Direct Recruitment of BCOR by AF9 forms a fuzzy complex required for MLL-AF9 leukemogenesis

Charles Robert Schmidt
Ashland, Pennsylvania

B.S., Lebanon Valley College, 2010
M.S., University of Virginia, 2014

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

Interdisciplinary Program in Biophysics

University of Virginia
December 2017

_____

_____

_____

_____

_____

Abstract

The Mixed Lineage Leukemia gene (*MLL*) is a target of chromosomal translocations leading to acute leukemias characterized by poor prognoses and early relapse. The mechanism of MLL leukemogenesis involves constitutive expression of MLL targets, and particularly *HOX* genes. Though a number of MLL translocation partners have been identified, the three most common (AF4, ENL, and AF9) account for roughly two-thirds of cases. In MLL-AF9 rearrangements, the AF9 C-terminal ANC1 homology domain (AHD) becomes fused in frame to the N-terminal portion of MLL. The AF9 AHD is intrinsically disordered and has the ability to interact with multiple partners, four of which have been identified. These include AF4 and DOT1L, whose importance in the activation and maintenance of MLL-fusion targets is well-documented. The others, CBX8 and BCOR, are traditionally associated with transcriptional repression. The role of CBX8, a component of polycomb repressive complex 1 (PRC1), in these leukemias has been the subject of several studies. The role of the direct BCOR-AF9 interaction remains uninvestigated.

We present the NMR solution structure of the BCOR-AF9 complex. In addition, we show that the minimal AF9-binding motif of BCOR interacts with AF9 only weakly, while an extended BCOR peptide has a much higher affinity for AF9. Based on this structural and binding information, we have designed a point mutation to selectively disrupt BCOR binding to AF9. This mutation, unlike those which affect the other binding partners, has no effect on colony formation, but instead affects proliferation and prevents development of leukemia *in vivo.* RNAseq analysis demonstrates a change in genes associated with hematopoiesis, cell adhesion, and the ERK cascade, which regulates processes including proliferation, differentiation, and survival.

The C-terminal extension of the BCOR peptide is unassignable in the NMR

experiments used for structural studies. We show that this region likely forms an alpha helix that transiently interacts with AF9 to modulate binding affinity. Interruption of the putative helix by a proline mutation recapitulates binding of the shorter peptide in NMR spectra. The structural disorder inherent in the BCOR C-terminus and its requirement for tight AF9 binding makes BCOR-AF9 a textbook example of a fuzzy complex.

Acknowledgements

Firstly, I'd like to thank Dr. John Bushweller for his mentorship and the opportunity to work in his lab. When I joined John's lab in May 2011, I was looking forward to apply what I had learned as an undergraduate to a problem with practical medical implications. Though I had experience with the biochemical and molecular biological techniques I would need for the ensuing years, my knowledge of the necessary computational methods was sorely lacking. John's mentorship allowed me to fill the gaps in my knowledge and be successful in a structural biology lab. He provided me with the tools needed to grow and develop as a scientist and to think critically about scientific problems and experimental approaches. Ultimately, thanks to working for John and the diverse experiments we perform, I have established a skill set that I could not have acquired elsewhere.

I'd also like to thank my committee members, Dr. Stefan Bekiranov, Dr. Zygmunt Derewenda, Dr. Jochen Zimmer, and Dr. Nancy Zeleznik-Le. Their discussions, suggestions, and questions over the course of my graduate career have been instrumental in driving the progress of my research. Stefan has provided useful advice about RNAseq, a realm which is certainly not my expertise nor the expertise of the Bushweller lab. Zygmunt and Jochen have provided equally useful advice from a structural perspective and on biophysical methods.

I have left Nancy for last intentionally, as she deserves special thanks. Her knowledge of the MLL field has been invaluable in guiding this project. Our collaboration with her lab historically has helped to shape our understanding of MLL leukemias, and it has been particular integral to the developing BCOR-AF9 story. Without the work done by Nancy, Nicholas Achille, and Shubin Zhang this project could not have been completed, and I am grateful not only for their efforts, but also for their efficiency and the

quality of their science. I am also greatly appreciative for the sections of this thesis Nancy provided and the editing she done on other sections in preparation for manuscript submission.

