcription, cell differentiation, embryogenesis, imprinting, and silencing of mobile genetic elements. Changes in DNA methylation patterns have been seen in cancers and other hereditary diseases. DNA methylation is also linked to chromatin modification, as proteins that bind to methylated CpGs recruit HDACs and other corepressors to inactivate chromatin. The CpG dinucleotide is underrepresented in the eukaryotic genome because deamination of a methylated cytosine results in a transition mutation to thymine. CpG islands, often appearing in the promoters of genes, are regions throughout the genome in which CpGs are not underrepresented. The CpGs in CpG islands are often unmethylated, which contributes to keeping the genes they are associated with transcriptionally active. DNA methylation patterns are set by de novo DNA methyltransferase enzymes, DNMT3A and
DNMT3B, and the maintenance methyltransferase DNMT1, but DNA de-methylation is not well understood.

DNA methylation mediated repression of transcription can occur through several mechanisms. As reviewed by Klose and Bird, in the promoters of genes, DNA methylation can inhibit binding of transcription factors to their DNA recognition site. Secondly, methyl-CpG binding proteins can bind via their methyl binding domain (MBD) to methylated DNA and recruit corepressors to silence gene transcription. Third, DNA methyltransferases, which are enzymes that methylate CpG DNA, can recruit HDACs and histone methyltransferases to modify chromatin into a repressed state. Finally, DNA methylation within a gene can inhibit RNA Polymerase II from binding and therefore inhibit transcriptional activation.11

It has recently been shown that a small percentage of CpGs are modified with hydroxymethyl groups. TET1, a CXXC domain containing protein and a rare fusion partner of MLL, can enzymatically modify methylated CpGs, converting the residues to hydroxymethylcytosines.85 Common assays, such as bisulfite sequencing, for measuring DNA methylation do not distinguish between methylated and hydroxymethylated CpGs. However, recent studies have found this modification present in many cell types and correlating with active transcription suggesting this mark has an important functional role that is not yet understood.86

Histone Tail Modifications

The amino-terminal ends of core histone proteins, called histone tails, protrude from compact nucleosomes, making them easily accessible to post-translational modifications.76 Acetylation of lysine residues in histone tails by histone
acetyltransferase, or HAT, enzymes such as CBP, p300, or MOF, is often associated with active gene transcription. Acetylated lysines in histone tails can be recognized by bromodomain-containing proteins such as chromatin remodeling proteins that help to facilitate transcription by relaxing the chromatin structure. Histone deacetylase or HDAC enzymes act to remove acetyl marks from histone tails. Methylation of certain lysine residues in histone tails is performed by histone methyltransferase enzymes that contain SET domains, which are named for three founding genes that have this domain: Su(var)3-9, Enhancer of zeste, and Trithorax. Different methyltransferase proteins often have specificity to methylate specific lysine or arginine residues on histone tails, as they can have different functions. In addition, lysine methylation can appear as mono-, di- or trimethylation, and each of these marks may be placed by different enzymes and have different functional outcomes. For example, the Polycomb Group protein EZH2 trimethylates histone H3K27, which is a mark of transcriptional repression. H3K9 trimethylation, catalyzed by the SUV39H and other proteins, is also a mark of repression. Methylation of H3K4 by MLL, Set1 and others, H3K36 by NSD1, and H3K79 by Dot1L are marks of activation. Chromodomain-containing proteins are one example of proteins that specifically bind to methylated histone residues. For example, HP1 recognizes methylated H3K9 via its chromodomain, and associates with the H3K9 methylase SUV39H1 to propagate the methyl mark and cause spreading of heterochromatin. Histone tail methylation is not a permanent mark; enzymes such as LSD1 and Jumanji-domain containing proteins have been shown to demethylate specific lysine residues. Additional histone tail modifications include phosphorylation of serine residues and ubiquitination of lysines.
CpG DNA Binding Domains

Like histone tail modifications, DNA methylation status can also be read by specific protein domains. The methyl-CpG binding domain, or MBD, specifically binds to methylated CpG DNA (Figure 3A). MBD domains are found in the MBD family of repressor proteins which include MeCP2 and MBD1, MBD2, MBD3 and MBD4. All of the MBD-containing proteins can bind to methylated CpG DNA except MBD3, which has changes in the amino acid sequence of its MBD domain that prohibit binding. The Kaiso protein is also known to bind to methylated DNA, though not through an MBD domain. The zinc finger domains of Kaiso can bind to two methylated CpGs to mediate target repression.

CXXC domains on the other hand, bind specifically to unmethylated CpG containing DNA (Figure 3B). The name of the CXXC domain comes from the eight cysteine residues that have a highly conserved spacing throughout the domain (Figure 3C). The domain folds into a saddle-like structure that loops around in such a manner that the cysteines coordinate two zinc ions, with four cysteine amino acids present around each ion. Structural data suggest that the domain interacts with target DNA such that a single CpG residue specifically forms hydrogen bonds with amino acids in the pocket of the domain, while a few DNA bases surrounding the CpG contribute non-specific electrostatic interactions. The first CXXC domain identified was in the DNA methyltransferase protein DNMT1. The MLL CXXC domain was identified by its homology to DNMT1, and since then CXXC domains have been found in proteins including CpG Binding Protein CGBP, repressor protein MBD1, histone H3K36 demethylase FBXL11, and LCX/TET1, a rare fusion partner of MLL that
Figure 3. Structures of the MBD1 MBD domain in complex with methylated CpG DNA and the DNMT1 CXXC domain in complex with unmethylated CpG DNA and alignment of CXXC domains

(A) Stereo ribbon diagram of the solution structure of the MBD1 MBD domain in complex with methylated CpG DNA, with methyl groups shown in red. Secondary structures are indicated. Structure figure was adapted from Ohki et al.92

(B) Stereo ribbon diagram of the DNMT1 CXXC domain in complex with unmethylated CpG DNA, with the CpG shown in yellow and zinc ions shown in magenta. Figure was adapted from Song et al.96

(C) CXXC domains are aligned to show amino acid sequence conservation: those identical in all, or present in two or more are shaded in dark and light grey, respectively. Sequence of the MBD1 MBD is also shown.

MLL

<table>
<thead>
<tr>
<th>MBD1</th>
<th>MBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAEDWLDCPALSFGKRRLEVFRKSGATCGRSQTYQSPTRIFQKVLRTYLGPAEFLTDFKGI</td>
<td>MAEDWLDCPALSFGKRRLEVFRKSGATCGRSQTYQSPTRIFQKVLRTYLGPAEFLTDFKGI</td>
</tr>
</tbody>
</table>

(A) Stereo ribbon diagram of the solution structure of the MBD1 MBD domain in complex with methylated CpG DNA, with methyl groups shown in red. Secondary structures are indicated. Structure figure was adapted from Ohki et al.92 (B) Stereo ribbon diagram of the DNMT1 CXXC domain in complex with unmethylated CpG DNA, with the CpG shown in yellow and zinc ions shown in magenta. Figure was adapted from Song et al.96 (C) CXXC domains are aligned to show amino acid sequence conservation: those identical in all, or present in two or more are shaded in dark and light grey, respectively. Sequence of the MBD1 MBD is also shown.
hydroxylates methylcytosines. CXXC domains are known to coordinate two zinc ions and most can specifically recognize and bind to one unmethylated CpG motif. By reading the DNA methylation status of a gene, CpG binding domains often play an important role in bridging DNA and histone modifications to cause further repression or activation of a gene locus.

**CpG DNA Binding Domain Proteins used in my research**

CpG binding protein CGBP/CPF1/CXXC-1 is a transcriptional coactivator protein that binds to nonmethylated CpG DNA via its CXXC domain. CXXC-1 is an essential gene, as null mice embryos die shortly after implantation. The CGBP protein localizes to nuclear speckles in the cell and binds to the nuclear matrix. CGBP is thought to be involved in global DNA methylation patterning during embryogenesis, probably through its interaction with DNMT1. Besides its CXXC domain, CGBP also contains plant homeodomains (PHD), acidic and basic domains, and a coiled-coil domain (Figure 4A). CGBP has been shown to interact with H3K4 histone methyltransferase proteins MLL1, MLL2, and hSET1, and may play a role in regulating certain MLL target genes. CGBP also interacts with and helps direct the genomic targeting of H3K4 methyltransferase Setd1A, and thus helps to bridge DNA and histone modifications. DNMT1 is the maintenance DNA methyltransferase enzyme that specifically methylates hemimethylated DNA. DNMT1 is present at the replication fork during S phase and provides the methyl mark to the unmethylated strand of newly replicated, hemimethylated DNA in order to propagate the epigenetic mark. The DNMT1 protein contains a DNA binding CXXC domain, an enzymatic methyltransferase domain, a replication focus-targeting domain, a nuclear localization signal, and two bromo-adjacent
Figure 4. CXXC Domain containing Proteins

A  CGBP

- PHD1
- CXXC
- Acidic
- Basic
- Coiled-coil
- PHD2

1 27 73 164 208 256 317 321 360 430 471 485 591

B  DNMT1

- Autoinhibitory linker
- NLS
- RFD
- CXXC
- BAH1
- BAH2
- Methyltransferase domain

mDNMT1 1 197 205 350 600 699 733 897 911 1107 1140 1551

C  MBD1

- MBD
- CxxC-1
- CxxC-2
- CxxC-3
- TRD

Mbd1a

Mbd1b

Mbd1c

Mbd1d

(A) The human CGBP protein schematic was adapted from Lee et al. 2001. CGBP contains two PHD domains, a coiled-coil domain, acidic and basic stretches and a DNA-binding CXXC domain. (B) Schematic of mouse DNMT1, which has 85% amino acid sequence identity to human DNMT1, was adapted from Song et al 2010. DNMT1 has a replication foci-targeting domain (RFD), a nuclear localization signal (NLS), a DNA-binding CXXC domain, two bromo-adjacent homology (BAH) domains, and a C-terminal methyltransferase domain that has specificity for the unmethylated strand of hemimethylated DNA. (C) Mouse Mbd1 splice variant isoforms were adapted from Jorgensen et al 2004. Several isoforms of human MBD1 also exist, similar to the mouse proteins shown here. MBD1 contains a methyl-binding domain (MBD), a transcriptional repression domain (TRD), and two or three CXXC domains, depending on the splice variant. In MBD1, CXXC-3 binds to unmethylated CpG-DNA, but the CXXC-1 and CXXC-2 domains are unable to bind to DNA.
homology (BAH) domains (Figure 4B). While it was once suggested that the CXXC domain of DNMT1 could recognize the hemimethylated DNA target of the enzyme, it was more recently shown to interact with unmethylated DNA, similar to most CXXC domains\(^9\) (Figure 3B). The DNMT1 CXXC domain is now thought to play an auto-inhibitory role such that when it encounters unmethylated CpG DNA, the CXXC domain recognizes and binds to it, thereby protecting it from becoming methylated by the adjacent enzymatic domain.\(^9\)

MBD1, a member of the methyl-CpG binding protein family, is a co-repressor protein known to bind to methylated DNA via its MBD domain.\(^9\) MBD1 also contains two or three CXXC domains, depending on the splice variant (Figure 4C).\(^9\) From murine Mbd1, the first two CXXC domains were shown to be unable to bind to CpG DNA, regardless of methylation status,\(^1\) but do interact with the Polycomb repressive complex 1 (PRC1) protein Ring1b.\(^1\) Mbd1’s third CXXC domain can bind to nonmethylated DNA\(^1\) and also interacts with PRC1 protein hPc2.\(^1\) MBD1 also interacts with histone H3K9 methyltransferase enzymes SETDB1\(^1\) and Suv39h1, which also recruits HDACs.\(^1\) These interactions between MBD1 and Polycomb group proteins and other chromatin modifying proteins show the importance of coupling DNA methylation to chromatin remodeling for gene repression. It is also known that MBD1 is present at the replication fork during S phase, through its interaction with the p150 subunit of chromatin assembly factor 1 (CAF-1).\(^1\) CAF-1 binds to histone proteins H3 and H4 to reassemble nucleosomes following DNA replication.\(^1\) Both DNMT1 and CAF-1 proteins are present at the replication fork via their interactions with PCNA (proliferating cell nuclear antigen).\(^1,1\) After DNMT1 provides the methyl mark on
newly replicated DNA, MBD1 can bind directly to the methylated DNA, facilitating repression through recruitment of histone methyltransferases and histone deacetylases.

**MLL CXXC Domain**

The CXXC domain of MLL was initially identified by its homology to DNMT1 and was first shown to bind specifically to unmethylated DNA through electrophoretic mobility shift assays. Because it is a DNA binding domain, the MLL CXXC domain was initially thought to function in contributing to the targeting of MLL and MLL fusion proteins to the genes that they regulate. Through deletion and point mutation experiments, the presence of the MLL CXXC domain was shown to be essential to MLL fusion protein function in colony formation assays. In the study by Ayton et al, the conserved cysteine residues of the MLL CXXC domain were mutated to alanines, which would disrupt zinc coordination and appropriate folding of the CXXC domain. From this, we can conclude that an intact CXXC domain is essential to MLL fusion protein function, but nothing can be concluded about the DNA binding function of the MLL CXXC domain as this function would be lost when the domain does not fold properly.

Our laboratory has shown that MLL binds to a region at the 5’ end of the Hoxa9 locus within a CpG island, and that a subset of these CpGs is protected from DNA methylation when MLL or MLL fusion proteins are present. As it is known to bind to unmethylated CpG DNA, we hypothesized that this methylation protection function could be facilitated by the MLL CXXC domain. By protecting the target DNA from methylation, the CXXC domain may help to keep the Hoxa9 locus in an active state.
CHAPTER 3

METHODS

Cloning of MSCVneo-MLL-AF9-Flag CXXC Domain Swap and Point Mutant Constructs

In order to swap CXXC point mutations and alternate CpG binding domains efficiently into MLL-AF9, I generated a modified version of our MSCVneo-MLL-AF9-Flag construct. I introduced a BamHI restriction site at the carboxyl end of the MLL CXXC domain (amino acid 1201; one glycine added) and removed an endogenous BamHI site at amino acid 1251 (no amino acid change) to facilitate domain swapping (see Appendix A). In order to do this, the 5’ MLL fragment (4086 base pairs) from a plasmid encoding MSCV-MLL-ELL fusion had been isolated by a previous graduate student (Donna Santillan) using the restriction enzymes EcoRI, which cuts at the 5’start of MLL, and SalI, which cuts at the 3’ end of the fragment of MLL. This is the 5’ MLL fragment which is included in MLL fusions proteins and encompasses a sequence corresponding with MLL amino acids 1-1403. All Restriction digests were carried out for 1-2 hours for plasmid DNA or overnight for PCR products, at 37° Celsius, and were followed by heat inactivation at 65° Celsius for 15 minutes. If the digest was performed on a vector, the linearized vector was treated with 1ul calf intestine alkaline phosphatase (NEB) for 30-60 minutes at 37° Celsius to prevent internal religation of the vector. All digests were followed by gel purification using Qiaex II beads (Qiagen). I ligated this 4.1 KB DNA fragment of MLL into a pMCS5 holding vector that I had digested with EcoRI.
and SalI. All ligations were carried out using 3:1 molar ratios of insert to plasmid, in
10ul reactions using 1ul T4 DNA ligase (Invitrogen), overnight at 14° Celsius. 1μl of
ligation mix was then transformed into 50 μl of Library Efficient DH5α bacteria cells.
Mini-preps of DNA from transformed bacterial colonies were performed, and correct
DNA ligation was verified by restriction digest.

Next I performed a series of restriction digests, PCRs, and ligations in order to
move the BamHI site in MLL from the corresponding amino acid site 1250 to 1201.
Primer sequences are listed in Table 1. First I digested the pMCS5-5’MLL plasmid with
BamHI, which linearized the DNA with a cut at the site of amino acid 1250. With an
undigested plasmid, I next PCR amplified MLL from aa site 1201 to aa site 1250,
creating with the primers a BamHI site at 1201, and a BglII site at 1250. PCRs for
cloning were carried out with Pfu polymerase in order to ensure high fidelity of
amplification. When DNA is digested with BglII, it creates the same overhangs that
BamHI creates, allowing two pieces of DNA that have been digested with each of the two
enzymes to be ligated onto each other. This creates a hybrid site that destroys the ability
of either enzyme to cut at that site. This allowed me to destroy the 1250 BamHI site
without altering the amino acid sequence produced by the codons involved in this area. I
then digested the PCR products with BamHI and BglII and ligated this fragment into the
BamHI digested pMCS5-5’MLL. This created a plasmid with the DNA corresponding to
aa1201-1250 being present two adjacent times, but with the second aa1250 BamHI site
now a hybrid BglII/BamHI site. In order to delete the duplicated DNA, I next returned to
the unmutated pMCS5-5’MLL plasmid, and PCR amplified the CXXC domain region
corresponding to amino acids 1150 to 1201. The endogenous AatII site was included in
the amplified products, and a BamHI site was added with the reverse primers to appear at
amino acid site 1201. This BamHI sequence inserted with the primers caused an extra
amino acid, glycine, to be introduced into the translated protein. As this glycine appears
C-terminal to the structured part of the CXXC domain, it was not expected to change the
function of the CXXC domain or MLL. Next, these PCR products and the engineered
MLL plasmid containing the duplicated DNA were digested with AatII and BamHI. This
digest removed the DNA sequence corresponding to amino acids 1150-1201 (the CXXC
domain) as well as the original 1201-1250 site, but not the PCR amplified region from
aa1201-1250 added in the last ligation that has the destroyed BamHI site. The AatII and
BamHI digested PCR products and vector were then ligated to each other to create the
new BamHI site at amino acid 1201. This produced a DNA sequence that encodes
5’MLL with AatII and BamHI restriction sites flanking the MLL CXXC domain. This
clone was sequenced and this altered MLL-AF9 fusion protein was also later verified to
transform mouse bone marrow cells with the same capacity as unmodified MLL-AF9.