My time in the Bushweller Lab would have been much duller and more difficult if not for the experience and instruction of those who preceded me. I thank Dr. Ben Leach and Dr. Aravinda Kuntimaddi for imparting their expertise in dealing with AF9 and its binding partners and Dr. Mike Regan, Dr. Yunpeng Zhou, and Xuan Qin for general guidance in learning techniques and developing as a scientist. I also thank Adam Boulton who, though not a predecessor, has been just as influential in my education. His computational knowledge has helped to advance mine, and his contributions to data analysis for this project should not go unmentioned. These people also helped to make the proverbial house a home, becoming not only colleagues, but the friends with whom I spent evenings as well as days. I am appreciative, also, of the work of our undergraduate students and particularly Caitlin Dalrymple, who is currently forging on into some of the future directions outlined here. Her attention to detail and dedication to the project have made her a valuable contributor. I am thankful for the work and support of the other members of our lab, including Annie Carlton, Anuradha Illendula, Kiran Tirumala, and Yan Gao. Again, I leave for last someone who deserves a special thanks. Without our lab manager Yan we could not hope to function. She is not only on top of stocking all necessary supplies, sometimes it almost seems as if she has ordered and received reagents before I've managed to hit send on the e-mail. How she manages to do this and experiments with equal competence and success, I'll never know, but my time in the Bushweller lab has been easier and more successful for it.

Further thanks are due to several instructors from my past. I owe to Mrs. Marybeth Sugrue and the late Mr. Vincent Radel, teachers from my time at North

Schuylkill High School, my interest in pursuing a career in science. They brought effective science education to a rural school district that otherwise would have sorely lacked it. I am also indebted to Dr. Walter Patton, my research advisor at Lebanon Valley College, who enabled me to progress to graduate school at the University of Virginia. I remember fondly his biochemistry lectures and his mentorship at the bench, and his teaching style can be credited with the development of my ability to address scientific problems. These three not only inspired me, but equipped me with the tools I needed to be successful as a graduate student.

Finally, I would like to thank my family, including my parents, sister, and extended family, and my girlfriend, Erin Kodis. Over the course of my time in graduate school there have been many instances where I have needed support and encouragement, and they have been there at every turn. I would truly be lost without Erin, who has endured me through good days and bad and supported me in innumerable ways as I've completed my degree.

List of Abbreviations

3C	Chromosome conformation capture
AEP	AF4 family/ENL family/P-TEFb
AF10	ALL1-fused gene from chromosome 10
AF4	ALL1-fused gene from chromosome 4
AF9	ALL1-fused gene from chromosome 9
AHD	ANC1 homology domain
ALL	Acute lymphoid leukemia
AML	Acute myeloid leukemia
APL	Acute promyelocytic leukemia
ASH1L	Asymmetric synthesis of HO 1 like
ASH2L	Asymmetric synthesis of HO 2 like
BBD	BCL6 binding domain
BCOR	BCL6 Corepressor
BM	Bone marrow
BTB	BR-C, ttk, and bab
CBP	CREB-binding protein
CBX8	Chromobox 8
CD	Circular dichroism
Cdk	Cyclin-dependent kinase
CDK9	Cyclin-dependent kinase 9
CREB	cAMP response element binding protein
CSP	Chemical shift perturbation
CTD	C-terminal domain
DOT1L	Disruptor of telomeric silencing 1 like

DotCom	DOT1L complex
E/Meg	Erythrocyte/megakaryocyte
ENL	Eleven-nineteen-leukemia
FA	Fluorescence anisotropy
FITC	Fluorescein isothiocyanate
FYRC	FY-rich, C-terminal
FYRN	FY-rich, N-terminal
GM	granulocyte/monocyte
H3K18ac	Acetylated histone H3 lysine 18
H3K27ac	Acetylated histone H3 lysine 27
H3K36me2	Dimethylated histone H3 lysine 36
H3K36me3	Trimethylated histone H3 lysine 36
H3K4me2	Dimethylated histone H3 lysine 4
H3K4me3	Trimethylated histone H3 lysine 4
H3K79me	Monomethylated histone H3 lysine 79
H3K79me2	Dimethylated histone H3 lysine 79
H3K79me3	Trimethylated histone H3 lysine 79
H3K9ac	Acetylated histone H3 lysine 9
HDAC	Histone deacetylase
HetNOE	Heteronuclear NOE
HOX	Homeobox
HSC	Hematopoietic stem cell
HSQC	Heteronuclear single quantum coherence
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region

IRF3	Interferon regulatory factor 3
KID	Kinase-inducible domain
KIX	Kinase-inducible domain interacting
LEDGF	Lens epithelium-derived growth factor
LMS	Lenz microphthalmia syndrome
MBP	Maltose binding protein
MES	4-Morpholineethanesulfonic acid
MLL	Mixed lineage leukemia
MoRE	Molecular recognition element
MSCV	Murine stem cell virus
MTM	Minimum targeting module
NCBD	Nuclear coactivator binding domain
NELF	Negative elongation factor
NLS	Nuclear localization signal
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
OFCD	Oculofaciocardiodental syndrome
OFMT	Ossifying fibromixoid tumors
P-TEFb	Positive transcription elongation factor b
PAF1	Polymerase associated factor 1
PAF1C	Polymerase associated factor 1 complex
PDB	Protein Data Bank
PCGF	Polycomb group RING finger
PHD	Plant homeodomain

PIC	Preinitiation complex
PPIase	Peptidyl prolyl isomerase
POZ	Pox virus and zinc finger
PRC1	Polycomb repressive complex 1
PTD	Partial tandem duplication
PTM	Post-translational modification
PUFD	PCGF ubiquitin-like fold discriminator
qPCR	Quantitative polymerase chain reaction
RARA	Retinoic acid receptor alpha
RAWUL	RING finger and WD40-associated ubiquitin-like
RBBP5	Retinoblastoma binding protein 5
RDC	Residual dipolar coupling
RNAP2	RNA polymerase II
RRM	RNA recognition motif
SD	Standard deviation
SEC	Super elongation complex
SET	Su(var)3-9, Enhancer-of-zeste, and Trithorax
SIRT1	Sirtuin1
SL1	Selectivity factor 1
SNL1	Speckled nuclear localization 1
SNL2	Speckled nuclear localization 2
SWI/SNF	Switch/sucrose non-fermentable
TAD	Transactivation domain
TBP	TATA-binding protein
TEV	Tobacco etch virus

TFIID	Transcription factor II D
TFIIF	Transcription factor II F
Trx	Trithorax
WDR5	WD repeat domain 5
WT	Wild type
YEATS	Yaf9, Enl, Af9, Taf14, Sas5

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 The Mixed Lineage Leukemia Protein

The mixed lineage leukemia (*MLL*) gene is located on chromosome 11 and encodes a multi-domain protein of nearly 4000 amino acids. It is a homolog of *Drosophila* trithorax (trx) that functions as a human equivalent of the *S. cerevisiae* Set1 protein (Guenther et al., 2005; Tkachuk et al., 1992). Though its association with over 5000 promoters suggests a role in global transcription, MLL is known for its role in positively regulating homeobox (*HOX*) gene expression through maintenance of transcription rather than activation (Guenther et al., 2005; Yu et al., 1995, 1998). Homozygous deficiency of *Mll* is embryonically lethal in mice, and mice with a single functional copy displayed a posterior shift in the anterior boundaries of *Hox* gene expression (Yu et al., 1998). MLL is subject to post-translational cleavage by Taspase1 and subsequent heterodimerization of the cleavage products (Hsieh et al., 2003a). Interaction of the N- and C-terminal fragments is mediated by the FY-rich, N-terminal and FY-rich, C-terminal (FYRN and FYRC, respectively) domains, and disruption of heterodimerization results in loss of stability of the N-terminal fragment and loss of subnuclear localization of the C-terminal fragment (Hsieh et al., 2003b).