Table 1. Primers used to generate MSCV-MLL-AF9-Flag construct with alternate
BamHI restriction site to use for swapping the CXXC domain

<table>
<thead>
<tr>
<th>Amplicon</th>
<th>Primers (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MLL (aa1201BamHI-1250BglIII)</strong></td>
<td>Forward: CGGGATCCAAAGCCTACCTGCAGA</td>
</tr>
<tr>
<td></td>
<td>Reverse: GAAGATCTTCTCTTGTGATGGGTAGG</td>
</tr>
</tbody>
</table>
| **MLL (aa1150AatII – 1201BamHI)** | Forward: GGGACGTCGATCGAGGC
|                                  | Reverse: CGGGATCCAGGCATCCATTGTGATTCTGAC            |
Table 2. Primers used to amplify CXXC or CpG-Binding Domains for substitution into MLL-AF9

<table>
<thead>
<tr>
<th>Amplicon</th>
<th>Primers (5’ to 3’)</th>
</tr>
</thead>
</table>
| **MLL CXXC (point mutations)** | Forward: GGGACGTCGATCGAGGCGGTGT  
Reverse: CCGGATCCAGGCGATCGAGGCGGTGT (For R1154A) |
|                           | Forward: GGGACGTCGATCGAGGGCGGTGT (For R1154A)                                     |
|                           | Reverse: CCGGATCCAGGCGATCGAGGCGGTGT                                               |
| **DNMT1 CXXC**            | Forward: GGGACGTCGAAAGCGCGCGCGCGATGTG                                              |
|                           | Reverse: CCGGATCCCTTCCATGCGGCTATGCGGAC                                              |
| **CGBP CXXC**             | Forward: GGGACGTCGAAAGCGCGCGCGCGATGTG                                              |
|                           | Reverse: CCGGATCCCTTCCATGCGGCTATGCGGAC                                              |
| **MBD1 CXXC**             | Forward: GGGACGTCGAAAGCGCGCGCGCGATGTG                                              |
|                           | Reverse: CCGGATCCCTTCCATGCGGCTATGCGGAC                                              |
| **MBD1 MBD**              | Forward: GGGACGTCGAAAGCGCGCGCGCGATGTG                                              |
|                           | Reverse: CCGGATCCCTTCCATGCGGCTATGCGGAC                                              |

The endogenous AatII restriction site at the N-terminus of the MLL CXXC domain (amino acid 1150), and the engineered BamHI site (amino acid 1201) were next used to swap point mutation containing MLL CXXC domains or alternate CXXC or MBD domains into MLL-AF9. MLL CXXC (amino acids 1147-1203) with point mutations R1154A, Q1187A, C1188D, C1188A, or K1185A were provided in pGex vectors by Bushweller Lab. I amplified these point mutant domains with primers that included the endogenous AatII site in the forward primer and an added BamHI site in the reverse primer (Table 2). The R1154A construct required a forward primer that contained this mutation as it appears near the N-terminus of the CXXC domain. The primers listed in Table 2 were also used to amplify the following domains: DNMT1 (NM_001379) (TB MeTase in PcDNA3.1 from Moshe Szyf) CXXC domain (amino acids 649-697), CGBP (NM_001101654) (pBS-CGBP from David Skalnik) CXXC (amino acids 163-215), MBD1 transcript variant 1 (NM_015846) (amplified from genomic DNA) CXXC (amino acids 172-221), and MBD1 transcript variant 1 (NM_015846) (pCS2-5MT-MBD1 from
Adrian Bird) MBD (amino acids 1-75). Following PCR, the PCR products and the pMCS5-5’MLL with altered BamHI site were digested with AatII and BamHI and gel purified. Ligations were performed to introduce the point mutant MLL CXXC or alternate CpG-Binding domains into 5’-MLL. Plasmids were transformed then sequenced.

With regards to the MBD1 MBD domain, its DNA sequence contained a BamHI restriction site within the region we wanted to amplify. In order to be able to use the same cloning scheme for this domain as for the others, I performed a site-directed mutagenesis on the MBD domain prior to ligation into MLL. The MBD domain was amplified with the primers listed in Table 2, then the PCR products were cloned into a pCR2.1 TOPO TA cloning vector following kit protocol, and the product was sequenced. Next, site directed mutagenesis (Stratagene) was performed in order to destroy the BamHI site in the MBD domain, while preserving the amino acid sequence produced from this site. Primers are listed in Table 3. Mutated DNA was transformed into XL-1 Blue bacterial cells and cells were plated on LB agar containing X-gal for blue-white selection according to protocol. Following sequencing, the MBD domain was digested out of the Topo vector with AatII and BamHI, and ligated into 5’MLL as described above.

**Table 3. Primers for site-directed mutagenesis of MBD1 MBD**

<table>
<thead>
<tr>
<th>Amplicon</th>
<th>Primers (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCR2.1 Topo-MBD1 MBD</td>
<td>Forward:GCCCCACAGGAGACAGAATCCGAAGCAAGAAGATTTGAG</td>
</tr>
<tr>
<td></td>
<td>Reverse: CTCAACTTTGCTTCGGATTCTGTCTCCTGTGGGCC</td>
</tr>
</tbody>
</table>
Following the generation of CXXC point mutations or domain swaps into 5’MLL, three way ligations were performed to create MSCVneo-MLL-AF9-Flag constructs. The pMCS5-5’MLL plasmids with altered CXXC domains were digested with EcoRI and Sall and the 5’MLL fragments were gel purified. To obtain the AF9-Flag fragment, MSCVneo-MLL-AF9-Flag was digested with SalI and Bgl2, and the 300 base pair band was gel purified. A purified SalI/EcoRI digested MSCVneo vector was obtained from a previous lab member (Amanda Ryan). Three-way ligations were performed overnight then 1μl was transformed into electrocompetent eDH5α cells by electroporation at 310 volts. All constructs were sequenced.

**Cloning of GST-tagged CXXC and MBD Domain Proteins**

Isolated CXXC or MBD domains were cloned into the pGex-4T1 vector. The designs of the domains were based off of our published work with the Bushweller Lab using 57 amino acids of the MLL CXXC domain (1147-1203) to solve the structure of the domain with DNA.10 Primers for the CXXC domains were designed to include the same number of amino acids before and after the CXXC cysteines as were used from MLL CXXC, and 75 amino acids for the MBD domain, as was previously used to solve the MBD domain structure.92 Primers are listed in Table 4 and amplified DNMT1 (NM_001379) CXXC amino acids 645-700, CGBP (NM_001101654) CXXC amino acids 161-217, MBD1 transcript variant 1 (NM_015846) CXXC-1 amino acids 168-224, and MBD1 (NM_015846) MBD amino acids 1-75. BamHI or BglII (for MBD1 MBD) restriction sites (in forward primers) and EcoRI restriction sites (in reverse primers) were used to clone the PCR products into the pGex4T1 vector. All clones were confirmed by sequencing.
Table 4. Primers used to amplify CpG-Binding Domains for creation of GST fusion proteins

<table>
<thead>
<tr>
<th>Amplicon</th>
<th>Primers (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNMT1 CXXC</td>
<td>Forward: CGGGATCCGAGAAGCCTTTAAGCGC Reverse: CGGAATTTCATCTGCTTCCTCATTGGCC</td>
</tr>
<tr>
<td>CGBP CXXC</td>
<td>Forward: CGGGATCCAGATCAAACGGTGACGCC Reverse: CGGAATTCCCTTGTACGATTCCGGGCG</td>
</tr>
<tr>
<td>MBD1 CXXC</td>
<td>Forward: CGGGATCCAGAATGTMTTAAAGCGTGTCG Reverse: CGGAATTCCCTCGCTCCTTTCCAC</td>
</tr>
<tr>
<td>MBD1 MBD</td>
<td>Forward: GAAGATCTATGGCTGAGGACTGCTGCTG Reverse: CGGAATTCCCTCGGGGCTGGATAGCAC</td>
</tr>
</tbody>
</table>

Bacterial Expression and Column Purification of GST-tagged CXXC and MBD Domain Proteins

GST-CpG Binding Domain constructs were transformed into BL21(DE3)pLysS bacterial cells (Stratagene). Starter cultures (4 flasks of 100mLs each per construct) were grown overnight in LB in the presence of ampicillin and chloramphenicol at 37°C shaking at 225 rpm. Cultures were expanded (into 4.4 liters total per construct) and allowed to grow for 8 hours at 37°C, then were induced with 1mM IPTG overnight at room temperature. Cells were pelleted at 6000 rpm in a JA-10 rotor, then lysed with 50mM Tris, pH 7.2, 400mM Sodium Chloride, 50μM Zinc Chloride, 1mM DTT, 1mg/ml lysozyme, 0.25mg/ml PMSF. Lysates were sonicated at 60% power on ice with stirring for 15 minutes using cycling of 30 seconds on, 30 seconds off. Cell debris was pelleted at 12,000 rpm (21,000 x g) using a JA-18 rotor for 1 hour at 4°C, followed by filtration of the supernatant, or debris was pelleted using an ultracentrifuge at 20,000 rpm (30,000 x g) for 30 minutes at 4°C. Soluble GST-tagged CXXC proteins were isolated using Glutathione agarose beads (Fluka 49739). Proteins were allowed to bind the column overnight, rocking at 4°C. After washing (50mM Tris, pH 7.2, 400mM Sodium Chloride, 50μM Zinc Chloride, 1mM DTT), GST-CXXC or GST-MBD proteins were eluted with
10mM Glutathione (G4251, Sigma, St. Louis, MO) in wash buffer. Protein purity was verified by Coomassie Blue staining.

**DNA binding experiments using Fluorescence Polarization (FP)**

Unmethylated C-terminal fluorescein labeled DNA (5' GGGTCGCGGGAG 3', Integrated DNA Technologies) and the purified GST-tagged CXXC or MBD domains were dialyzed into FP buffer (50mM Tris-HCl, 150mM KCl, 1mM DTT, 10uM ZnCl₂). Fluorescein labeled DNA was added to 96-well black COSTAR (Corning Life Sciences, Lowell, MA, USA) plates. The proteins were separately mixed with fluorescein labeled DNA and serially diluted 1:2 onto the DNA-containing COSTAR plates. The plates were incubated in the dark at room temperature for 1 hour. A PHERAstar microplate reader (BMG Labtech, Durham, NC, USA) was used to measure fluorescence polarization with excitation at 494 nm and emission at 525 nm. The DNA binding experiments were performed three times. To calculate Kᵋ for each protein, anisotropy values (mA) were plotted versus the log of protein concentration (µM), and the resulting plots were fit to a one-site sigmoidal binding curve using Origin 7.0 (MicroCal, Northampton, MA, USA). Final FP experiment included in this study was performed by Aravinda Kuntimaddi.

**Immunoprecipitation and Western Blot Analysis**

Phoenix-Eco cells (Garry Nola, Stanford CA / Orbigen) were harvested three days after transfection with MSCV-MLL-AF9-Flag or domain swap constructs. Cells were lysed with RIPA buffer (50mM Tris, 150mM NaCl, 1% Triton X, 0.1% SDS, 1% Na-deoxycholate) containing Protease Inhibitors (Sigma, St. Louis, MO) at 1:100 for 10 minutes on ice. Lysates were sonicated briefly, then insoluble material was removed by centrifugation at 14,000 rpm for 30 minutes at 4°C. Anti-Flag M2 affinity gel (Sigma)
was washed with PBS-Tween, then beads were resuspended in RIPA buffer with Protease Inhibitors. Cell lysates were incubated overnight with anti-Flag beads at 4°C with rotation to immunoprecipitate MLL fusion proteins. Proteins were eluted by boiling samples in SDS sample buffer at 95°C for 5 minutes, then were resolved by SDS-PAGE on a 5% gel. Proteins were transferred for 2 hours at 70 volts to PVDF (Millipore) membranes using 1X CAPS transfer buffer with 10% methanol (10X CAPS buffer: 100mM CAPS, 0.1% SDS, pH = 11). Membranes were blocked in 5% milk for one hour up to 2 days. MLL fusion proteins were detected by staining with anti-Flag (M2) antibodies (Sigma) in 3% milk at 1:1000 dilution for 2-2.5 hours at 4°C, followed by three 5 minute washes with PBS-0.05% Tween, then secondary anti-mouse-HRP (Amersham, Pittsburgh, PA) at 1:3000 in 3% milk for 2 hours, and three 15-20 minute PBS-Tween washes. Blots were rinsed with water then developed with ECL (Pierce/Thermo Scientific Rockford, IL).

**Retrovirus Production**

Phoenix-Eco cells (Garry Nolan / Orbigen) were grown in DMEM 10% FBS 1% Pen/Strep. Four million cells per dish were plated into five 10cm dishes one day prior to transfection. Using CalPhos Mammalian Transfection Kit (Clontech, Mountain View, CA), the cells were transfected with 25μg of DNA per 10cm dish. Sixteen hours post-transfection, media was removed from the cells and 7ml of fresh media added. Cells were put into a 32°C incubator with 5% CO₂. The following day, media was collected from the cells. An additional 7ml of the media was added to the cells and the dishes were returned to the 32°C incubator. Media was collected again 48 hours post-transfection and combined with the media already collected, then spun down to remove any cell debris.
Retroviruses were concentrated about 6x using Centricon Plus 70 filters (Millipore) by centrifugation in a swing bucket JS 7.5 rotor. Concentrated retrovirus was aliquoted, snap frozen, and stored at -80° Celsius.

**Retroviral Titering**

2.5x10⁴ Rat-1A cells per well were plated in a 6 well dish a day before infection. Serial dilutions of retroviruses (1:2 up to 1:1x10⁶) in 1mL DMEM 10%FBS were added to cells with 16mg/mL polybrene. Cells were incubated for 2 hours at 37°C 5%CO₂, then 1ml DMEM 10%FBS was added to each well. Cells were incubated for 24h at 37°C 5%CO₂. Media was removed then replaced with 2ml DMEM 10%FBS with 0.5mg/mL G418 to select for cells that had taken up the virus. Cells were kept under selection for 10 days, then washed with PBS and stained with 1mL of a solution of 0.1 gram methylene blue in 60 mL Methanol for 5 minutes. Plates were rinsed and dried, then colonies were counted. Ideal titers were in the 1x10⁵ to 1x10⁶ virus particles/mL range, but lower titers were acceptable if larger volumes of virus were used for infection.

**Primary Mouse Bone Marrow Isolation**

One C57Bl/6 mouse, age 8-12 weeks, was sacrificed by CO₂ inhalation. Tibias and femurs were collected in a cell culture dish with RPMI 2%FCS on ice. Working in the hood, excess tissue was removed from bones with sterile gauze. 8 mL RPMI 2% FCS was added to a 10cm plate on ice. Using a 25 gauge 1.5” needle and media in the dish, bone marrow was flushed into the plate until bones appeared white. Cells were spun down at 1500 rpm for 5 min at 4°C, then resuspended in 5mL 1X cold sterile red cell lysis buffer (10X RBC lysis buffer: 83 grams NH₄Cl, 10 grams NaHCO₃, 0.37 grams EDTA in 1 liter H₂O) and incubated on ice for 5 min. Cells were spun down at 1500 rpm
for 5 min at 4°C, then resuspended in 10 ml PBS 2%FCS to create a single cell suspension to count the cells.

**C-Kit Positive Selection of Mouse Bone Marrow Progenitor Cells**

To select for the c-Kit+ progenitor cells in the mouse bone marrow, I followed the Easy Sep Mouse CD117 (c-Kit) Positive Selection Kit (#18757 Stem Cell). Cells were resuspended at 100 x 10^6 cells per ml in PBS 2% FBS, in a Falcon 5ml Polystyrene Round-Bottom Tube. CD117 antibody (with Fc blocker) was added to cells at 70ul/ml cells. Solution was mixed and incubated at room temperature for 15 minutes. Easy Sep PE selection Cocktail was added at 70μl/ml cells. Solution was mixed and incubated at room temperature for 15 minutes. Magnetic nanoparticles were mixed by pipetting 5 times, then added to the cell solution at 50μl/ml cells. Solution was mixed and incubated at room temperature for 10 minutes. Cell suspension was next brought up to 2.5mL with PBS 2% FBS. Cells were mixed by pipetting 2-3 times. Tube without cap was placed in chilled Easy Sep magnet and let sit for 5 minutes in hood. Holding tube in the magnet, it was inverted and supernatant was allowed to pour into waste tube. Tube was removed from magnet and 2.5 mL PBS 2%FBS was added to cells, pipetting 2-3 times to mix. Magnet separation was repeated for a total of four 5 min separations. Cells were resuspended in 1.5mL PBS 2%FBS and counted. Cells were plated overnight in a round bottom 96 well dish at 1 million per 1 mL of RPMI 1640 media with 10% FBS, 1% Pen-Strep, and supplemented with 0.05 mM β–mercaptoethanol (Sigma), 100ng/ml SCF, 10ng/ml IL3, and 10ng/ml IL6. Sterile H2O was added to wells surrounding cells to keep them from drying out.
**Retroviral Transduction**

C-Kit+ cells were infected with MSCV-MLL-AF9-Flag or CXXC mutant-expressing retroviruses or vector control on two consecutive days. 30,000 c-Kit+ cells were centrifuged with 750μl of retrovirus, 7.5μl of 100mM HEPES, 1μl of 4mg/ml polybrene (Sigma) and RPMI 1640 10%FBS to bring volume to 1mL in Beckman GS-6KR centrifuge at 33°C for 4 hours. Double amounts of virus and reagents were used if virus stocks were of low titer. After the first day of spinoculation, cell were cultured in 96 well round bottom dish in RPMI 10% FBS, 1% Pen/Strep, 0.05 mM β-mercaptoethanol (Sigma), 10ng/ml of IL-3, 10 ng/ml of IL-6 and 100ng/ml of SCF. Cells were grown overnight at 37°C in 200μl of media.

**Colony Assays**

Following the second day of spinoculation, cells were plated at a concentration of 1,000 to 10,000 cells per 35mm petri dish, depending on transformation potential, in methylcellulose (Stem Cell M3234). Methylcellulose was supplemented with glutamine, 10ng/ml of IL-3, 10ng/ml of IL-6, 100ng/ml of SCF and 10ng/ml of GM-CSF. For selection of infected cells, G418 was added at concentration of 1.25mg/ml for the first week. Cells were grown in humid chambers consisting of a 15cm dish containing an uncovered 35mm dish holding sterile water. After one week, colonies were counted, then broken up by pipetting, and the total number of cells were determined. Cells were replated at 1-10,000 cells per 35mm dish for four weeks or until cell numbers were too low to replate. Experiments were repeated at least 6 times. Pictures of colonies in methylcellulose were taken through air on a Leica model DMIL microscope (Wetzlar, Germany), through a 4x/0.10 NA lens, with a Canon PowerShot S40 digital camera.
Images were acquired with Canon ZoomBrowser EX, version 8. Using week 4 colony number means, statistical significance was determined by performing two-sample t-tests with a p-value cutoff of 0.05. All constructs were tested for a significant increase in colony number as compared to MSCVneo and a significant decrease in colony number as compared to MSCV-MLL-AF9.

**Cytospins**

50 – 200,000 colony assay cells were spun onto poly L-lysine coated slides (Scientific Device Laboratory, Des Plaines, IL) at 22,000 rpm for 15 minutes using a Cytofuge (StatSpin Technologies, Norwood, MA) and stained with Hema 3 (Fisher Scientific).