The N-terminal fragment of MLL comprises several domains in addition to the FYRN domain, including AT hooks, two nuclear localization domains, a CXXC domain, four PHD fingers, and a bromodomain (Figure 1-1A). Two of these domains confer DNA--binding ability. The AT hooks, which share homology with minor groove-binding AT hooks, bind cruciform DNA independent of sequence (Zeleznik-Le et al., 1994). The zinc finger CXXC domain binds unmethylated CpG islands at previously active promoters, protecting them from methylation (Birke et al., 2002; Cierpicki et al., 2010; Erfurth et al., 2008). Between these two domains are two speckled nuclear localization domains,

SNL1 and SNL2, which drive the punctate distribution of MLL in the nucleus (Yano et al., 1997). C-terminal to the CXXC domain are four plant homeodomain (PHD) fingers, with a bromodomain between PHD3 and PHD4. The third finger performs the canonical PHD function of binding histone H3 lysine 4 di- and tri-methylation marks (H3K4me2/3), an activity augmented by the bromodomain, which unlike others of its kind does not read acetyllysine (Milne et al., 2010; Park et al., 2010; Wang et al., 2010). In addition, the cassette formed by PHD1-3 has been reported to homodimerize (Fair et al., 2001).

In addition to the FYRC domain, the C-terminal fragment of MLL is contains a transactivation domain (TAD) and a Su(var)3-9, Enhancer-of-zeste, and Trithorax (SET) domain. The SET domain is a histone methyltransferase responsible for writing the histone H3K4 mark at the promoters of MLL target genes such as *HOXA9* and *MEIS1* (Guenther et al., 2005; Milne et al., 2002).

MLL interacts with several other proteins. The N-terminus is involved in interactions with MENIN and lens epithelium-derived growth factor (LEDGF), the latter of which binds di- and trimethylated histone H3 lysine 36 (H3K36me2/3) via its PWWP domain (Botbol et al., 2008; Yokoyama et al., 2005). LEDGF demonstrates a preference for the trimethylated species, a mark whose presence is maintained by asymmetric synthesis of HO 1 like (ASH1L) (Zhu et al., 2016). The CXXC domain interacts with the polymerase associated factor 1 (PAF1) complex (PAF1C), an elongation complex required for targeting to the *HoxA9* locus in mice (Milne et al., 2010). Recruitment of MLL to its target promoters requires a minimum targeting module (MTM) consisting of the LEDGF PWWP domain and the MLL CXXC domain (Okuda et al., 2014). In addition to H3K4me2/3, MLL PHD3 binds the RNA recognition motif (RRM) of CYP33(Fair et al., 2001; Park et al., 2010; Wang et al., 2010). The peptidyl prolyl isomerase (PPIase) domain of CYP33 catalyzes the *cis-trans* isomerization of a proline residue in the linker

between PHD3 and the bromodomain. This opens the interface between the two domains, allowing for CYP33 RRM binding to PHD3. RRM binding in turn modulates repression of MLL targets via recruitment of histone deacetylases (HDACs) and a decrease in the affinity of PHD3 for H3K4me2/3, exposing it to potential demethylation (Park et al., 2010; Wang et al., 2010). The TAD interacts with the kinase-inducible domain interacting (KIX) domain of CREB-binding protein (CBP), facilitating its binding to phosphorylated cAMP response element binding protein (CREB) to effect MLL-dependent transcriptional activation (Ernst et al., 2001). The C-terminal region of MLL forms a histone methyltransferase complex containing WD repeat domain 5 (WDR5), asymmetric synthesis of HO 2 like (ASH2L), and retinoblastoma binding protein 5 (RBBP5) (Yokoyama et al., 2004).

1.2 MLL-rearranged leukemias

MLL rearrangements of three types have been observed to cause transcriptional misregulation. These include reciprocal translocations, complex translocations, and partial tandem duplications (PTDs) (De Braekeleer et al., 2010; Yu et al., 1996). Reciprocal translocations are the most common rearrangements, and most involve the fusion of the N-terminal regions of MLL in frame to the C-terminal portion of its partner. There are a few examples of the reverse, including AF4-MLL, which are leukemogenic, but these remain largely unstudied (Bursen et al., 2010; Rowley, 1992). The translocations of the former type now involve over 90 identified fusion partners (Meyer et al., 2017). Despite the large number, 9 translocation partners account for approximately 90% of MLL leukemias (Meyer et al., 2017). Expression of these chimeric proteins leads to acute myeloid and acute lymphoid leukemias (AML and ALL, respectively) characterized by poor prognoses and early relapse (DiMartino and Cleary, 1999; Rowley, 1992). Though not an absolute determinant, the identity of the MLL fusion

partner may have some correlation to the lineage of the leukemia (Krivtsov and Armstrong, 2007). While MLL fusion proteins account for approximately 10% of all human leukemias, they are found in more than 70% of infant leukemias and are over-represented in adults treated with topoisomerase II inhibitors (Biondi et al., 2000; Muntean and Hess, 2012).

The break point in MLL varies, but it is virtually always found from exon 8 to intron 11 and is highly concentrated just before exon 12 in those treated with topoisomerase II inhibitors (Wright and Vaughan, 2014). As a result, MLL fusion proteins retain the N-terminal regions of MLL and thus the interaction with MENIN and LEDGF (Figure 1-1B). LEDGF is required for leukemogenesis, but not for normal hematopoiesis (El Ashkar et al., 2017). In addition, the fusions retain the CXXC domain and its interaction with PAF1C. In fact, the LEDGF PWWP domain and the MLL CXXC domain alone fused to a translocation partner are sufficient for recognition of MLL target chromatin (Okuda et al., 2014). Importantly, the fusion proteins lack PHD finger 3, reinsertion of which drastically decreases the serial replating capacity of MLL-ENL cells (Chen et al., 2008).

In MLL-rearranged leukemia, transcription is activated and maintained by recruitment of the AF4 family/ENL family/P-TEFb (AEP) complex (also known as the super elongation complex, or SEC) and the DOT1L-AF10 family-ENL family complex (DOT1L complex or DotCom) to these targets (Mohan et al., 2010a; Okuda et al., 2017; Yokoyama et al., 2010). Members of the AEP complex account for nearly 70% of MLL rearrangements (Meyer et al., 2017). Fusion to one of these proteins bypasses the need for AEP recruitment and leads to constitutive activation of MLL targets (Yokoyama et al., 2010).

Positive transcription elongation factor b (P-TEFb), composed of cyclin-dependent kinase 9 (CDK9) and Cyclin T1/2, phosphorylates the C-terminal domain (CTD) of RNAP2 leading to release of negative elongation factor (NELF) and the transition from transcription initiation to elongation (Peterlin and Price, 2006). This function led to the hypothesis that P-TEFb recruitment by MLL-AEP fusions may direct the activation of transcription by release of promoter-proximal pausing (Mohan et al., 2010b). However, recruitment of P-TEFb alone and fusion of the MLL MTM directly to the P-TEFb binding platform of AF4 were both insufficient for transformation (Okuda et al., 2015; Yokoyama et al., 2010). The AF4 pSER domain interacts with selectivity factor 1 (SL1), which comprises TATA-binding protein (TBP) and TAF1A-D, to load TBP onto the TATA element. This is the rate limiting step in MLL fusion-dependent gene activation and leads to the formation of the RNA polymerase II (RNAP2) preinitiation complex (PIC) (Okuda et al., 2015). DOT1L is responsible for the mono-, di-, and tri-methylation of histone H3 lysine 79 (H3K79me/me2/me3), an activity that is promoted by association with AF10 (Deshpande et al., 2014; Feng et al., 2002; Jones et al., 2008). The resultant high level of H3K79me2/me3 protects against sirtuin1 (SIRT1)-mediated gene silencing and maintains an active transcriptional environment (Chen et al., 2015). Both transcriptional activation through AF4 recruitment and transcriptional maintenance through DOT1L recruitment are required for MLL leukemias (Kuntimaddi et al., 2015; Lokken et al., 2014; Okuda et al., 2017).