**Leukemia Assays**

Bone marrow cells were collected from 5-6 week old B6.SJL (CD45.1+) mice, and c-Kit positive progenitor cells were isolated and infected with retrovirus as described above. C57Bl/6 recipient mice were exposed to 9 Gy radiation from a Gammacell 40 machine. Subsequently, 1.5 – 2 x10^5 of the B6.SJL cells (above) plus 2x10^5 whole bone marrow cells (C57Bl/6 rescue cells), were injected retroorbitally into recipient mice. To monitor for disease progression, mice were examined 3-5 times per week for signs of illness, and peripheral blood samples were obtained once per month. Complete blood counts were measured with a Hemavet 950 (Drew Scientific, Oxford, CT). In addition, expansion of the CD45.1+ blood cell population was monitored by flow cytometry. Sick mice were sacrificed by CO₂ inhalation. Spleen, heart, lung, kidney, and liver tissues were fixed in buffered formalin then were sent to be sectioned and stained with hematoxylin and eosin in the core histology facility. Peripheral blood and bone marrow
smears were stained with Hema 3 stain (Fisher Scientific). Flow cytometry on peripheral blood and bone marrow cells was performed as described below. All experiments on mice in this study were performed with the approval of and in accordance with the Loyola University Medical Center Institutional Animal Care and Use Committee, in accord with Federal guidelines. Pictures of bone marrow and blood smears were taken on an Olympus BH-2 microscope (Tokyo, Japan), under a 100x/1.25 NA oil-immersion lens, with a Sony 3CCD camera, model DXC-76OMD. Pictures of organ sections were taken through air with a 10x/0.25 NA lens on the same microscope. Images were acquired with Adobe Premier software, version 4.2.1. All images were processed using Adobe Photoshop CS3, version 10.0.

**Flow Cytometric Analysis of Leukemic Cells**

Peripheral blood and bone marrow samples (1-5x10^6 cells) from mice were resuspended in 5 ml PBS 2%FBS, then spun down. Supernatants were removed, then cells were resuspended in 5 ml of 1X RBC lysis buffer. Cells were incubated for 5 minutes, then spun down and washed with 5 ml PBS 2%FBS. Bone marrow and blood cells were resuspended in 250 ul each PBS 2%FBS. 50 μL aliquots of cells were moved to 96 well dish. 1 μL anti-CD16/32 was added to each well to block cells for about 10 minutes. One well for each cell type was kept as a no antibody negative control. Several panels of antibodies were designed to assess cell phenotypes present. Five or six different fluorochromes were included in each group, and included FITC (Fluorescein Isothiocynate), PE (Phycoerythrin), PE-Cy5 (Phycoerythrin-Cyanine-5), Pe-Cy7 (Phycoerythrin-Cyanine-7), APC (Allophycocyanin), APC-Cy7 (Allophycocyanin-Cyanine-7). For the APC-Cy7 fluorochrome, I used a streptavidin-conjugated-APC-Cy7
secondary to biotin labeled primary antibodies, while the rest of the fluorochromes were conjugated directly to the primary antibodies (See Table 5). Most groups included the CD45.1 antibody so that the transplanted cells expressing MLL-AF9 could be gated upon for analysis. Group 1 antibodies included myeloid and bone marrow progenitor markers typical of MLL-AF9 myeloid leukemia. Group 2 antibodies added in some B cell markers, while Group 3 included T cell markers, to determine if there was any mixed-lineage phenotypes present among the leukemic cells. Group 4 antibodies included markers of erythroid, megakaryocyte and NK cell types. Antibody groups were pre-mixed then added (approximately 5μl per well) to cells. Antibodies were incubated with cells for 30 minutes on ice in the dark. Positive and negative compensation beads (Becton-Dickinson) were also incubated with antibodies (one well of each fluorochrome type or no antibody for negative beads). Cells were washed one time with 150μl PBS 2%FBS, then resuspended in 100μl PBS 2%FBS. Secondary streptavidin-APC-Cy7 conjugate was then added to cells. Cells were incubated for 30 minutes on ice in the dark. Cells were washed two times, then resuspended in a final volume of ~200μl PBS 2%FBS and moved to FACS tubes. A FACSCanto flow cytometer (Becton-Dickinson) was used to analyze stained cells, and data was analyzed with FlowJo software.

Table 5. Fluorochrome-conjugated Antibodies used for Flow Cytometry Analysis of Mouse Bone Marrow and Peripheral Blood Cells

<table>
<thead>
<tr>
<th>FACS</th>
<th>FITC</th>
<th>PE</th>
<th>PE-Cy5</th>
<th>PE-Cy7</th>
<th>APC</th>
<th>APC-Cy7</th>
<th>Biotin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>CD45.1</td>
<td>CD11b</td>
<td>Gr-1</td>
<td>CD117</td>
<td>SA</td>
<td>Sca-1</td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>CD11b</td>
<td>CD45.1</td>
<td>B220</td>
<td>Gr-1</td>
<td>CD117</td>
<td>SA</td>
<td>CD19</td>
</tr>
<tr>
<td>Group 3 PB</td>
<td>CD3e</td>
<td>CD8a</td>
<td>CD4</td>
<td>CD25</td>
<td>CD117</td>
<td>SA</td>
<td>CD44</td>
</tr>
<tr>
<td>Group 3 BM</td>
<td>CD45.1</td>
<td>CD8a</td>
<td>CD4</td>
<td>CD25</td>
<td>CD3e</td>
<td>SA</td>
<td>CD44</td>
</tr>
<tr>
<td>Group 4</td>
<td>CD45.1</td>
<td>CD41</td>
<td>TER119</td>
<td>CD117</td>
<td>SA</td>
<td>CD49b/DX5</td>
<td></td>
</tr>
<tr>
<td>pos BEADS</td>
<td>CD3e</td>
<td>CD11b</td>
<td>CD4</td>
<td>Gr-1</td>
<td>CD3e</td>
<td>SA</td>
<td>Sca-1</td>
</tr>
</tbody>
</table>
RNA Isolation and cDNA Preparation

RNA was isolated from week 1 methylcellulose colony assay cells (see above) using TRI Reagent (Sigma T9424) according to the manufacturer’s protocol. cDNA was prepared by reverse transcription with the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems #4368814) using random primers. 2μg of RNA was used for each sample. For each reaction 2μl of 10X RT Buffer, 0.8μl of 100mM dNTPs, 2μl of 10X Random Primers, 1μl of MultiScribe Reverse Transcriptase, and nuclease-free H2O up to 20μl was used. cDNA was synthesized using following incubation protocol: 10 minutes at 25°C followed by 120 minutes at 37°C, then 5 seconds at 85°C. cDNA was stored at -20°C until further analysis. No Reverse Transcriptase reactions were also performed to ensure that no genomic DNA was present in RNA samples.

Quantitative RT-PCR

Real-time PCR to analyze gene expression using cDNA described above, was performed using Taqman kits (Applied Biosystems Master Mix kit #4324018, Hoxa9 kit Mm00439364_m1, Hprt kit Mm00446968_m1). Each sample was analyzed in triplicate using 2μL cDNA, 10μL 2X Taqman Master Mix, 1μL 20X Gene Expression Assay (Primers/Probe mix), and nuclease free water to 20μL total. The standard program of one cycle of 50°C for 2 minutes, one cycle of 95°C for 10 minutes, 40 cycles of 95°C for 15 seconds-60°C for 1 minute was run and data was analyzed with ABI Prism 7300. Primer sequences are withheld by Applied Biosystems. Relative Hoxa9 expression was normalized to Hprt expression levels, and calculated using the 2-ΔΔCt method. Experiment was repeated two times.
MicroRNA Detection Protocol

RNA was isolated from week one colony assay cells using TRI Reagent (Sigma) according to the manufacturer’s protocol. Reverse transcription was performed using TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems). For the RT step, a total of 10ng of RNA was used for each control and testing sample. In a microcentrifuge tube 5μL of RNA was mixed with 0.15μL of 100mM dNTPs, 1μL of MultiScribe Reverse Transcriptase (50U/μL), 1.5μL of 10X Reverse Transcription Buffer, 0.19μL of RNase Inhibitor (20U/μL) and 4.16μL of nuclease-free water. To each control sample 3μL of the RNU6B RT primer (Applied Biosystems) was added. To each testing sample 3μL or mir-196b RT primer (Applied Biosystems) was added. The tubes were centrifuged briefly and kept on ice for 5 minutes. The tubes were loaded into the PCR Express cycler (Hybaid) and the following program was run: 1 cycle of 16°C for 30 minutes, 1 cycle of 42°C for 30 minutes, 1 cycle of 85°C for 5 minutes, followed by 4°C hold.

Real-time PCR was performed using ABI 7300 machine using the standard program described above. Each sample was performed in triplicate and contained 10μL of TaqMan Universal PCR Master Mix (Applied Biosystems), 7.67μL of nuclease-free water and 1.33μL of the RT reaction. To each control sample 1μL of RNU6B primer/probe mix was added (Applied Biosystems) and to each testing sample 1μL of mir-196b primer/probe mix was added (Applied Biosystems). Relative mir-196b expression was normalized to RNU6B levels and calculated using the $2^{-\Delta\Delta C_t}$ method. Relja Popovic performed the miR196b Taqman Expression real time PCR experiment included in this dissertation.
Chromatin Immunoprecipitation Assay

The chromatin immunoprecipitation (ChIP) assay was performed using EZ-Magna ChIP G (Upstate/Millipore, Temecula, CA) according to the manufacture’s protocol. For the CXXC point mutation constructs, Phoenix-Eco cells (Orbigen, San Diego, CA) were transfected with MSCV-MLL-AF9-Flag constructs using the CalPhos Mammalian Transfection Kit (Clontech, Mountain View, CA). One dish of transfected Phoenix cells (17-25x10^6 cells) was used for each IP. 70 hours after transfection, cells were washed with PBS then trypsinized and moved to falcon tubes. Cells were crosslinked for 10 minutes at room temperature with 37% formaldehyde (final concentration 1%). The reaction was quenched with 10X glycine solution for 5 minutes at room temperature. Cells were washed with cold PBS twice, then spun down, snap frozen and stored at -80°C until ready to proceed. Cells were thawed on ice, then like cell types were combined and resuspended in 1mL SDS Lysis Buffer containing 1X Proteinase Inhibitor Cocktail II. Cells were allowed to lyse for 15 minutes on ice, vortexing every 5 minutes. After lysis, cells were spun down and resuspended in nuclear lysis buffer (700-850μL depending on cell concentration) with 1X Protease Inhibitor Cocktail II. Lysates were sonicated on ice five times 7 seconds at power 4, resting 30-40 seconds between cycles, using a Branson Sonifier 250. After sonication, lysates were centrifuged at 12,000xg for 10 minutes at 4°C to remove insoluble material. Supernatants were then diluted 10X with Dilution Buffer plus 1X Proteinase Inhibitor Cocktail II and divided up into four tubes, for the number of IPs to be done. 1% of each sample was saved as Input chromatin. Before pulldowns, the sample was pre-cleared to reduce background binding of any proteins to the beads. 10μL of magnetic Protein G beads were added to
each aliquot and samples were placed on wheel at 4°C for 1 hour. Samples were then placed in the magnetic separator to remove the beads, and lysates were moved to new tubes. 20μL fresh magnetic Protein G beads were added to each sample with the appropriate antibody. Chromatin was immunoprecipitated overnight on wheel at 4°C using 5μg anti-Flag (Sigma, St. Louis, MO), or 1μg of anti-Histone H3 trimethyl K9 (Abcam, Cambridge, MA), anti-RNA Polymerase II (Upstate/Millipore) or mouse IgG (Upstate/Millipore). Following incubation, Protein G beads were pelleted with magnetic separator and supernatant was removed. Beads were washed for 5 minutes on ice with each solution: low salt buffer, high salt buffer, LiCl buffer and TE buffer. To elute DNA/Protein complexes and reverse crosslinks to free the DNA, beads were resuspended in 100 μL ChIP Elution Buffer plus 1 μL Proteinase K. Samples were incubated at 62°C for 2 hours with agitation, then 95°C for 10 minutes. After allowing the samples to cool to room temperature, beads were separated with magnet and supernatants were moved to new tubes. DNA was then purified using spin columns, eluted in 50μL and stored at -20°C. Results were analyzed by qPCR as described below. Enrichment was normalized to GAPDH and input chromatin. Primer sequences for the **HOXA9** and **GAPDH** DNA listed in Table 6.

For the CXXC domain swap project, ChIP was carried out as described above with the following changes. C-Kit+ mouse bone marrow cells were infected with MSCV-MLL-AF9-Flag domain swap retroviral constructs as described above. Cells were harvested and fixed with 1% formaldehyde one week after infection and selection. 8-10x10⁶ cells were used for Flag IPs and 2-5x10⁶ cells were used for IgG or p16 IPs. Lysates were sonicated for six cycles of 10 seconds at power 4. Results were analyzed
by qPCR as described below. Enrichment of anti-Flag over IgG binding was normalized to input chromatin. Primer sequences for *Hoxa9* DNA are listed in Table 6.

**Quantitative PCR for Chromatin Immunoprecipitation Assay**

For quantitative PCR for Chromatin Immunoprecipitation experiments, DNA was analyzed using ABI Prism 7300 sequence detector. Each 25 μL reaction contained 4μL ChIP DNA, 12.5μL of 2X iTaq SYBR Green Supermix with Rox (BioRad, Hercules, CA) and 0.2 μM forward and reverse primers. All reactions were performed in triplicate and DNA levels were detected using SYBR Green reagents. All qPCR reactions were performed using the following protocol: one cycle of 50°C for 2 minutes, one cycle of 95°C for 10 minutes, 40 cycles of 95°C for 15 seconds-60°C for 1 minute, one cycle of 95°C for 15 seconds-60°C for one minute-95°C for 15 seconds-60°C for 15 seconds. DNA levels were normalized to input and fold difference was calculated using 2^{-ΔΔCt} method. Primers used for Chromatin Immunoprecipitation assays can be found in Table 6.

**Table 6. Primers used for Chromatin Immunoprecipitation Analysis**

<table>
<thead>
<tr>
<th>Amplicon</th>
<th>Primers (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>HOXA9 AB Region</em></td>
<td>Forward: GAAGCCACATAGTAAGCAAGCAAGTC</td>
</tr>
<tr>
<td></td>
<td>Reverse: GGCTGACTAGGAGATCTGATTAGG</td>
</tr>
<tr>
<td><em>GAPDH</em></td>
<td>Forward: TACTAGCGGTTTTACCAGGCAG</td>
</tr>
<tr>
<td></td>
<td>Reverse: TCGAACAGGAGGACAGAGCGAGCGA</td>
</tr>
<tr>
<td><em>Hoxa9 AB Region</em></td>
<td>Forward: CGGTGATTCTAGTTTCTCTCGTGAG</td>
</tr>
<tr>
<td></td>
<td>Reverse: CACACCGGCGGAAGGAAGAC</td>
</tr>
</tbody>
</table>

**Nucleofection of *Mll* -/- MEFs**

Two million *Mll* -/- MEF (Stanley Korsemeyer) cells were resuspended in MEF2 solution (Amaza) and transfected by electroporation (Amaza) with no more than 10μg DNA in a 5μL volume. MSCVneo, MSCV-MLL(C1188A)-AF9, or MSCV-
MLL(C1188D)-AF9 point mutation constructs were co-transfected with a pSuper plasmid at a 1:5 molar ratio to confer puromycin resistance. Program A-23 was used to nucleofect the cells with the Amaza machine, then cells were immediately plated in pre-warmed (37°C) growth media. On the second day after transfection, puromycin was added to (1.4μg/mL) select for the transfected cells. After one week of selection, genomic DNA was harvested (Puregene DNA kit, Gentra Systems). If possible, cells were cultured out to four weeks post-transfection, and DNA was harvested at week 2 and 4 timepoints, but only week one data was found to have significance and was fully analyzed.

**Bisulfite Sequencing to Measure DNA Methylation**

2μg DNA from transfected Mll -/- MEFs was subjected to sodium bisulfite treatment with the EpiTect Bisulfite Kit (Qiagen). Conversion was run in PCR machine with the following cycling: 5 minutes at 99°C, 25 minutes at 60°C, 5 minutes at 99°C, 85 minutes at 60°C, 5 minutes at 99°C, 175 minutes at 60°C, then 20°C hold, and was followed by DNA cleanup according to the kit. Nested PCR was performed using HotStar Taq(Qiagen) on each sample to amplify the upstream region of Hoxa9 that our lab had previously shown to have MLL dependent CpG-DNA methylation protection. In order to PCR amplify bisulfite converted DNA, we must keep in mind that bisulfite treatment converts any unmethylated cytosine to uracil, so we assume that a bisulfite converted DNA sequence will contain thymines at any position that previously had a cytosine that was not within a CpG. For the cytosines that are part of CpGs, they may or may not have been methylated and bisulfite converted. If a primer crosses a CpG, the primers should allow for either a cytosine or thymine at this position in the forward
primer, or an adenine or guanine in the reverse primer. Primers were designed to amplify
the top strand only of bisulfite converted DNA, and are listed in Table 7. PCR products
were sequenced using the same inside primers that were used to amplify the products.
ImageJ (NIH) software was used to analyze each CpG peak in the sequence tracings.
Area under the curve values were calculated for each CpG cytosine peak, which
represents a methylated CpG residue, and its corresponding thymine peak, which
represents an unmethylated CpG. Error due to background levels of conversion was
determined by measuring non-CpG cytosine conversion. Relative methylation
percentage at each CpG peak was then calculated. PCR products were also cloned into
TOPO vectors (Invitrogen) for sequencing.

**Table 7. Primers used for Bisulfite Sequencing**

<table>
<thead>
<tr>
<th>Amplicon</th>
<th>Primers (5’ to 3’)</th>
</tr>
</thead>
</table>
| Outside *Hoxa9* AB CpGs PCR product: 225 bases | Forward: TTTT(C/T)GTTAGGTAA(C/T)G(C/T)GTTT  
Reverse: AAACAAAAACCTCCCTAAAC(A/G)A |
| Inside *Hoxa9* AB CpGs PCR product: 185 bases   | Forward: GTTTTTTTGTT(C/T)GT(C/T)GGAGG  
Reverse: AAAAAAC(A/G)CAAAAAACAACAAC |
CHAPTER 4

RESULTS

Aim 1: Determine if the DNA binding ability of the MLL CXXC domain is essential to MLL fusion protein function

Identification of MLL CXXC Point Mutations critical to DNA Binding

It was previously shown through deletion studies that the presence of the MLL CXXC domain is essential for MLL fusion proteins to confer increased proliferative capacity upon transduced bone marrow progenitor cells in *in vitro* colony assays. It was also demonstrated that point mutations within the CXXC domain that altered the highly conserved, zinc-coordinating cysteines or the KFGG motif also disrupted colony formation. These point mutations were later shown to cause an isolated MLL CXXC domain to lose coordination of its two zinc ions, indicating misfolding of the domain. As a misfolded CXXC domain would likely lose all function, these point mutation studies did not specifically address whether or not the DNA binding function of the MLL CXXC domain is essential to MLL fusion proteins. In collaboration with our laboratory, John Bushweller’s laboratory at the University of Virginia recently solved the structure of the MLL CXXC domain and mapped its DNA-binding interface to determine which amino acid residues within the domain contact the target DNA (Figure 5A and B). The DNA used for this structure had a palindromic sequence with a central unmethylated CpG. With this information, the Bushweller laboratory introduced point mutations into an isolated MLL CXXC domain (amino acids 1147-1203) and measured the resulting
Figure 5. Structure of the MLL CXXC domain in complex with DNA and CXXC point mutations that disrupt DNA binding

(A) Structure of the MLL CXXC domain is shown in green in complex with DNA with six central base pairs shown in blue. The magenta amino acid loop mediates CXXC base specific contacts with DNA. Zinc ions are shown as black spheres. (B) MLL CXXC in complex with DNA with specific amino acids indicated. Green residues are involved in electrostatic interactions with DNA, the magenta Q1187 forms a hydrogen bond with the guanine base of the CpG, and C1188 in yellow is in close proximity to the DNA. (C) Relative dissociation constants ($K_d$) for binding of isolated wild type or mutant MLL CXXC to unmethylated CpG DNA. The $K_d$ for the C1188D mutant could not be determined due to very weak binding. Figures were provided by the Bushweller Laboratory.
changes in binding affinity by NMR titration to the nonmethylated CpG-containing target DNA molecule for each CXXC point mutant (Figure 5C). They also used NMR to determine that each introduced point mutation did not cause the domain to undergo any structural changes as compared to the unmutated MLL CXXC domain.

The MLL CXXC point mutations that we decided to pursue for further functional studies were R1154A, K1185A, Q1187A, C1188D, and C1188A. The R1154A and K1185A mutations disrupt electrostatic interactions between their amino acid side chains and the sugar phosphate backbone of the target DNA. These mutations decrease CXXC binding affinity by 10.5 and 14.6 fold, respectively. Lysine1185 also contributes a hydrogen bond interaction between the carbonyl group of the lysine amino acid protein backbone with the amine group of the CpG cytosine; however, mutation of this residue to alanine or any other amino acid would not disrupt this interaction as it does not involve the specific amino acid side chain. The Q1187A mutation disrupts a CXXC side chain hydrogen bond interaction with the guanine base of the CpG residue, which decreases the binding affinity by 6.4 fold. Cysteine1188 is not involved in any DNA contacts; however, it is in close proximity to the DNA. We hypothesized that mutation of this residue to the negatively charged amino acid aspartate could introduce a repulsive interaction between the CXXC domain and the DNA. When the C1188D mutation was introduced into the MLL CXXC domain, DNA binding was completely abolished. At the same amino acid, a C1188A mutation has no negative effect on CXXC binding to DNA (Figure 5C). In order to determine if DNA binding is a property of the MLL CXXC domain that is essential to MLL fusion proteins in their ability to promote leukemia, I introduced the above mutations into the CXXC domain of an MLL-AF9 fusion protein.
Cloning of MLL CXXC Point Mutations into MSCV-MLL-AF9-Flag

As MLL-AF9 cDNA is quite long, it is not practical to perform site-directed mutagenesis to introduce mutations, as the entire 4.4 kilobase coding region plus 6.5 kilobase vector construct would need to be sequenced following PCR. Mutations had also already been introduced into isolated MLL CXXC domains in pGex bacterial expression vectors by the Bushweller Laboratory. In order to efficiently introduce the point mutations into MLL-AF9, two restriction enzymes were utilized that could cut at either end of the approximately 150 nucleotide (50 amino acid) CXXC domain. MLL contains an endogenous AatII restriction site at the 5’ start of the CXXC domain which corresponds to MLL amino acid 1149. To cut at the 3’ end of the CXXC domain, I created a BamHI restriction site at the DNA site corresponding to MLL amino acid 1201, and destroyed an endogenous BamHI site at MLL amino acid 1250. Introduction of the BamHI site inserted a glycine residue into the MLL amino acid sequence carboxy-terminal to the CXXC domain. As this glycine appears in the unstructured region of the CXXC domain, it was not expected to alter the function of MLL or MLL-AF9. I later performed side-by-side colony assays to compare the transformation potential of the altered MLL-AF9 with an unmutated MLL-AF9 and determined that there was no difference. I next PCR amplified the point mutation-containing CXXC domains, and digested the products with AatII and BamHI to swap them into the 4.1 kilobase 5’ MLL fragment that is retained in leukemic MLL fusions. After sequencing the regions that had been PCR amplified and digested, I performed three-way ligations to create CXXC point mutation containing-MSCV-MLL-AF9-Flag constructs.
Retrovirus Production and Titers

In order to stably express MSCV-MLL-AF9 constructs in primary murine bone marrow cells, retroviruses containing the constructs were used to infect the cells. This retroviral transduction causes the MSCV-MLL-AF9 DNA to be integrated into the genome of the cells. Ecotropic Phoenix cells, which were engineered with retroviral packaging genes and were derived from the human embryonic kidney (HEK 293) cell line, were employed to produce retrovirus. I used calcium phosphate to transfect MSCVneo (empty vector), MSCV-MLL-AF9-Flag, or MSCV-MLL (CXXC point mutations)-AF9-Flag constructs into Phoenix-Eco cells, and collected and concentrated the retrovirus that was produced. The transfected Phoenix-Eco cells were also saved to perform Western Blots to confirm stable protein expression of the mutant MLL-AF9 constructs. To ensure that the viruses collected were of sufficient concentration to infect primary bone marrow cells, I performed retroviral titering using Rat1A cells. Ideal titers were in the $1 \times 10^5$ to $10 \times 10^6$ virus particles/mL range, but titers as low as $5 \times 10^4$ particles/mL were acceptable if larger volumes of virus supernatant were used for infection (Figure 6A).

Confirmation of Stable Expression of MLL-AF9-Flag CXXC Mutant Proteins

While the Bushweller laboratory had shown that the CXXC point mutations did not disrupt folding or expression of the isolated MLL CXXC domain in vitro, it was also important to show that introduction of the point mutations did not disrupt the stability or expression of an MLL-AF9 fusion protein in cells. Protein lysates were obtained from Phoenix cells which had been transfected with MSCVneo, MSCV-MLL-AF9-Flag, or MSCV-MLL-AF9-Flag with CXXC point mutations. Anti-Flag agarose beads were used
Figure 6. Retroviral Titering and Expression of MLL-AF9 CXXC Point Mutation Proteins

(A) Methylene blue stains of retrovirally infected Rat1A cells after 10 days of G418 selection, with virus dilutions indicated. (B) Detection of MSCV-MLL-AF9-Flag or CXXC point mutation constructs in transfected or untransfected Phoenix cell lysates subjected to immunoprecipitation and western blotting with anti-Flag antibody.
to concentrate the Flag-tagged MLL-AF9 proteins, and an anti-Flag antibody was used to
detect the fusion proteins by Western Blot. The absolute level of expression was not
important, as expression in the bone marrow cells would depend on virus titer used to
infect the cells and the sites of genomic integration of the DNA in individual cells.
Instead it was important to see that the proteins were all present and at the expected size
on the blot, which indicates that there is no degradation of the proteins or any previously
undetected cloning mistakes (Figure 6B).

**MLL CXXC point mutations that disrupt DNA binding reduce the
transformation potential of MLL-AF9 in bone marrow colony assays**

When cultured in semi-solid methylcellulose media containing the appropriate
cytokines, normal bone marrow progenitor cells will form colonies in the first week of
plating, then undergo differentiation and die within a couple of weeks. Bone marrow
progenitor cells that express MLL fusion proteins, however, have enhanced proliferation
and are able to be serially replated in methylcellulose colony assays. In order to
determine if the DNA binding function of the MLL CXXC domain is critical for MLL-
AF9 to confer an enhanced proliferative capacity to bone marrow progenitor cells, I
performed *in vitro* colony assays. I hypothesized that the MLL CXXC DNA binding
function is critical to MLL-AF9 function, so I expected that those point mutations that
conferred reduced DNA binding ability in the isolated MLL CXXC domain binding
assays would cause reduced colony formation ability in the context of MLL-AF9.

C-Kit+ bone marrow progenitor cells were isolated from mice then were infected
by spinoculation with retroviral particles containing MSCVneo, MSCVneo-MLL-AF9-
Flag or MSCVneo-MLL-AF9-Flag with R1154A, K1185A, Q1187A, C1188D, or
C1188A CXXC point mutations. Transduced bone marrow cells were then plated in methylcellulose with cytokines IL-6, IL-3, GM-CSF and SCF. For the first week of plating, the transduced cells were selected for survival with geneticin (G418), as the MSCVneo vector contains a neomycin resistance cassette. After each week, colonies and cells were counted and were replated if colonies were present and cell numbers were sufficiently high. MLL-AF9-expressing bone marrow cells are highly proliferative, and colony formation can be limited by the nutrients and amount of methylcellulose available in individual dishes. Therefore, while the non-transforming constructs were plated at 10,000 cells per dish, after the initial week of selection, MLL-AF9 and any other highly transforming constructs were plated at concentrations of 1,000 to 3,000 cells per dish, then the resulting colony and cell numbers were normalized for comparison to the constructs that had 10,000 cells plated. Colony assays were all performed in duplicate and were repeated at least six times. Results are presented as averages of all colony assays in which at least 10 colonies were present at the end of the first week of selection. Colony assay results often vary from experiment to experiment, due to differences in viral titers used to infect the cells, the specific state of differentiation of the cells that had been infected, and the locations of genomic integration of the constructs which can confer high or low expression.

As seen in Figure 7A, MLL-AF9 expressing bone marrow cells produced 1500 colonies on average during week four of the colony assay, which was a statistically significant increase as compared to MSCVneo cells. Almost 16% of the MLL-AF9 plated cells gave rise to a colony, and the total cell numbers increased on average by 900 to 1000-fold each week. MLL CXXC point mutation C1188A did not perturb DNA
binding activity of the isolated CXXC domain. MSCV-MLL(C1188A)-AF9 expressing bone marrow cells produced many colonies and showed no statistical difference to the unmutated MLL-AF9, as expected. Both constructs produced large, round, densely packed colonies typical of bone marrow progenitor cells (Figure 7B). At the end of the fourth week of replating, cells were taken for morphological analysis. Cytospins of the cells onto slides were performed, followed by staining with Hema-3, a Wright-Giemsa-like stain. Digital pictures of the cells indicated the MLL-AF9 and MLL(C1188A)-AF9 colony assay cells had a blast-like morphology, with large nuclei and little cytoplasm. Cells taken from week 4 of the colony assay were stained with a panel of antibodies directed against a variety of mouse hematopoietic cell surface markers in order to determine the phenotype of the cells. MLL-AF9 and MLL(C1188A)-AF9 cells co-expressed myeloid markers CD11b and Gr-1, and expressed c-Kit at a low level (Figure 8). These cells also expressed B-cell marker CD19, but had very low expression of the erythroid cell marker TER119 and T-cell marker CD3e.

Point mutations that caused reduced binding of the isolated CXXC domain to DNA also significantly reduced MLL-AF9 colony forming ability. The number of colonies produced by MLL-AF9 with point mutants R1154A, K1185A, and Q1187A, which reduced MLL CXXC binding to DNA by 10.5, 14.6, and 6.4 fold, respectively, in the final week of the assay varied, and ranged from zero up to 300. However, compared to unmutated MLL-AF9, all three mutants showed statistically significant decreases in clonogenic activity. The morphologies of these point mutant colonies differed significantly from those of MLL-AF9 and were more typical of differentiated bone marrow cells (Figure 7B). The colonies were most often less dense and smaller than the
Figure 7. MLL-AF9 CXXC Point Mutation Bone Marrow Colony Assay

(A) Average numbers of colonies for weeks two through four after re-plating in methylcellulose are shown for bone marrow progenitor cells expressing MLL-AF9 or MLL-AF9 with indicated CXXC domain point mutations, with error bars showing standard error. Relative Kd values for binding of the wild type and mutated MLL CXXC domains to DNA are also shown. (B) Digital photographs showing colony (above) and cell (below) morphologies of transduced bone marrow cells at the end of week 4 of the colony assay.
MLL-AF9 colonies (Figure 7B), although they did show some variability. As compared to MSCVneo colony numbers, MLL-AF9 with point mutations R1154A and K1185A, but not Q1187A, showed a small but statistically significant increase in average colony number. The proliferative capacity of these point mutant-transduced cells was also greatly diminished, with total cell numbers increasing on average ten-fold less than with MLL-AF9. Surviving week 3 or 4 colony assay cells expressing MLL-AF9 with point mutations R1154A, K1185A, or Q1187A showed a different phenotype than the MLL-AF9 cells, as seen by flow cytometry analysis. While CD11b and Gr-1 remained high, these cells showed increased expression of TER119 and decreased expression of CD117 as compared to the MLL-AF9 cells (Figure 8).

Point mutation C1188D, which completely abolished detectable DNA binding in the isolated MLL CXXC domain, showed the most severe phenotype in the colony assays. The MLL(C1188D)-AF9 transduced bone marrow cells lacked colony forming ability and showed no statistically significant difference in week 4 colony numbers as compared to the MSCVneo negative control (Figure 7A). The C1188D mutant cells proliferated on average only 1.7 fold, and colonies rarely survived into the fourth week of the colony assay. The surviving colonies were always very diffuse with small numbers of differentiated cells (Figure 7B). Immunostaining of week 3 colony assay MLL(C1188D)-AF9-expressing cells revealed lower levels of Gr-1 and CD11b as compared to the MLL-AF9 cells, but very high levels of c-Kit, which is typical of differentiated mast cells (Figure 8). Overall, the colony assay results show a high correlation with the DNA binding affinities of the isolated point mutation-containing MLL CXXC domains (Figure 7A). From these colony assays, we concluded that the
Figure 8. FACS Profiles of MLL-AF9 CXXC Point Mutation Bone Marrow Colony Assay Cells
Bone marrow cells taken from week 3 or 4 of the colony assay, expressing the indicated MLL-AF9 CXXC point mutation constructs, were analyzed by flow cytometry using fluorochrome-conjugated antibodies directed against CD11b, CD3e, TER119, Gr-1, CD117, and CD19.
DNA binding function of the MLL CXXC domain is critical for MLL-AF9 to confer enhanced proliferative capacity upon murine bone marrow progenitor cells \textit{in vitro}.

**MLL CXXC point mutations that disrupt DNA binding abolish the ability of MLL-AF9 to promote leukemia \textit{in vivo}\**

In order to determine if the DNA binding function of the MLL CXXC domain is critical for an MLL-AF9 fusion protein to cause leukemia \textit{in vivo}, mouse leukemia assays were performed. As previously described,\textsuperscript{119} c-Kit positive murine bone marrow progenitor cells were isolated from CD45.1+ B6.SJL donor mice and were retrovirally transduced with MSCVneo, MSCV-MLL-AF9-Flag, MSCV-MLL(C1188A)-AF9-Flag, or MSCV-MLL(C1188D)-AF9-Flag. Along with whole bone marrow cells for radioprotection, the transduced cells were transplanted into lethally irradiated C57Bl/6 recipient mice. As the bone marrow cells from the recipient mice do not express CD45.1, this cell surface marker could be used to track those transplanted bone marrow cells that had received the retrovirus transductions. The mice were observed several times per week, and blood samples were taken periodically (every 2-6 weeks). Complete blood counts were analyzed using a Hemavet (Drew Scientific, Oxford, CT), and flow cytometry using an anti-CD45.1 antibody was used to monitor expansion of the donor bone marrow cells in the peripheral blood of the mice. Mice were sacrificed when they started to become moribund. Sick mice were often underweight, with ruffled fur and slow mobility, and had elevated white blood cell counts.

As seen in Figure 9A, similar to the MSCV-MLL-AF9 mice, all of the MSCV-MLL(C1188A)-AF9 mice developed leukemia. In this experiment, the disease in the MLL(C1188A)-AF9 mutant mice (n = 8) progressed at an even faster rate than the MLL-
AF9 mice (n = 3), with average latencies of 39.5 days and 94 days, respectively. However, similar to the MSCVneo mice (n = 5), none of the MSCV-MLL(C1188D)-AF9 mice (n = 9) developed leukemia during the course of the experiment, which extended for 200 days. Peripheral blood and bone marrow smears indicated an increased percentage of blast cells in the MSCV-MLL(C1188A)-AF9 leukemic mice (Figure 9B and C).

Median complete blood cell counts at the time of sacrifice was 23.9 K per μL for the MSCV-MLL(C1188A)-AF9 mice and 18 K per μL for the MLL-AF9 mice. The normal whole blood count for healthy mice ranges from 1.8 - 10.7 K cells per μL. The sick mice often had enlarged spleens and livers, pale femurs, and infiltrates in multiple organs. The weight of the spleens from the MSCV-MLL(C1188A)-AF9 mice averaged 8 times that of the healthy MSCV-MLL(C1188D)-AF9 mice at time of sacrifice. Organs including spleen, liver, lungs, heart, kidneys and thymus from all sick and some healthy mice for comparison, were fixed in formalin and processed by the histology core facility at Loyola University Medical Center for hematoxylin and eosin (H&E) staining. The livers, lungs and spleens of the leukemic mice typically showed infiltration of leukemic cells (Figure 9C). Peripheral blood and bone marrow cells from the leukemic mice were stained with a panel of antibodies directed against a variety of mouse hematopoietic cell surface markers in order to determine the phenotype of the leukemia cells. An anti-CD45.1 antibody was included in the antibody mix so that the cells expressing MLL-AF9 or MLL(C1188A)-AF9 could be gated upon during data analysis. The leukemic cells expressed high levels of myeloid markers CD11b and Gr-1, moderate levels of the progenitor cell marker c-Kit, and were negative for T-cell marker CD3e and B-cell marker B220 (data not shown).
Figure 9. *In vivo* MLL-AF9 CXXC Point Mutation Leukemia Assay

(A) Survival curve of mice transplanted with bone marrow progenitor cells infected with MSCVneo, MSCVneo-MLL-AF9, MSCVneo-MLL(C1188A)-AF9, or MSCVneo-MLL(C1188D)-AF9(C1188D). Numbers of mice per group are indicated.

(B) Peripheral blood from mice at time of sacrifice (MLL(C1188A)-AF9) or at two months after bone marrow transplants (MSCVneo and MLL(C1188D)-AF9).