The *HOX* genes are, as mentioned above, important targets of MLL. They are transcription factors which direct hematopoietic differentiation and whose expression is high in hematopoietic stem cells (HSCs), but becomes decreased as differentiation progresses (Slany, 2009). Maintenance of aberrant *Hoxa9* expression is essential for MLL leukemogenesis *in vivo* (Ayton and Cleary, 2003). Transformation also requires

continued expression of the *Hoxa9* dimerization partner *Meis1* and the redundant function of the related proteins Pbx2 and Pbx3 (Wong et al., 2007).

It has been reported that expression of the wild type *MLL* allele is required for leukemogenesis, specifically as induced by MLL-AF9 (Thiel et al., 2010). However, more recent work has demonstrated no change in colony forming ability upon *MLL1* knockdown, but rather indicates a dependence on MLL2 (Chen et al., 2017). Loss of MLL2 negatively affected cell growth and survival, while loss of both MLL1 and MLL2 impaired colony formation and further reduced cell viability. In addition to MLL2, a requirement has been demonstrated for TET1, another CXXC domain-containing protein (Huang et al., 2013). Knockdown of *TET1* reduced colony formation, while overexpression had the opposite effect. Interestingly, *TET1* knockdown could be rescued by expression of MLL targets *HOXA9*, *MEIS1*, or *PBX3*.

Recently, two attempts at targeted MLL leukemia therapy have shown some promise. EPZ-5676, a small molecule targeting the enzymatic region of DOT1L and inhibiting its activity, leads to loss of H3K79me2 at MLL fusion target loci and extension of survival in a mouse xenograft model (Daigle et al., 2011). Two small molecules, MI-463 and MI-503, targeting the MENIN-MLL interaction block progression of MLL leukemia without affecting normal hematopoiesis (Borkin et al., 2015). They extend survival in a mouse model and reduce clonogenic efficiency in primary patient samples. These targeted therapies may be the beginning of a new era in MLL treatment, as MLL fusions provide several attractive targets for inhibition.