(C) Hematoxylin- and eosin-stained sections of liver, lung and spleen or bone marrow smears from an MLL-AF9 (C1188A) mouse (at time of sacrifice) or a healthy MSCVneo mouse two months after bone marrow transplant.
MLL CXXC DNA binding activity is essential for MLL fusion proteins to protect target Hoxa9 CpG DNA from methylation in Mll-/- mouse embryonic fibroblasts

Our laboratory previously discovered that the DNA methylation status of a cluster of CpGs in a CpG island upstream of the AB exon of Hoxa9 is MLL dependent. Using Mll null mouse embryonic fibroblasts (MEFs), we showed that this cluster of eight CpG residues has high levels of DNA methylation when Mll is absent, but when MLL is added back to the cells, DNA methylation levels decrease. When the MLL fusion protein MLL-AF4 was added back to the Mll-/- MEFs, DNA methylation of the first five of the eight CpGs in the cluster was reduced. We hypothesized that this DNA methylation protection function of MLL is facilitated by the MLL CXXC domain as it is known to bind to unmethylated CpGs. To test this hypothesis, I transfected the MSCVneo, MSCV-MLL(C1188A)-AF9-Flag, or MSCV-MLL(C1188D)-AF9-Flag constructs into the Mll-/- MEFs. I expected that the MSCV-MLL(C1188A)-AF9 construct would confer reduced levels of DNA methylation on the cluster of CpGs in the upstream region of Hoxa9, while in MSCVneo and MSCV-MLL(C1188D)-AF9 transfected cells, DNA methylation levels would remain high at this cluster. As the Mll-/- MEFs are resistant to transfection, the Amaxa nucleofection technique was used to transfect the MSCV-MLL-AF9 point mutant constructs along with a pSuper vector. The Mll-/- MEFs were engineered with a neomycin resistance cassette, so the pSuper vector was included so that puromycin could be used to select for the cells into which the plasmids were successfully introduced.

After one week of antibiotic selection, genomic DNA was isolated from the cells and was treated with sodium bisulfite, which converts any unmethylated cytosines to thymines. The upstream region of Hoxa9 was then amplified by PCR using primers that
had been designed to amplify sodium bisulfite converted DNA. Two successive PCR reactions using nested primers were needed to generate sufficient PCR products for analysis, as bisulfite treatment can be harsh on DNA integrity. PCR products were sequenced directly in order to get an accurate average percentage of DNA methylation from the whole population of cells from which the genomic DNA was harvested. Relative levels of DNA methylation were calculated for each CpG in the upstream *Hoxa9* cluster. As seen in Figure 10, MEFs transfected with MSCVneo or MLL(C1188D)-AF9, which does not have CXXC DNA binding activity, show higher levels of DNA methylation at all CpGs previously shown to be protected from methylation by MLL. This was measured as compared to the cells transfected with MLL(C1188A)-AF9, which has normal CXXC DNA binding activity. The first five of the eight CpGs show increased protection from methylation by MLL(C1188A)-AF9, as was observed in our earlier studies using the MLL-AF4 fusion protein. This experiment was repeated three times, with one representative experiment shown. These results confirmed our hypothesis that the DNA binding activity of the MLL CXXC domain is critical to the MLL function of protecting *Hoxa9* from DNA methylation.

**MLL CXXC DNA binding activity is essential for MLL fusion proteins to promote overexpression of *Hoxa9* and miR196b in bone marrow progenitor cells**

The presence of unmethylated CpG islands in the promoters of genes is often correlated with active gene expression. As MLL CXXC DNA binding is necessary to keep CpGs in the MLL target gene *Hoxa9* in the unmethylated state, we hypothesized that it is also necessary to promote expression of *Hoxa9*. RNA was isolated from week one murine bone marrow colony assay cells expressing MLL(C1188A)-AF9 or
Figure 10. *Hoxa9* DNA Methylation Levels in MLL(C1188A)-AF9 and MLL(C1188D)-AF9-expressing *Mll*-/− Mouse Embryonic Fibroblasts

Relative methylation levels of CpGs in the upstream *Hoxa9* locus in *Mll* null MEFs transfected with either MSCVneo-MLL-AF9(C1188A) or MSCVneo-MLL-AF9(C1188D) and pSuper, after one week of puromycin selection. Bisulfite treatment, PCR and sequencing on genomic DNA samples were performed three times, and the results of one representative experiment are shown.
MLL(C1188D)-AF9 (as described above). RNA was converted to cDNA, then was assayed for *Hoxa9* and miR196b expression using Taqman probes in a real-time RT-PCR assay (performed by Relja Popovic). MiR196b is a microRNA that arises from the upstream AB exon of *Hoxa9*, and like *Hoxa9*, is overexpressed in MLL leukemias. As seen in Figure 11A and B, bone marrow cells expressing MLL(C1188A)-AF9 show high levels of *Hoxa9* and miR196b, as would be expected for an MLL fusion protein. However, the CXXC DNA binding-deficient MLL(C1188D)-AF9, similar to MSCVneo, is not able to promote overexpression of these *Hoxa9* transcripts. This suggests that the CXXC binding activity is critical to the function of an MLL fusion protein to overexpress *Hoxa9*.

**MLL CXXC DNA binding activity is not critical for MLL fusion proteins to localize to the *Hoxa9* Locus**

We have established the importance of the DNA binding function of the MLL CXXC domain to MLL fusions in promoting leukemia, protecting *Hoxa9* DNA from becoming methylated, and in promoting expression of *Hoxa9*. It remains to be addressed if the DNA binding function of the MLL CXXC domain is critical for MLL fusion proteins in their localization to target genes such as *Hoxa9*. While the CXXC domain is one of only two known DNA binding domains within MLL, the CXXC domain centers its binding on only one CpG motif at a time, indicating a lack of sequence specificity beyond the unmethylated CpG. However, as the MLL(C1188D)-AF9 protein was unable to activate *Hoxa9*, I expected that this could be due to loss of localization of MLL(C1188D)-AF9 to the *Hoxa9* locus. I performed chromatin immunoprecipitation (ChIP) assays in transfected Phoenix-Eco (human) cells to test for binding of
Expression levels of Hoxa9 (A) and mir196b (B) in bone marrow progenitor cells transduced with MSCVneo vector, MLL-AF9(C1188A), or MLL-AF9(C1188D). Cells were harvested after one week culture in methylcellulose and expression levels of Hoxa9 and mir196b were quantified with real-time RT-PCR. Shown are average relative expression levels, with error bars indicating standard deviation. Relja Popovic provided graphs shown in A and B. (C) ChIP assay performed on Phoenix cells transfected with FLAG tagged MSCV-MLL-AF9(C1188A) or MSCV-MLL-AF9(C1188D). Chromatin was immunoprecipitated with the indicated antibodies and real time PCR was performed with primers that localize near mir-196b in the upstream region of the HOXA9 locus. Samples were run in triplicate and were normalized to GAPDH and input chromatin, with error bars showing range.
MLL-AF9(C1188D) and MLL-AF9(C1188A) to the upstream AB exon of HOXA9. Immunoprecipitations were performed with anti-Flag antibody for the mutant MLL-AF9 proteins, IgG as a negative control, anti-RNA Polymerase II as a mark of active transcription, and anti-H3K9-trimethylated as a mark of repressed transcription. Real-time PCR followed the pulldowns to determine the enrichment of the proteins or chromatin modifications at the upstream AB exon of HOXA9. Surprisingly, my results showed no significant difference in the binding of the MLL(C1188A)-AF9 and MLL(C1188D)-AF9 proteins to HOXA9 (Figure 11C). RNA Polymerase II showed a small but insignificant decrease in the MLL(C1188D)-AF9 cells; the presence of Polymerase II could indicate a paused state of transcription in these cells. The repressive trimethyl-H3K9 mark, however, did show a significant increase in the MLL(C1188D)-AF9 cells compared to the MLL(C1188A)-AF9 cells. Two independent ChIP experiments were performed, with similar results seen.

The results indicate that while the CXXC binding deficient MLL(C1188D)-AF9 protein can still localize to HOXA9, the target gene remains in an inactive, repressed state. We concluded that the DNA binding function of the MLL CXXC domain is not absolutely critical for localization of MLL fusion proteins to target genes. Instead, we expect DNA interactions through the MLL AT hooks, and perhaps interactions with proteins such as Menin and LEDGF which also bind to DNA, to be important for MLL target localization.
Aim 2: Determine whether CpG DNA binding domains from other proteins can substitute for the MLL CXXC domain within the context of MLL-AF9

Cloning, Bacterial Expression and Protein Purification of CpG-Binding Domains in GST-Fusion Constructs

Several individual isolated CXXC and MBD domains have been previously tested for DNA binding ability and specificity. However, the affinities of these CpG-binding domains to the same target DNA molecule have never been measured in a side by side experiment; doing so would allow the DNA binding affinities of the domains to be directly compared to each other. We chose to study the MLL CXXC, CGBP CXXC, DNMT1 CXXC, MBD1 CXXC and MBD1 MBD domains (Figure 12A). Previous work by us and others has shown that the MLL CXXC domain preferentially binds to unmethylated CpG DNA.8-10,117 The CGBP CXXC domain, from the co-activator CpG-Binding Domain Protein (CGBP/Cfp1/CXXC1), is also known to bind to unmethylated CpG DNA,121 so was expected to bind with similar affinity to the target DNA as MLL CXXC. The DNMT1 DNA maintenance methyltransferase protein provides enzymatic DNA methylation activity on hemi-methylated DNA substrates.109 Its CXXC domain was recently shown to interact with unmethylated DNA in an auto-inhibitory manner in order to protect any unmethylated DNA that is encountered by DNMT1 from its enzymatic activity.96 Therefore, the DNMT1 CXXC domain was also expected to be able to bind to the nonmethylated DNA probe. From the repressor protein Methyl-Binding Domain Protein-1 (MBD1), the MBD domain preferentially binds to methylated DNA,93 acting opposite to most CXXC domains. The MBD1 protein contains either two or three CXXC domains depending on the splice variant (Figure 4C). We chose to study the MBD1 CXXC domain with the least amino acid similarity to MLL CXXC (Figure
Figure 12. Alignment of CXXC domains and Expression of CXXC-GST or MBD-GST fusion proteins

**A**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLL</td>
<td>LKGRSRCCQFPGQVPGEDGVTCLNLHFKFGGRNIKKQCKLRKQCHLQWMPK</td>
</tr>
<tr>
<td>CGBP</td>
<td>QIKRSEMCGEAERTEGCHCDFRDFMKFPGPNKIPQICLRQCOLRARATYK</td>
</tr>
<tr>
<td>DNMT1</td>
<td>ENAKFRBCVTECCPPSCGFKCAKXDMVQFKGSGSRKFAOERGRFINAMKEADD</td>
</tr>
<tr>
<td>MBD1</td>
<td>QRMFKRVCQCEAACTEDGACSTCLLQLPVDVAFSGLFKECEERCLRIVERSRG</td>
</tr>
<tr>
<td>Consensus</td>
<td>XXXXXXX</td>
</tr>
</tbody>
</table>

**B**

(A) CXXC domains are aligned to show amino acid sequence conservation: those identical in all, or present in two or more are shaded in dark and light grey, respectively. Sequence of the MBD1 MBD is also shown. (B) Digital photographs of Coomassie Blue stained SDS-PAGE gels of CXXC-GST fusion proteins in fractions eluted from GST beads following purification.
12A). It was previously shown that the Mbd1 mouse isoform of this CXXC domain is unable to bind DNA regardless of methylation status.\textsuperscript{111} We expected that neither the MBD nor the CXXC domain from MBD1 would be able to bind to the unmethylated CpG DNA probe with a sequence derived from an MLL binding site in \textit{Hoxa9}.\textsuperscript{6}

The MLL CXXC domain (amino acids 1147-1203) in pGEX-4T2 was provided by the Bushweller laboratory and is the same construct that was used to solve the structure of the domain. I cloned the isolated DNMT1 CXXC, CGBP CXXC, MBD1 CXXC, and MBD1 MBD domains into pGEX-4T1. I included 57 amino acids from the CXXC domains, as was used for MLL CXXC, and 75 amino acids for the MBD1 MBD domain as was used to solve the structure of this domain\textsuperscript{92} (Figure 12A). pGEX vectors produce GST-tagged fusion proteins and allow for inducible expression in bacteria for large-scale protein purification. After cloning, the constructs were sequenced and expressed in bacteria on a small scale to verify GST-tagged protein expression. The proteins were then expressed in BL21 bacteria cells on a large scale, and were purified using glutathione agarose affinity chromatography. Purified proteins were analyzed by SDS-PAGE (Figure 12B).

\textbf{MLL CXXC domain binds with a higher affinity to unmethylated CpG DNA than other CXXC and MBD domains tested}

Binding affinity was measured by fluorescence polarization to a fluorescein-labeled DNA probe with a sequence derived from an MLL binding site in \textit{Hoxa9}.\textsuperscript{6} This DNA probe has two central nonmethylated CpGs. After I expressed the proteins and shipped them to Virginia, Aravinda Kuntimaddi (University of Virginia) performed the fluorescence polarization experiment presented here. The experiment was repeated three
times, with one representative experiment shown. Increasing concentrations of the proteins were added to a constant amount of DNA until saturation of binding was approached. This allowed us to generate curves (Figure 13A) that we used to calculate absolute and relative binding affinities (Figure 13B). The MLL CXXC domain has the highest DNA binding affinity of the domains tested, at 0.64 μM, to the unmethylated target DNA, and its relative binding affinity was set to 1. CGBP CXXC domain has the second highest DNA binding affinity at 4.44 μM, or about 7-fold lower than the MLL CXXC. The DNMT CXXC domain came in third with a binding affinity of 21.9 μM, or 34-fold lower binding than MLL CXXC. As expected, neither MBD1 CXXC nor MBD1 MBD is able to bind to the unmethylated DNA. In conclusion, while most of the CXXC domains are able to bind to the unmethylated CpG-containing DNA, there are significant differences in the DNA binding affinities measured.

**Cloning of alternate CXXC or MBD domains in the place of the MLL CXXC domain into MSCV-MLL-AF9-Flag Constructs**

Domain swap experiments are those in which a functional protein domain from one protein is removed and replaced with a similar or different type of domain from a different protein. Functional experiments can be performed on the domain swapped proteins to give insight into function and level of specificity and uniqueness of a particular protein domain to its protein. Determination of common features among similar domains from alternative proteins can also be assessed in order to more fully understand a protein domain and perhaps its evolutionary history.

To swap in alternate CpG-Binding domains in place of the MLL CXXC domain in MLL-AF9, the pMCS5-5’MLL construct with the BamHI site moved from amino acid
Figure 13. Relative binding affinities of isolated CXXC domains to unmethylated CpG-containing DNA

(A) Representative binding curves for fluorescein labeled DNA titrated with increasing concentrations of GST-tagged CXXC domains and measured by fluorescence polarization. (B) Absolute and relative $K_d$ values for each of the GST-tagged CXXC domains. Absolute $K_d$ values were determined from the fluorescence polarization titration curves. Relative $K_d$ values were determined by comparison to GST-MLL CXXC (set to 1). Figures were provided by Aravinda Kuntimaddi, University of Virginia.
1250 to 1201 was employed, as was described above. The AatII and BamHI restriction sites were again used to swap in the CXXC or MBD domains into 5’ MLL. The CGBP CXXC (CGBP amino acids 163-215), DNMT1 CXXC (DNMT1 amino acids 649-697), MBD1 CXXC (MBD1 amino acids 172-221) and MBD1 MBD (MBD1 amino acids 1-75) were PCR amplified, using primers to add the AatII and BamHI restriction sites onto the ends of the domains. The domains were amplified to insert a similar number of amino acids into MLL as were taken out with the MLL CXXC domain by digesting 5’MLL with AatII at amino acid site 1149 and BamHI site 1201. After swapping in the alternate domains and sequencing, three way ligations were performed to create MSCVneo-MLL-AF9-Flag CXXC domain swap constructs (Figure 14A).

Retrovirus Production and Titers and Confirmation of Stable Expression of MLL-AF9-Flag CXXC Domain Swap Proteins

MSCV-MLL(MLL CXXC)-AF9-Flag, MSCV-MLL(CGBP CXXC)-AF9-Flag, MSCV-MLL(DNMT1 CXXC)-AF9-Flag, MSCV-MLL(MBD1 CXXC)-AF9-Flag, MSCV-MLL(MBD1 MBD)-AF9-Flag, and MSCVneo DNA constructs were used to transfect Phoenix-Eco cells and produce and titer retrovirus as previously described in the point mutation section (Figure 14B). The same Phoenix-Eco cells were collected and assayed for MLL fusion protein expression by Immunoprecipitation and Western Blot analysis with anti-Flag antibody, also as described previously. All domain swap constructs produced MLL fusion proteins of similar molecular weight (Figure 14C), suggesting there were no problems with the cloning of the constructs or stability of the mutant proteins. In some cases, two bands were observed (Figure 14C), with the lower
Figure 14. Schematic and Expression of MLL-AF9 CXXC Domain Swap Proteins

(A) Schematic of CXXC domain swaps into MLL-AF9-Flag.
(B) Methylene blue stains of retrovirally infected Rat1A cells after 10 days of G418 selection, with virus dilutions indicated.
(C) Western blots of MLL-AF9-Flag proteins with indicated CXXC domain swaps from transfected Phoenix-Eco cells. Fusion proteins were immunoprecipitated and blotted with anti-Flag antibody. Upper bands represent the fusion proteins, while the lower bands are degradation products.
mobility band corresponding to the full length MLL-AF9 fusion protein, and the smaller band likely a degradation product.