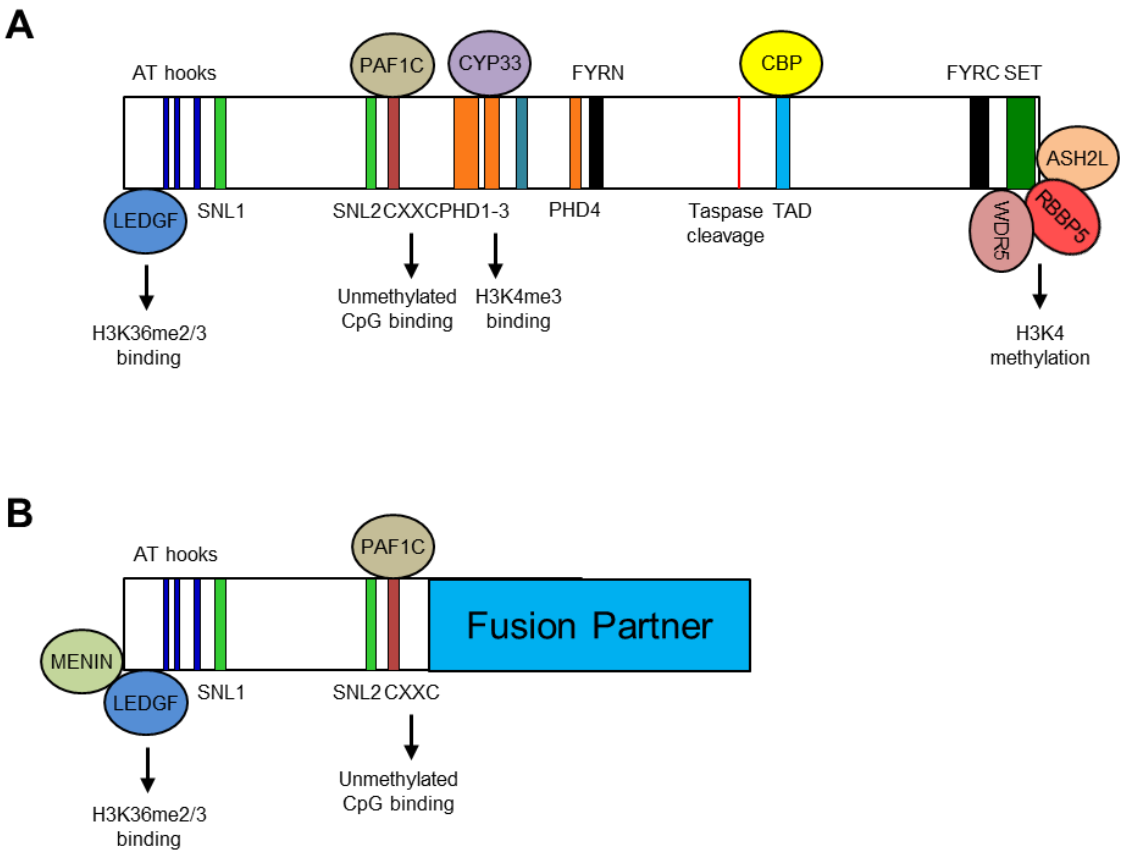


Figure 1-1. Schematic diagram of MLL wild type and fusion proteins.

(A) Schematic diagram of wild type MLL and its constituent domains, including protein, DNA, and histone interactions.

(B) Schematic diagram of the MLL protein as fused to one of over 90 fusion partners.

1.3 The MLL fusion partner AF9

AF9 shares significant homology with eleven-nineteen-leukemia (ENL) and is a member of the YEATS family. Both comprise an N-terminal YEATS domain (named for its founding members, Yaf9, Enl, Af9, Taf14, and Sas5) and a C-terminal ANC1 homology domain (AHD) separated by a serine- and proline-rich region containing a nuclear localization signal (NLS) (Cairns and Henry, 1996; Iida et al., 1993; Nakamura et al., 1993; Yokoyama et al., 2010). AF9 and ENL are among the most common MLL fusion partners and, as YEATS family members, can both be components of the AEP complex (Muntean and Hess, 2012).

ANC1, a yeast protein with significant similarity to AF9 and ENL, forms part of the SWI/SNF, TFIID, and TFIIF complexes and is required for transcriptional control (Cairns and Henry, 1996). Overexpression of AF9 in HSCs results in increased expression of *GATA1, GATA2, SCL, and GFI1b*, genes which are involved in erythrocyte and megakaryocyte (E/Meg) development, and therefore an increased number of erythroid and megakaryocytic progenitors (Pina et al., 2008). AF9 overexpression additionally leads to decreased expression of *PU.1* and *GFI1*, which are associated with granulocyte/monocyte (GM) development (Pina et al., 2008). Null mutation of *Af9* in mice results in perinatal lethality, but has no effect on hematopoiesis (Collins et al., 2002). Instead, these mice were marked by skeletal anomalies due to changes in the posterior boundary of *Hoxd4* expression.

The AF9 YEATS domain consists of the first 140 residues of the human protein and, like other YEATS domains, is involved in recognition of chromatin modifications (Schulze et al., 2009). The first modification bound by the AF9 YEATS domain to be identified was acetyllysine, specifically histone H3 lysine 9 (H3K9ac), with lower affinity for the same modification on histone H3 lysines 27 and 18 (H3K27ac and H3K18ac,

respectively) (Li et al., 2014). Structurally, this involves an eight-stranded immunoglobin fold and a serine-lined aromatic cage for acetyllysine binding, distinct from other known acetyllysine readers such as the bromodomain. The YEATS domain provides a link between histone acetylation (read by AF9 YEATS) and H3K79 methylation by DOT1L (recruited by AF9 AHD, discussed below). More recently, it was shown that the AF9 YEATS domain has a preference for crotonylated lysine over the acetylated form (Li et al., 2016). The interaction between the YEATS domain and crotonylated lysine positively regulates gene expression.