**DNMT1 CXXC domain can functionally replace the MLL CXXC domain in an MLL fusion protein in an in vitro immortalization assay**

In order to determine if the alternate CpG binding domains can function in the place of the MLL CXXC domain in an MLL-AF9 fusion protein to confer an enhanced proliferative capacity on murine bone marrow progenitor cells, *in vitro* bone marrow colony assays were performed. As described previously for the CXXC point mutation constructs, the MLL-AF9 CXXC domain swap retroviruses were introduced by spinoculation into c-Kit+ murine bone marrow progenitor cells then plated in methylcellulose with appropriate cytokines. After one week, colonies and cells were counted, then cells were replated for up to four weeks. We hypothesized that only those MLL-AF9 constructs containing CpG-binding domains that could bind to unmethylated CpG DNA *in vitro* would be able to function in MLL-AF9 in a colony assay to promote cell proliferation. As expected, the MLL(MLL CXXC)-AF9 construct was highly transforming (see point mutation data) and the MSCVneo infected cells died in the first couple of weeks (Figure 15A). MLL(DNMT1 CXXC)-AF9 showed robust *in vitro* colony forming activity, with no statistically significant difference to MLL-AF9. However, MLL(CGBP CXXC)-AF9, MLL(MBD1 CXXC)-AF9, and MLL(MBD1 MBD)-AF9 were not transforming and showed very poor colony forming ability. The average numbers of colonies at week 4 for these constructs showed no statistically significant difference to MSCVneo but did show a statistically significant decrease as compared to MLL-AF9. Those few colonies that persisted for these constructs were
Figure 15. MLL-AF9 CXXC Domain Swap Bone Marrow Colony Assay

(A) Average colony numbers for weeks 2-4 after plating primary mouse bone c-Kit+ progenitor cells expressing MLL-AF9 or MLL-AF9 with indicated CXXC domain swaps in methylcellulose, with error bars representing standard error from 6-8 independent biological replicates. (B) Digital photographs showing colony and cell morphologies at week 4 of the colony assay.
often diffuse in nature and typical of differentiated bone marrow cells, with few cell numbers (Figure 15B). The MLL-AF9 and MLL(DNMT1 CXXC)-AF9 cells formed dense compact round colonies typical of bone marrow progenitor cells. Cell counts over the course of the replating experiment revealed that MLL-AF9 and MLL(DNMT1 CXXC)-AF9 both conferred an approximately 870-fold expansion of cell number at each replating, whereas for the remaining constructs the cell expansion ranged from 3 to 7-fold. Cytospins of the bone marrow cells taken from week 3 or 4 of the colony assays also confirm that the transformed MLL-AF9 and MLL(DNMT1 CXXC)-AF9 cells primarily resemble leukemic blasts, while the other constructs produce a more heterogeneous differentiated bone marrow cell population (Figure 15B).

Because the domains from MBD1 are not able to function in MLL-AF9, our hypothesis that MLL CXXC must maintain the ability to bind to unmethylated DNA in order to transform bone marrow cells was supported. However, we expected that both CGBP and DNMT1 CXXC domains would replace the MLL CXXC domain in the context of the MLL-AF9 fusion because the isolated domains bind to the nonmethylated CpG-containing DNA \textit{in vitro}. The ability of a CXXC domain to replace the MLL CXXC domain did not strictly correspond with DNA binding affinity, however, because the CXXC domain from CGBP has a higher affinity for DNA binding as compared to the DNMT1 CXXC domain (Figure 13). These results suggest that CXXC DNA binding affinity alone is not the only function of this domain necessary for MLL-AF9 transformation.
 DNMT1 CXXC domain can functionally replace the MLL CXXC domain in an MLL fusion protein in an in vivo leukemia assay

In order to determine if the alternate CpG binding domains can function in the place of the MLL CXXC domain in an MLL-AF9 fusion protein to promote leukemogenesis, in vivo leukemia assays were performed in mice. MSCVneo, MSCV-MLL(MLL CXXC)-AF9-Flag, MSCV-MLL(CGBP CXXC)-AF9-Flag, MSCV-MLL(DNMT1 CXXC)-AF9-Flag, MSCV-MLL(MBD1 CXXC)-AF9-Flag, and MSCV-MLL(MBD1 MBD)-AF9-Flag retroviruses were used to infect CD45.1+ c-Kit+ murine bone marrow progenitor cells. As described for the CXXC point mutations, the infected cells were transplanted into lethally irradiated recipient mice, and then the mice were monitored for leukemia development.

As seen in Figure 16A, all of the MLL-AF9 mice (n = 7) and six of the seven MLL(DNMT1 CXXC)-AF9 mice developed acute myeloid leukemia with mean latencies of 89 and 73 days, respectively. Median whole blood counts assayed with the Hemavet at time of sacrifice were 16.2 K cells per μL for the MLL-AF9 mice and 17 K per μL for the MLL(DNMT1 CXXC)-AF9 mice. None of the MLL(CGBP CXXC)-AF9 (n=7), MLL(MBD1 CXXC)-AF9 (n=8), and MLL(MBD1 MBD)-AF9 (n=6), or MSCVneo mice (n=4) developed leukemia during the course of the experiment which extended 260 days. The median whole blood count of all the healthy mice at time of sacrifice was 8.8 K per μL. Average whole blood counts for all mice are summarized in Table 8. A small number of the mice that did not develop leukemia (one mouse each receiving MLL(DNMT1 CXXC)-AF9, MLL(MBD1 CXXC)-AF9, or MSCVneo and two mice receiving the MLL(MBD1 MBD)-AF9), died from other causes (thymus or liver tumors).
Table 8. MLL leukemia development in mice reconstituted with MLL-AF9 or MLL(DNMT CXXC)-AF9

<table>
<thead>
<tr>
<th>Construct</th>
<th>Average Latency (days ± s.d.)</th>
<th>Average CBC (K per μL ± s.d.)</th>
<th>Range of CBCs (K per μL)</th>
<th>Average Spleen Weight (grams ± s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLL-AF9 n = 7</td>
<td>89 ± 32</td>
<td>43 ± 50</td>
<td>8 - 140</td>
<td>0.39 ± 0.19</td>
</tr>
<tr>
<td>MLL(DNMT CXXC)-AF9 n = 7</td>
<td>73 ± 36</td>
<td>20 ± 6</td>
<td>14 - 29</td>
<td>0.29 ± 0.13</td>
</tr>
<tr>
<td>MLL(CGBP CXXC)-AF9 n = 7</td>
<td>-</td>
<td>7.5 ± 4</td>
<td>2.6 - 14</td>
<td>0.07 ± 0.01</td>
</tr>
<tr>
<td>MLL(MBD1 CXXC)-AF9 n = 8</td>
<td>-</td>
<td>7.6 ± 4</td>
<td>3.6 - 15</td>
<td>0.08 ± 0.01</td>
</tr>
<tr>
<td>MLL(MBD1 MBD)-AF9 n = 6</td>
<td>-</td>
<td>9.3 ± 6</td>
<td>3 – 16</td>
<td>0.06 ± 0.01</td>
</tr>
<tr>
<td>MSCVneo n = 4</td>
<td>-</td>
<td>11.4 ± 2.4</td>
<td>9 - 13</td>
<td>0.08 ± 0.01</td>
</tr>
</tbody>
</table>

during the course of the experiment. However, these tumor cells did not express the CD45.1 marker from the donor cells. Therefore, the tumors were unrelated to MLL-AF9 expression and were likely a side effect of the radiation dose that the recipients received prior to transplant. The complete blood counts and spleen weights of the mice that died of other causes were not included in the analysis seen in Table 8. The mice that developed leukemia had an increased number of blast cells in their peripheral blood and bone marrow as seen in Figure 16B. Spleens of the leukemic mice were enlarged and had an average weight five times that of the healthy mice (Table 8). Organs sections of the spleen, liver, lungs and kidneys from the MLL-AF9 and MLL(DNMT1 CXXC)-AF9 mice displayed infiltration of leukemic cells, while the other mice had normal organ histology (Figure 16C).
Figure 16. *In vivo* MLL-AF9 CXXC Domain Swap Leukemia Assay
(A) Survival curve of mice transplanted with bone marrow progenitor cells expressing MLL-AF9 or MLL-AF9 CXXC domain swap fusion proteins. (B) Peripheral blood smears and bone marrow samples taken from indicated mice at time of sacrifice. (C) Digital photographs of spleen, liver, lungs, and kidney sections stained with Hematoxylin and Eosin and taken from MLL-AF9 or MLL(DNMT CXXC)-AF9 mice at time of sacrifice due to leukemia or healthy MSCVneo, MLL(CGBP CXXC)-AF9, MLL(MBD1 CXXC)-AF9, or MLL(MBD1 MBD)-AF9 nine months after bone marrow transplants.
Peripheral blood and bone marrow cells taken from the mice at time of sacrifice were subjected to flow cytometry using a variety of antibodies directed at mouse hematopoietic cell surface markers. Anti-CD45.1 antibody was included with most of the groups to use for gating of the donor MLL-AF9 or CXXC domain swap expressing cells. The first group of antibodies was directed at markers of bone marrow myeloid cells and progenitor cells and included anti-CD45.1, CD11b, Gr-1, CD117, and Sca-1. These marks would be expected to be present with MLL-AF9 myeloid leukemia. The second group of antibodies added in marks for B cells and included anti-CD45.1, CD11b, B220, Gr-1, CD117 and CD19. The third group incorporated markers for T cells and included anti-CD45.1 (for bone marrow cell analysis only), CD8a, CD4, CD25, CD3e, CD44 and CD117 (for peripheral blood cell analysis only). The fourth group of antibodies included markers for erythroid, megakaryocyte and NK cells and included anti-CD45.1, CD41, TER119, CD117, and CD49b/DX5. These other antibodies were included to test if any mixed lineage leukemias or other hematopoietic proliferative disorders had developed in any of the mice. As seen in Figure 17, the bone marrow of the leukemic mice most often showed a very high percentage of cells expressing CD45.1, indicating that the transplanted cells expressing MLL-AF9 or MLL(DNMT1 CXXC)-AF9 had a proliferative advantage over any other bone marrow cells present and had repopulated the bone marrow. The healthy mice all showed a very low percentage of CD45.1+ cells in their bone marrow (Figure 17). The CD45.1+ bone marrow and peripheral cells in the leukemic mice were gated upon for analysis. Nearly 100% of these CD45.1+ cells expressed myeloid markers CD11b and Gr-1, while a lower percentage of cells expressed progenitor markers c-Kit/CD117 and Sca-1, as would be expected for acute myeloid
Figure 17. FACS profiles of bone marrow cells from leukemic or healthy mice with MLL-AF9 or CXXC domain swapped MLL-AF9

CD45.1 indicates transplanted cell population expressing the MLL-AF9 or CXXC domain swap construct. The second and third panels show the percentage of CD45.1 positive cells from the MLL-AF9 and MLL(DNMT CXXC)-AF9 mice or total bone marrow cells from the healthy MLL(MBD1 CXXC)-AF9 mouse that express myeloid markers CD11b and Gr-1, and progenitor markers CD117 and Sca-1.
leukemia (Figure 17). The CD45.1+ bone marrow and blood cells were negative for B cell markers B220 and CD19, T cell markers CD3e, CD8a, CD4 and CD25, magakaryocyte marker CD41, erythroid marker TER119, and NK cell marker CD49b. All bone marrow cells expressed CD44 as expected. Whole bone marrow cells taken from a healthy MLL(MBD1 CXXC)-AF9 mouse showed a more mixed population of cells as compared to the leukemic mice with lower percentages of cells expressing CD11b, Gr-1, CD117 and Sca-1 (Figure 17), along with populations of cells expressing the B cell, T cell and other markers (data not shown). The results of the leukemia assay corroborated the colony assay results in that of the CpG-binding domains tested, DNMT1 CXXC alone can functionally replace the MLL CXXC domain in an MLL-AF9 fusion protein. While DNA binding activity of the CXXC domain appears to be essential for MLL-AF9 to function in causing leukemia, it is not sufficient, as the CGBP CXXC domain, which is able to bind unmethylated DNA, is not able to function in the context of MLL-AF9.

**DNMT1 CXXC domain can functionally replace the MLL CXXC domain in an MLL fusion protein to promote overexpression of Hoxa9 in bone marrow progenitor cells**

MLL fusion proteins frequently cause upregulation of HOXA9 in MLL leukemias, and this HOXA9 overexpression has been linked to disease progression. Due to the importance of HOXA9 upregulation to MLL leukemia, we hypothesized that only those MLL-AF9 CXXC domain swap proteins which transform bone marrow cells would be able to cause overexpression of Hoxa9. In order to determine if the MLL-AF9 CXXC domain swap proteins cause upregulation of Hoxa9 expression, RNA was isolated from bone marrow progenitor cells expressing the domain swap fusion proteins after one week
of culture in methylcellulose. Quantitative real-time RT-PCR was performed using Taqman probes to determine Hoxa9 expression levels. MLL-AF9 and MLL(DNMT CXXC)-AF9 proteins promote high levels of expression of Hoxa9, while the other domain swap fusion proteins that were non-transforming do not cause overexpression of Hoxa9, similar to MSCVneo vector infected cells (Figure 18A). The low levels of Hoxa9 expression help explain why the MLL-AF9 fusion proteins with the CGBP CXXC, MBD1 CXXC or MBD1 MBD domain were unable to transform bone marrow progenitor cells.

**MLL fusion proteins can localize to the Hoxa9 locus regardless of which CpG binding domain replaces the MLL CXXC domain**

I had previously shown that while MLL CXXC DNA binding is essential to protect target CpGs in Hoxa9 from becoming methylated and to cause overexpression of Hoxa9 transcription, DNA binding activity of the CXXC domain is not essential for MLL-AF9 proteins to localize to the Hoxa9 locus. Therefore, we hypothesized that all of the CXXC domain-swapped MLL-AF9 proteins would be able to localize to Hoxa9. In order to determine if MLL-AF9 domain swap constructs which do not cause overexpression of Hoxa9 still bind to the locus, chromatin immunoprecipitation (ChIP) assays were performed with mouse bone marrow progenitor cells expressing the MLL-AF9 domain swap constructs. Anti-Flag antibody or IgG control was used to immunoprecipitate chromatin. DNA was isolated then analyzed by real time PCR using primers that localize to the alternate upstream AB exon of Hoxa9, a region which we have previously shown MLL binding. As seen in Figure 18B, all of the MLL-AF9 CXXC domain swap proteins bind to the Hoxa9 locus, regardless of their ability to
Figure 18. *Hoxa9* expression and fusion protein binding to the *Hoxa9* locus in bone marrow cells expressing MLL-AF9 or MLL-AF9 with substituted CXXC domains

(A) Quantitative RT-PCR for *Hoxa9* in bone marrow cells expressing the indicated constructs. Shown are mean relative expression levels from two independent experiments each done in triplicate, with error bars indicating standard deviation. (B) Chromatin immunoprecipitation assay from c-Kit+ mouse bone marrow cells expressing the indicated constructs. Anti-Flag or IgG was used to IP chromatin and real-time PCR was performed in triplicate with primers that amplify an MLL-binding region in the *Hoxa9* locus. Results were normalized to input chromatin. Fold increase in binding with anti-Flag antibody over IgG to the *Hoxa9* locus is shown. Experiments were repeated three times, with error bars indicating standard deviation.
activate transcription of the locus. As expected, the MSCVneo vector control does not show enrichment of anti-Flag over IgG binding to this region of Hoxa9 (data not shown). Surprisingly, the MLL(MBD1 CXXC)-AF9 and MLL(MBD1 MBD)-AF9 proteins seem to have an increased presence on Hoxa9 compared to MLL-AF9 and the other domain swap proteins. These results support the hypothesis that the specific DNA binding activity of the MLL CXXC domain is not absolutely required for MLL-AF9 protein localization to gene targets.
CHAPTER 5
DISCUSSION

Many transcription factors contain protein domains that bind to specific DNA sequence motifs.\textsuperscript{122} This allows transcriptional activation to be highly regulated in cells. When a certain transcription factor is expressed, it will in turn initiate the transcription of a specific set of target genes to which the transcription factor can bind. For example, the general transcription factor TBP, or TATA-binding protein, binds specifically to TATA box-containing gene promoters.\textsuperscript{122} HOX proteins bind to specific target genes through their DNA-binding homeodomains which recognize the DNA motif TAAT in gene promoters.\textsuperscript{69} In \textit{Drosophila}, Polycomb Group and Trithorax proteins are recruited to regulatory sequences called Polycomb response elements, PREs, and Trithorax response elements, TREs, respectively.\textsuperscript{123} MLL and MLL fusion proteins, however, have long stumped researchers as to how they recognize and bind to target genes, and a specific DNA binding consensus sequence has not been identified. The DNA binding function of the MLL CXXC domain was initially hypothesized to contribute, along with the MLL AT hooks, to MLL’s localization to specific target genes. We now know from structural studies that MLL CXXC domain binding centers specifically on one nonmethylated CpG dinucleotide, while flanking DNA sequences contribute to CXXC domain-DNA binding via non-specific interactions only.\textsuperscript{9,10} This suggests that the MLL CXXC domain could interact with any nonmethylated CpG-containing DNA, rather than a specific consensus
sequence that may appear in the promoters of MLL target genes. In this study, we set out to examine the functional roles of the MLL CXXC domain as it contributes to MLL fusion protein leukemogenesis. Through CXXC point mutation and domain swap studies, we hoped to learn how the MLL CXXC domain contributes to MLL-AF9 target localization and promotion of target gene overexpression which leads to leukemia development. We hypothesized that the DNA binding function of the MLL CXXC domain would be essential for MLL fusion proteins to bind to target genes, protect target CpG DNA from methylation, cause overexpression of target genes, give bone marrow progenitor cells an enhanced proliferative capacity \textit{in vitro}, and to cause leukemia \textit{in vivo}.

In the domain swap experiments, we hypothesized that only those DNA binding domains with known properties most similar to the MLL CXXC domain would be able to function in an MLL-AF9 fusion protein.

Previous studies have shown that the presence of the MLL CXXC domain is essential for MLL fusion proteins to promote transformation of bone marrow progenitor cells \textit{in vitro}.\textsuperscript{8} However, these previous studies did not address maintaining the structural integrity of this domain. In the first aim of this study, I wished to determine if the DNA binding function of the MLL CXXC domain is essential for MLL fusion proteins to function. John Bushweller’s laboratory at the University of Virginia solved the structure of the MLL CXXC domain in complex with DNA. This allowed them to identify the CXXC domain amino acids which specifically contribute to binding to the unmethylated CpG DNA, and to generate point mutations that would disrupt these interactions. Importantly, The Bushweller laboratory used NMR to check that each amino acid
substitution would still allow proper zinc-coordination and folding of the CXXC domain. Therefore, I would be able to conclude that any functional effects seen would be due specifically to loss of DNA binding rather than disruption of the structure of the CXXC domain.

I introduced several CXXC point mutations into an MSCVneo-MLL-AF9-Flag retroviral construct and performed colony formation assays with murine bone marrow progenitor cells in order to determine if the DNA binding function of the MLL CXXC domain is critical to MLL fusion proteins to promote oncogenic transformation. The point mutations at DNA interacting amino acids included R1154A, K1185A and Q1187A, which reduced the isolated MLL CXXC DNA binding affinity by 6-15 fold compared to a wild type MLL CXXC domain. The C1188D point mutation, at one of the non-conserved and non-zinc-coordinating cysteines, abolished all detectable binding of the isolated CXXC domain to DNA. While cysteine1188 does not contact target DNA directly, it is in very close proximity to the CpG and it is expected that mutation to aspartic acid introduces a repulsive interaction with the DNA that is strong enough to ablate CXXC DNA binding ability. Mutation of cysteine1188 to alanine (C1188A), however, was inert in that it did not alter the isolated MLL CXXC DNA binding affinity as compared to the wild type domain. The results of the colony assays correlated highly with the \textit{in vitro} DNA binding affinities of the point mutant MLL CXXC domains. The MLL(C1188A)-AF9 mutant behaved similar to an unmutated MLL-AF9, producing many bone marrow colonies. The R1154A, K1185A and Q1187A point mutations gave MLL-AF9 a much reduced colony forming ability, while the C1188D point mutant in MLL-AF9 gave a phenotype similar to MSCVneo, often with no colonies surviving into
the third or fourth week of the assay. From these experiments, I concluded that the DNA binding ability of the MLL CXXC domain is essential for MLL fusion proteins to promote the transformation of bone marrow progenitor cells in vitro.

In order to determine if the MLL CXXC DNA binding function was also essential for MLL fusion proteins to promote leukemogenesis in vivo, the MLL(C1188A)-AF9 and MLL(C1188D)-AF9 retroviral constructs were used to infect murine bone marrow progenitor cells which were then transplanted into irradiated recipient mice. These mice were then monitored for leukemia development. Confirming the in vitro DNA binding and colony assay results, all of the mice with the normal DNA binding C1188A point mutation developed MLL leukemia. These mice had a shorter latency to leukemia development than mice expressing unmutated MLL-AF9. The mice expressing the DNA binding deficient MLL(C1188D)-AF9 construct or MSCVneo did not develop leukemia throughout the course of the experiment. These results confirm that the DNA binding function of the MLL CXXC domain is an essential function that contributes to MLL fusion proteins ability to promote leukemia.

These experiments were followed by in vitro assays designed to understand the mechanism of action of the MLL CXXC domain. The MLL-AF9 CXXC point mutations C1188A and C1188D were again used to determine how these mutant proteins affect and act upon the MLL target gene Hoxa9, which has previously been shown to be an important gene that gets upregulated in MLL leukemias.7,72 Our laboratory has also shown that MLL binds to the Hoxa9 locus, and protects specific target CpGs in this region from becoming DNA methylated.6 MLL(C1188A)-AF9 and MLL(C1188D)-AF9 constructs were expressed in Mll-/- MEFs to assay their ability to protect target CpG
DNA from DNA methylation. As expected, the CpGs that previously showed MLL and MLL-AF4 dependent DNA methylation protection showed higher levels of DNA methylation in cells transfected with MLL(C1188D)-AF9 or the MSCVneo vector, as compared to cells expressing MLL(C1188A)-AF9. From this experiment, we concluded that the MLL and MLL fusion protein function of DNA methylation protection on target CpG DNA is facilitated by the DNA binding activity of the MLL CXXC domain. DNA methylation protection results also correlated with Hoxa9 expression. Real time RT-PCR on bone marrow colony assay cells showed that MLL(C1188A)-AF9 was able to promote overexpression of Hoxa9 and miR196b, while the MLL fusion with the binding deficient CXXC point mutation, MLL(C1188D)-AF9, was unable to cause overexpression of Hoxa9 over the level of cells infected with MSCVneo vector alone. This suggests that the DNA binding function of the MLL CXXC domain, perhaps through its DNA methylation protection function, is necessary for MLL fusion proteins to promote overexpression of target genes.

Following these experiments, the question remained as to whether the DNA binding function of the CXXC domain of MLL is necessary for an MLL fusion protein to localize to its target genes. As the CXXC domain is one of two known DNA binding domains in the MLL protein, one might hypothesize that its binding function must remain intact in order for MLL to locate and bind appropriately to its target genes. However, MLL and MLL fusion proteins interact with many other DNA binding proteins, and the CXXC domain does not bind to a specific DNA sequence outside of one unmethylated CpG, a motif which appears in the promoters of many genes. In order to address this question, Phoenix cells were transfected with MLL(C1188A)-AF9 and MLL(C1188D)-
AF9 constructs and chromatin immunoprecipitation assays were performed. Antibodies including anti-Flag for the fusion proteins, anti-RNA Polymerase II as a mark of active transcription, and anti-trimethylated H3K9 as a mark of repression were used to immunoprecipitate DNA-protein chromatin complexes, then real time PCR was performed using primers to the upstream region of Hoxa9 that shows MLL-dependent DNA methylation protection. Interestingly, we saw no significant differences in the binding of MLL(C1188A)-AF9 and MLL(C1188D)-AF9 to Hoxa9. Similarly, RNA Polymerase II showed just a small insignificant decrease in binding to Hoxa9 in cells expressing MLL(C1188D)-AF9 as compared to MLL(C1188A)-AF9. The biggest difference observed was in the increased presence of the trimethylated H3K9 mark, which is a mark of transcriptional repression, on Hoxa9 chromatin in cells expressing MLL(C1188D)-AF9 compared to cells expressing MLL(C1188A)-AF9.

From these experiments, we concluded that the DNA binding function of the MLL CXXC domain is not critical for MLL fusion proteins to localize to Hoxa9. This therefore suggests that the DNA binding MLL CXXC domain is not involved in the targeting of MLL or MLL fusion proteins to MLL target genes. Instead, we expect DNA interactions via the MLL AT hooks, possibly in combination with interactions with proteins such as Menin and LEDGF, which also bind to DNA and chromatin, to be important for MLL target localization. The MLL CXXC domain, however, does have a necessary function once an MLL fusion protein locates its target gene. The MLL CXXC domain binds to specific unmethylated CpGs in the gene locus and acts to keep these CpGs in the unmethylated state. This in turn helps promote an active state of transcription in the gene locus. When the DNA binding activity of the MLL CXXC
domain is disrupted, as in the case of the MLL(C1188D)-AF9 protein, while the fusion protein still localizes to Hoxa9, the CXXC domain cannot protect the target CpGs from becoming methylated. This DNA methylation then promotes a state of repression for Hoxa9, which includes the recruitment of histone H3K9 trimethylation which helps keep the gene silenced (Figure 19).

Overall, to summarize the results from my first aim, MLL-AF9 requires CXXC DNA binding in order to give bone marrow cells an enhanced proliferative capacity in vitro, to cause leukemia in vivo, to promote overexpression of target gene Hoxa9 and to protect target Hoxa9 CpGs from becoming DNA methylated, but not to localize to Hoxa9. We can conclude that the DNA binding function of the MLL CXXC domain is a critical function for MLL fusion proteins. This conclusion suggests that the MLL CXXC domain could be a potential therapeutic target for MLL leukemia. If a specific inhibitor could be developed that could interact with the DNA binding amino acids within the CXXC domain such that they would no longer be available for DNA binding, then the MLL fusion proteins could not promote leukemogenesis. Our collaborators in the Bushweller laboratory are currently searching for an appropriate small molecule inhibitor. This inhibitor could be tested first in vitro in CXXC binding assays then in MLL-AF9 bone marrow colony assays to determine efficacy, then in vivo in mice to determine any potential side effects.

One potential negative aspect of this approach is that a CXXC inhibitor may also target the CXXC domain of the wild type MLL protein in patients. As MLL is an essential gene for proper blood cell development, this could be detrimental to the normal hematopoiesis process in the patient and affect blood cell counts. However, most current
Figure 19. Model of the regulation of *Hoxa9* locus transcription by the CXXC domain of MLL-AF9

(A) The CXXC domain of MLL-AF9 protects specific CpG sequences within the *Hoxa9* locus from methylation and maintains transcription within the locus; (B) disruption of DNA binding function of CXXC domain by C1188D mutation results in increased methylation of the same CpGs, increased H3K9 methylation, and silencing of *Hoxa9* and *mir196b*. Models were developed in collaboration with John Bushweller, University of Virginia.
chemotherapy treatments also negatively affect blood cell populations, so the proper dosage may be able to distinguish effects on wild type versus MLL fusion proteins. In addition to targeting the MLL CXXC domain, a potential small molecule inhibitor would need to be tested for binding to other proteins that contain CXXC domains. If the drug also binds to any of these proteins, such as DNMT1 or CGBP, additional effects would need to be analyzed, such as changes to global genomic DNA methylation and histone methylation patterns, and the possibility of future diseases, such as cancer, arising from these changes.

During the course of the MLL-AF9 CXXC point mutation in vivo leukemia assays, it was noted that all of the MLL(C1188A)-AF9 mice developed leukemia rather quickly as compared to the unmutated MLL-AF9 mice. The average latencies were 39.5 days for the MLL(C1188A)-AF9 mice and 94 days for the MLL-AF9 mice. The final average white blood cell counts and spleen weights at time of sacrifice were also slightly increased in the MLL(C1188A)-AF9 mice as compared to the MLL-AF9 mice. This more severe phenotype suggests that the MLL(C1188A)-AF9 construct has a more potent oncogenic activity compared to the unmutated MLL-AF9. While it is possible that these effects could be due to differences in viral titers used to infect the donor bone marrow cells, we would not expect this to cause such a big difference in latency. MLL fusion proteins are directly responsible for promoting transformation of bone marrow cells, but it is also thought that a number of other genetic mutations must occur in a cell before a mouse or a human develops leukemia. Therefore, it was unexpected to see all of the MLL(C1188A)-AF9 mice develop leukemia so quickly, and we hypothesized that the C1188A mutation may have an effect on MLL-AF9 function.
The amino acid cysteine contains a thiol group in its side chain, and can often create a disulfide bond with a nearby partner cysteine.\textsuperscript{124} Cysteine1188 in MLL is not thought to have a cysteine partner, as it is not involved with the other CXXC cysteines in coordinating the zinc ions, and because it is in close proximity to the CpG when the domain is bound to DNA, it is not likely to form bonds with cysteines from other protein domains. We hypothesize that cysteine1188 may be acting as a redox sensor for the MLL CXXC domain. Under certain cellular conditions, cysteine1188 may be oxidized. The addition of such a post-translational modification would make cysteine1188 more bulky and would likely inhibit CXXC DNA binding affinity, similar to when cysteine1188 was mutated to aspartic acid. Alternatively, when cysteine1188 is in the reduced state, with its normal unmodified thiol group, it allows the CXXC domain to bind to DNA. We suspect that the introduction of an alanine, which cannot be oxidized, in place of cysteine1188 allows DNA binding to occur at all times; the redox switch thereby becomes bypassed and DNA binding can occur under all oxidative cellular states. Therefore we hypothesize that the MLL(C1188A)-AF9 mice develop leukemia faster because the CXXC domain is able to bind to target DNA at all times, rather than being subject to regulation as a redox sensor or by some other post-translational modification.

MLL CXXC domain swap experiments were performed to gain further understanding of the role of the CXXC domain and how it contributes to MLL leukemogenesis. These experiments allowed us to understand if the MLL CXXC domain is unique in its ability to help MLL fusion proteins promote leukemia, or if similar CXXC domains from other proteins share enough similarities with the MLL CXXC domain that they can also function in MLL-AF9. Several CpG-DNA binding domains, which have
known properties both similar to and different from the MLL CXXC domain, were chosen to swap into MLL-AF9. The first domain chosen was the CXXC domain from the CpG binding protein, CGBP, also known as CFP1. This domain was expected to have the most similarity to the MLL CXXC domain because it is also known to bind only to nonmethylated CpG DNA, and is also present in a protein that promotes active transcriptional chromatin states. The second domain chosen for the domain swap project was the CXXC domain of the maintenance DNA methyltransferase protein DNMT1. It was initially thought that the DNMT1 CXXC domain could bind to hemimethylated DNA, which is the preferential substrate of DNMT1 enzymatic activity. Recent structural data, however, shows that the DNMT1 CXXC domain, like most CXXC domains, binds to unmethylated CpG DNA. Finally two domains were chosen from the repressor methyl binding domain protein MBD1. The MBD1 transcripts can undergo splicing which results in the proteins having either two or three CXXC domains. These CXXC domains from the mouse Mbd1 protein were tested in vitro for DNA binding activity, and only the third CXXC domain, which shows the most sequence identity to the MLL CXXC domain, showed an ability to bind to unmethylated DNA. The first two CXXC domains of Mbd1 were unable to bind to CpG-containing DNA targets, regardless their DNA methylation status. We chose to study the first CXXC domain from MBD1, which has the least amino acid sequence identity to the MLL CXXC domain and high identity to its murine Mbd1 homolog, for swapping into MLL-AF9. Finally, we also chose to swap in the methyl binding domain, or MBD domain, from MBD1. This domain is different from CXXC domains structurally and is only able to bind to methylated CpG DNA. We hypothesized that because the CGBP CXXC
domain seems to have the most similarities to the MLL CXXC domain, it would be able to function in MLL-AF9, while the other domains would behave differently and negatively affect MLL-AF9 function.

While the binding specificities of the different CXXC and MBD domains used in this study were previously characterized, their binding affinities relative to each other were never directly compared. In order to determine the differences in DNA binding affinities of the CpG DNA binding domains, the isolated proteins were purified from bacterial lysates and fluorescence polarization was used to determine their binding to unmethylated CpG DNA. The sequence of the DNA probe was derived from a region of *Hoxa9* that our lab previously showed to interact with the MLL CXXC domain. The results showed that the MLL CXXC domain bound with the strongest affinity to the unmethylated DNA, followed by the CGBP CXXC domain with 7 fold lower affinity, then the DNMT1 CXXC domain at 34 fold lower binding affinity than the MLL CXXC domain. The MBD1 MBD and CXXC domains were unable to bind to the unmethylated DNA as expected.

The MBD1 CXXC domain lacks a KFGG motif that is conserved in many CXXC domains (Figure 20A). This motif has also been shown to be essential to MLL fusion protein function in colony assays, as mutation of this motif to AAAA causes lack of bone marrow colony formation. In addition, the KFGG to AAAA mutation causes the MLL CXXC domain to be unable to coordinate zinc ions. This suggests that the KFGG motif is necessary for proper folding of the MLL CXXC domain, and perhaps other CXXC domains. The KFGG amino acids appear at the sharp turn at the top of the domain and the two glycines are thought to be necessary to facilitate the folding of this structure due
to their small molecular size. In addition to the folding considerations, the lysine in the KFGG motif, which appears at amino acid 1178 in MLL, makes contact with the DNA phosphate backbone and therefore also contributes to CXXC DNA binding. Lysine1178 is a proline in the MBD1 CXXC domain which would be unlikely to functionally substitute for a lysine for DNA binding.

It was interesting that the MLL, CGBP and DNMT1 CXXC domains showed such a large range of DNA binding affinities to the unmethylated DNA, which is the preferred DNA binding target of all of the proteins. We learned from our structural studies which of the MLL CXXC amino acids contribute to DNA binding. The differences in binding affinities between the CXXC domains could be due to differences in these amino acids (Figure 20A). The main recognition motif for the unmethylated CpG in the MLL CXXC domain is the amino acid loop 1182-1188; these amino acids either directly bind to or structurally accommodate the CpG DNA motif. Of those amino acids that directly contribute to binding, lysine1185 and isoleucine1184 appear only in MLL, while glutamine1187 is present in MLL, CGBP and DNMT1, and lysine1186 appears in MLL and DNMT1 but not CGBP. Of the other amino acids in the 1182-1188 loop, arginine1182 and cysteine1188 are present only in MLL, and asparagine1183 is in MLL and CGBP. Of the other amino acids in the MLL CXXC domain that contribute non-specific binding interactions with the backbone or the minor groove of the DNA, several are not conserved in DNMT1 or CGBP: serine1152, lysine1176, lysine1190, lysine1193 and leucine1197. Arginine1150 is conserved between MLL and CGBP, arginine1154 is conserved between MLL and DNMT1, and just two amino acids are conserved between all three proteins: lysine1178 and
Figure 20. CXXC amino acids highlighted by conservation to explain functional specificity of the domain in the context of an MLL-AF9 fusion

(A) CXXC domains are aligned to show sequence conservation. Identical residues to MLL are shaded in grey, while boxed residues highlight differences. Red residues are identical in MLL and DNMT1 but not present in CGBP of MBD1 CXXC. Green residues are conserved in CGBP and MBD1 but not present in MLL CXXC. The blue Cysteine 1188 is conserved in vertebrate MLL paralogs but is alanine in DNMT1. (B) Solution structure of MLL CXXC domain in complex with palindromic CpG DNA (PDB ID: 2KKF). MLL CXXC domain is shown as a space filling model in contact with the ball and stick structure of DNA, in four orientations. The same color scheme is used as in (A). Protein models were provided by Noah Birch.
arginine1192. So, while the spacing of the zinc-coordinating cysteines and a few other amino acids are highly conserved between CXXC domains, many of the amino acids that are involved in binding to DNA in the MLL CXXC are not present in the other CXXC domains. The variation in amino acid sequences between CXXC domains likely contributes to the different DNA binding affinities measured.

The CXXC and MBD domains were introduced into MLL-AF9 fusion proteins in the place of the MLL CXXC domain and tested in vitro in bone marrow colony assays and in vivo in murine leukemia assays. I previously showed that DNA binding activity is essential to MLL CXXC domain function. Therefore, it was expected that the CGBP CXXC domain, which has the closest DNA binding affinity to MLL CXXC, would be able to function in MLL-AF9 bone marrow colony assays and leukemia assays, but the other domains would confer a reduced function to MLL-AF9 and therefore show reduced colony formation and ability to cause MLL leukemia in mice. As expected, the MLL(MBD1 CXXC)-AF9 and MLL(MBD1 MBD)-AF9 constructs were not able to promote bone marrow colony formation or in vivo leukemia development. This confirmed my previous observations that the CXXC domain in MLL must be able to bind to unmethylated CpG DNA for MLL fusion proteins to transform bone marrow cells in vitro and cause leukemia in vivo. Surprisingly, MLL(DNMT1 CXXC)-AF9 was able to promote colony formation and leukemia while MLL(CGBP CXXC)-AF9 was not. Because CGBP CXXC domain has a higher DNA binding affinity than DNMT1 CXXC domain, we can conclude that CpG-DNA binding ability, while essential, is not sufficient for a CXXC domain to contribute to MLL-AF9 function. This suggests that the DNMT1 CXXC domain must share an important function with the MLL CXXC domain that is not
present in the CGBP CXXC domain. I expected that this shared function between the MLL and DNMT1 CXXC domains is the ability to recruit some cofactor that is essential for the activation of MLL target gene expression.

_Hoxa9_ expression in MLL-AF9 CXXC domain swap-expressing bone marrow cells correlated with the colony assay and leukemia assay results. Cells transduced with retroviruses encoding MSCVneo, MLL(CGBP CXXC)-AF9, MLL(MBD1 CXXC)-AF9 and MLL(MBD1 MBD)-AF9 all expressed low levels of _Hoxa9_, while MLL(MLL CXXC)-AF9 and MLL(DNMT1 CXXC)-AF9 transduced cells showed overexpression of _Hoxa9_ as expected. I next asked whether the non-transforming domain swap proteins would be able to localize to _Hoxa9_. In the point mutant studies, I saw that MLL-AF9 localizes to _Hoxa9_ regardless of whether or not the MLL CXXC domain can bind to DNA. However, because the domain swap fusion proteins have many amino acid differences with MLL and may be involved in different protein-protein interactions, I was unsure if they would still be able to bind to the _Hoxa9_ locus. The results of the chromatin immunoprecipitation assays, performed in bone marrow cells from the colony assays, showed that all of the domain swap MLL-AF9 proteins could localize to _Hoxa9_ regardless of which CXXC or MBD domain was present. Surprisingly, the MLL(MBD1 CXXC)-AF9 and MLL(MBD1 MBD)-AF9 proteins showed an increased presence on _Hoxa9_ as compared to MLL-AF9 and the other domain swap proteins. Overall these results again suggest that the CXXC domain of MLL is not required for the targeting of MLL to its target genes. Rather the MLL CXXC domain is more important for binding to unmethylated DNA and possibly recruiting co-factors necessary for gene activation.
To explain why the MLL(MBD1 CXXC)-AF9 and MLL(MBD1 MBD)-AF9 proteins bind so strongly to *Hoxa9*, I hypothesized that the MBD1 MBD and CXXC domains, in the context of MLL-AF9, are able to recruit co-repressor proteins that produce a dominant effect to silence the locus. It has been previously shown that the MBD1 MBD domain binds to co-repressor proteins HP1 and Suv39h1, a histone H3K9 methyltransferase that also recruits repressive histone deacetylase enzymes. This recruitment would act to further enforce transcriptional repression of gene targets such as *Hoxa9* by promoting histone H3K9 methylation and histone deacetylation.\(^{114}\) With regard to the MBD1 CXXC domain swap construct, a region of MBD1 that includes the MBD domain and its first two CXXC domains was previously shown to be able to bind the chromatin assembly factor-1 (CAF1) and the histone methyltransferase SETDB1 which trimethylates histone H3K9 on newly replicated, methylated DNA.\(^{113}\) These studies suggest that both the MBD and CXXC domains of MBD1 are able to recruit co-repressor proteins to target genes, and may also be able to do so in the context of MLL-AF9 when it localizes to *Hoxa9*. It is also possible that it appears that, as compared to MLL-AF9, an increased number of the MLL(MBD1 CXXC)-AF9 and MLL(MBD1 MBD)-AF9 proteins bind to *Hoxa9* because the presence of these proteins causes the formation of heterochromatin with an increased number of nucleosomes and H3 histone proteins present at *Hoxa9*. A chromatin immunoprecipitation experiment using an anti-H3 antibody would need to be performed to confirm this hypothesis.

The repression domain of MLL, which is retained in MLL fusion proteins and includes the CXXC domain, has been shown to interact with multiple proteins. The repression domain includes amino acids 1101-1400 of MLL, and has been further split
into RD1 (amino acids 1101-1250), which includes the CXXC domain (amino acids 1149-1201), and RD2 (amino acids 1251-1400). Our laboratory previously found that multiple repressor proteins bind to the MLL repression domain including the histone deacetylases HDAC1/2, C-terminal Binding Protein CtBP, and the Polycomb group proteins BMI-1 and HPC2.\textsuperscript{32} Full length MLL is primarily thought of as an activator of transcription, but the repression domain of MLL can also act to repress transcription of target genes.\textsuperscript{23} It is possible that the decision for MLL to act as an activator or repressor of transcription depends on which other proteins are recruited to MLL at target genes. So, when proteins such as HDAC, CtBP, and the Polycomb proteins bind to full length MLL, they may facilitate the repressive state of MLL function. In addition, the cyclophilin protein Cyp33 binds to the third PHD finger of MLL and its presence promotes repression of MLL target genes \textit{HOXC8} and \textit{HOXC9}.\textsuperscript{38} Cyp33 also enhances HDAC1 binding to the MLL repression domain,\textsuperscript{32} an effect that would be lost in MLL fusion proteins which do not retain the MLL PHD fingers. HDAC1/2 and BMI-1 bind to both the RD1 and RD2 of MLL, while HPC2 binds to the RD1 region. However, the minimal interaction domains within the MLL repression domain for binding to these repressor proteins has not been resolved. It remains to be determined if the CXXC domain of MLL contributes to binding and if these repressor proteins bind to MLL fusion proteins, or if they can only be recruited to a full length MLL protein.

A series of domain swaps between MLL and MLL family member MLL2, which is not involved in chromosome translocations that cause leukemia, was performed in the context of an MLL-ENL fusion to test for the minimal region of MLL needed for bone marrow cells to transform in colony assays \textit{in vitro}.\textsuperscript{125} The results indicate that the MLL
CXXC domain and a post-CXXC region (through MLL amino acid 1337) are both critical to colony formation.\textsuperscript{125} This suggests that the repression domain of MLL outside of the CXXC domain also plays an essential role in MLL leukemia.

Two recent studies have shown that the MLL repression domain mediates binding of MLL fusion proteins to the activating polymerase associated factor or PAF complex.\textsuperscript{33, 34} The PAF complex (PAFc) promotes transcriptional elongation by recruiting ubiquitin ligases which ubiquitinate histone H2BK120,\textsuperscript{126} a mark which then promotes histone H3K4 and H3K79 methylation.\textsuperscript{127, 128} The study by Muntean et al isolated the minimal regions of MLL required for PAF1 (a component of PAFc) binding to regions flanking the MLL CXXC domain.\textsuperscript{33} They found that MLL amino acids 1115-1154, immediately upstream of the CXXC domain, and a region in RD2 that includes amino acids 1200 through 1400, both bind to PAF1. They also determined that the minimal region of MLL needed to transform bone marrow cells \textit{in vitro} in colony assays includes MLL amino acids 1115-1299 fused to AF9. Since this minimal region includes the regions that bind to PAF1, they concluded that MLL fusion proteins must bind to PAF1 in order to function. In another study, Milne et al found that PAF1 binds to MLL amino acids 1100-1250, the RD1 region, but not to MLL RD2. They also tested the ability of PAF1 to bind to a region of CGBP that includes its CXXC domain and corresponds in length to the full MLL repression domain, and did not see any interaction.\textsuperscript{34} Arginine1153, which is located two amino acids before the first zinc-coordinating cysteine in the MLL CXXC domain, was determined to be an important amino acid in the interaction between MLL and PAFc, as binding decreased in the presence of an R1153A mutation.\textsuperscript{34} Interestingly, MLL arginine1153 is conserved in both the CGBP and DNMT1 CXXC domains. As
CGBP CXXC does not function in the context of MLL-AF9, and PAF1 did not bind to a protein fragment containing the CGBP CXXC domain, we can conclude that the presence of arginine1153 is not sufficient for CXXC domains to function in the context of an MLL fusion protein.

It is possible that PAF1 binding is one factor which may contribute to functional specificity of the either the larger repression domain region or the CXXC domain to MLL fusion function. It is unknown if DNMT1 can interact with PAF1, but if MLL(DNMT1 CXXC)-AF9 retains PAF1 binding but MLL(CGBP CXXC)-AF9 does not, it could explain the difference in transformation potential between the two domain swap proteins. However, additional work is needed to reconcile the differences between the two papers that discovered the MLL-PAF1 interaction and to definitively determine the regions of MLL that are required for PAF1 binding. As neither study isolated the 50 amino acid MLL CXXC domain as responsible for binding to PAF1, it seems that the larger repression domain, perhaps including regions in both RD1 and RD2, is responsible for recruiting the PAF complex. If the CXXC domain is not involved in binding to PAF1, then it would be expected that both MLL(DNMT1 CXXC)-AF9 and MLL(CGBP CXXC)-AF9 would retain PAF1 binding.

Other possible explanations as to why MLL(DNMT1 CXXC)-AF9 can transform bone marrow cells, but MLL(CGBP CXXC)-AF9 and MLL(MBD1 CXXC)-AF9 do not, could include recruitment of other as yet unidentified proteins which bind to both MLL and DNMT1 CXXC domains but are not recruited by CGBP CXXC or MBD1 CXXC. CGBP and MBD1 CXXC domains may also bind to factors inhibitory to MLL-AF9 function. Amino acid differences between the tested CXXC domains are highlighted in
Figure 20. Residues shaded in red are identical between MLL and DNMT1 but not the other CXXC domains, suggesting they may be essential to MLL CXXC function. Residues depicted in green are identical between CGBP and MBD1 CXXC domains, but are not present in MLL or DNMT1, which suggests that the amino acids at these positions in CGBP or MBD1 may be inhibitory to CXXC function in MLL-AF9. The red arginine1154 and lysine1186 appear on the DNA binding face of the CXXC domain, and as previously discussed, contribute to MLL CXXC DNA binding. The red proline1164 and asparagine1196, and green glutamine1157 and glycine1160 appear on the outer face of the domain. These amino acids on the outer surface of the CXXC domain may affect other functions, including protein-protein interactions. As proline1164 and asparagine1196 are common between the MLL and DNMT1 CXXC domains, they may be critical to the recruitment of an essential co-activator, while glutamine1157 and glycine1160 which appear in CGBP CXXC and MBD1 CXXC may inhibit this interaction.

MLL CXXC Cysteine1188, the non-structural, nonconserved cysteine which was discussed in a previous section, is located in close proximity to the CpG when the domain is bound to DNA (amino acid shaded in blue, see Figure 20B). John Bushweller’s laboratory showed by NMR that mutation of cysteine1188 to aspartate causes loss of DNA binding of the isolated CXXC domain. I confirmed that this C1188D mutation causes loss of function of MLL-AF9, likely due to introduction of a repulsive charge from the aspartate that repels DNA from the CXXC domain. However, mutation of cysteine1188 to alanine was inert in that it did not decrease DNA binding of the MLL CXXC domain or reduce transformation of MLL-AF9. Interestingly, the DNMT1
CXXC has an alanine at this 1188 position, which may contribute to the ability of DNMT1 CXXC to function in MLL-AF9.

The DNMT1 and MLL CXXC domains share an ability to bind to unmethylated CpG DNA and can function with similar effects in the oncogenic MLL-AF9 fusion protein, but the two domains come from proteins that have very different functions. MLL principally acts to maintain active transcription of target genes and through its SET domain trimethylates histone H3K4 as a mark and possible promoter of active transcription. My dissertation work suggests that the main function of the MLL CXXC domain is to interact with unmethylated CpG DNA in the promoters of target genes and protect the target CpG residues from becoming DNA methylated, helping to promote an active transcriptional state of the target gene locus. The DNMT1 DNA methyltransferase protein functions to maintain genomic DNA methylation patterns by propagating this transcriptionally repressive mark through DNA replication and cell division. It was originally thought that the DNMT1 CXXC domain could recognize and bind to hemimethylated DNA, which is the target of DNMT1 enzymatic activity. Structural data has now indicated that the DNMT1 CXXC domain binds only to unmethylated CpG DNA. It has been hypothesized that the DNMT1 CXXC domain functions to recognize and bind to unmethylated CpGs that are encountered by DNMT1 following DNA replication, and to protect or sequester these CpGs from the enzymatic activity of the DNMT1 catalytic domain. This autoinhibitory role of the DNMT1 CXXC domain ensures that only newly replicated DNA that is hemimethylated could be targeted by the DNMT1 enzymatic activity. Therefore, both the MLL and DNMT1 CXXC domain share a similar function to protect unmethylated CpG DNA from DNA methylation, but by
different mechanisms. The MLL CXXC domain functions to prevent unmethylated target CpGs from becoming methylated by *de novo* DNA methyltransferase enzymes, while the DNMT1 CXXC domain prevents unmethylated CpGs from becoming methylated by the DNMT1 protein itself.

In summary, the CXXC domain of MLL contributes essential functions to MLL fusion proteins. In this work, I showed that the CXXC domain of MLL is not absolutely required for the localization of MLL fusion proteins to gene targets, as was once hypothesized. It is likely that full length MLL proteins also do not require the CXXC domain for target localization. The DNA binding function of the MLL CXXC domain remains critical to MLL fusion proteins, however. Point mutations that cause loss of MLL CXXC DNA binding cause MLL fusion proteins to lose their ability to cause overexpression of target gene *Hoxa9* and to protect target CpG motifs in the *Hoxa9* locus from becoming DNA methylated, and therefore are unable to promote transformation of bone marrow cells *in vitro* or cause leukemia in mice *in vivo*. CXXC domain swap experiments furthered our understanding of the MLL CXXC domain. I showed that CXXC binding to unmethylated CpG DNA is not the only critical function of the MLL CXXC domain, as the CGBP CXXC domain which retains this binding function is unable to function in an MLL fusion protein. The DNMT1 CXXC domain, while possessing a much lower binding affinity for an unmethylated DNA target compared to the MLL CXXC domain, was still able to functionally replace the MLL CXXC domain in MLL-AF9. This suggests that the MLL CXXC domain is not unique in its ability to function in an MLL fusion protein, and that it must share some essential characteristics with the DNMT1 CXXC domain. These shared functions may include the ability to bind to the
same set of proteins which are necessary for MLL fusion proteins to promote leukemogenesis. By examining the amino acid sequences of the CXXC domains, in collaboration with Noah Birch, a biochemistry student in the laboratory, I identified several differences and similarities between the domains which may contribute to their functions.

The results of this project suggest several avenues for future experiments to explore. It remains to be determined what functional characteristics are shared by the MLL and DNMT1 CXXC domains that are not present in the CGBP CXXC domain. If minimal protein-protein interaction domains within the MLL repression domain to other proteins known to bind in this region, such as PAF1, could be carefully mapped, then the CXXC domain’s contribution to protein binding could be determined. Additionally it would be interesting to use the MLL CXXC domain as bait in a yeast two-hybrid screen or other screening process to determine if any other as yet unidentified proteins interact directly with the MLL CXXC domain. If certain proteins are identified that interact with the minimal MLL CXXC domain, then the other isolated CXXC domains could also be tested to see if they retain this binding ability as well. In addition, the CXXC domain swap MLL-AF9 fusion proteins could also be tested for these interactions to determine if specific proteins bind in this context. Another area currently under investigation by another member of our laboratory is the determination of the role of the redox state of cysteine1188 to MLL fusion protein function. Finally, as the CXXC domain remains an essential feature to MLL fusion proteins, I hope that small molecule inhibitors directed toward the CXXC domain could lead to an effective targeted therapy for MLL leukemia patients. The search for a CXXC domain inhibitor molecule is currently underway.
APPENDIX A:

CLONING STRATEGY FOR CREATING AN MLL CONSTRUCT WITH ALTERED
RESTRICTION ENZYME SITES TO BE USED FOR
SWAPPING CXXC DOMAINS INTO MSCV-MLL-AF9
5’ MLL in pMCS5

Digest MLL with BamHI –

PCR amplify aa1201-1250, adding restriction sites to PCR products with primers

Digest PCR products with BamHI and BglIII
Ligate PCR products into the plasmid digested with BamHI:

```
EcoRI  AatII  No site  BamHI overhangs  SalI
aa 1150 aa 1201 aa 1250 aa 1362

CXXC domain
```

Extra 50 bp of DNA!!!

```
EcoRI  AatII  No site  BamHI site  SalI
aa 1150 aa 1201 1250 aa 1250 aa 1362

CXXC domain

1201 1250 1250
BamHI site
BglII site
```

PCR amplify CXXC Domain:

```
EcoRI  AatII  No site  BamHI  SalI
Start of MLL  aa 1150  aa 1201 aa 1250 aa 1362

CXXC domain
```

```
AatII  BamHI
1150  1201
CXXC
```

```
AatII  BamHI
1150  1201
CXXC
```

Digest PCR products with AatII and BamHI
Take ligation construct with extra DNA:

Digest with AatII and BamHI

Ligate with new PCR products

Final product 5’ MLL with BamHI site moved in pMCS5:
REFERENCES


23. Zeleznik-Le NJ, Harden AM, Rowley JD. 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the


discovery and class prediction by gene expression monitoring. Science.
1999;286:531-537.

72. Ayton PM and Cleary ML. Transformation of myeloid progenitors by MLL

73. Popovic R, Erfurth F, Zeleznik-Le N. Transcriptional complexity of the HOXA9

74. Borrow J, Shearman AM, Stanton VP,Jr, et al. The t(7;11)(p15;p15) translocation in
acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I

75. Yekta S, Shih IH, Bartel DP. MicroRNA-directed cleavage of HOXB8 mRNA.

76. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of

77. Kornberg RD and Lorch Y. Twenty-five years of the nucleosome, fundamental


2000;403:41-45.

81. Jenuwein T and Allis CD. Translating the histone code. Science. 2001;293:1074-
1080.

1981;212:1350-1357.

21.

84. Bird AP and Wolffe AP. Methylation-induced repression--belts, braces, and

85. Tahiliani M, Koh KP, Shen Y, et al. Conversion of 5-methylcytosine to 5-
hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science.
2009;324:930-935.


102. Lorsbach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, Downing JR. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia. 2003;17:637-641.


108. Tate CM, Lee JH, Skalnik DG. CXXC finger protein 1 restricts the Setd1A histone H3K4 methyltransferase complex to euchromatin. FEBS J. 2010;277:210-223.

109. Fatemi M, Hermann A, Pradhan S, Jeltsch A. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J Mol Biol. 2001;309:1189-1199.


VITA

The author, Laurie Ellen (Riesbeck) Risner, was born in Palos Heights, IL on June 6, 1982 to John and Nancy Riesbeck. She received a Bachelor of Science with a Major in Science Preprofessional Studies and a Minor in Science, Technology and Values from the University of Notre Dame, in Notre Dame, Indiana, in May of 2004. As an undergraduate student, Laurie studied the molecular genetics of retinal degeneration using zebrafish as a model organism in laboratory of Dr. Lei Li.

In August of 2004, Laurie joined the Program in Molecular Biology at Loyola University Medical Center in Maywood, Illinois. Shortly thereafter, she joined the laboratory of Dr. Nancy Zeleznik-Le, where she studied the role of the MLL CXXC domain in MLL leukemia. While at Loyola, Laurie received a pre-doctoral training grant from the National Institutes of Health and the Arthur J. Schmitt Dissertation Fellowship, and had the opportunity to present her work at local, regional, and national conferences.

In May of 2009, Laurie married her husband Arthur Risner. They are learning the joys of homeownership in Alsip, IL and enjoy taking their dog Penny on long walks.

After completing her Ph.D., Laurie will join the laboratories of Dr. Lucy Godley and Dr. Nanduri Prabhakar in a joint post-doctoral position at the University of Chicago to study epigenetic changes that are induced by hypoxia and cancer.