The AF9 AHD, which is retained in MLL fusions, is intrinsically disordered, but undergoes coupled folding and binding upon interaction with one of its binding partners (Leach et al., 2013). Fused to MLL, it is necessary and sufficient for immortalization of hematopoietic progenitors (Collins et al., 2000). This domain is able to recruit AF4 and DOT1L, both of which support transcriptional elongation (Bitoun et al., 2007; Erfurth et al., 2004; Steger et al., 2008; Zhang et al., 2006). Interestingly, the AF9 AHD also recruits the BCL6 corepressor (BCOR) and chromobox homolog 8 (CBX8), which are implicated in transcriptional repression (Bárdos et al., 2000; Hemenway et al., 2001; Huynh et al., 2000; Leach et al., 2013; Srinivasan et al., 2003).

The roles of AF4 and DOT1L in MLL leukemias have been discussed above, and BCOR will be discussed at length, as it forms the basis of this work. CBX8 is one of five mammalian orthologs of *Drosophila* polycomb, each of which contains an N-terminal chromodomain capable of binding trimethylated histone H3 lysine 9 and 27 (H3K9me3 and H3K27me3) with variable affinity (Kaustov et al., 2011). In mouse embryonic stem cells (ESCs), a Cbx7-containing polycomb repressive complex 1 (PRC1) maintains pluripotency by repressing expression of targets including Cbx8 (Morey et al., 2012; O'Loghlen et al., 2012). Replacement of Cbx7 with Cbx8 in these complexes is required

for ESC differentiation and the transition to gene activation that accompanies it (Creppe et al., 2014). CBX8 contributes to repression of the *INK4a/ARF* tumor suppressor, bypassing oncogene- and stress-induced senescence (Dietrich et al., 2007; Maertens et al., 2009). It is up-regulated in esophageal and colorectal cancers and associates with WDR5, a histone methyltransferase component, to promote breast cancer tumorigenesis (Chung et al., 2016; Tang et al., 2014; Xiao et al., 2014). Additionally, CBX8 is recruited by poly (ADP ribose) polymerase (PARP) to sites of DNA damage, where it is involved in repair (Chou et al., 2010; Oza et al., 2016; Xiao et al., 2014). It has been studied in connection with MLL rearrangements and has been reported to be required for both MLL-AF9 and MLL-ENL leukemias (Maethner et al., 2013; Zeisig et al., 2011).

The means by which each of these proteins is bound and released at appropriate times is not known, but could be influenced by local effective concentrations or post-translational modifications (PTMs). Indeed, phosphorylation of Af9 at S435 has been shown to prevent interaction with Dot1a in mice, and phosphorylation of AF4 at T766 greatly reduces its affinity for AF9 (Leach et al., 2013; Zhang et al., 2007).

1.4 BCOR, the BCL6 Corepressor

1.4.1 BCOR function

BCOR is an alternatively spliced transcriptional corepressor that is ubiquitously expressed in adult mice and humans (Ahmad et al., 2003; Huynh et al., 2000; Srinivasan et al., 2003; Wamstad and Bardwell, 2007). The canonical isoform is a 1755 amino acid protein containing a BCL6 binding domain (BBD), an AF9 binding domain, three ankyrin repeats, and a PCGF ubiquitin-like fold discriminator (PUFD) domain (Ghetu et al., 2008; Huynh et al., 2000; Junco et al., 2013; Srinivasan et al., 2003). Alternative splicing of BCOR results in four isoforms in mice and humans, designated a through d in the former. Full length BCOR, including all 10 exons, constitutes isoform a. Exon 5 can be

spliced out, resulting in isoform b, or an alternative splice acceptor site can be used near the 5' end of exon 8, resulting in isoform c. Alternatively, both of those can occur, producing isoform d (Wamstad and Bardwell, 2007). Yeast two-hybrid results reveal that only isoforms a and b, which retain the entirety of exon 8, can interact with AF9 (Srinivasan et al., 2003). BCOR was initially identified in another yeast two-hybrid screen using the first 418 amino acids of BCL6 as bait (Huynh et al., 2000). It augments BCL6-mediated repression 2- to 3- fold in reporter assays, is unable to repress transcription on its own, and lacks a recognizable DNA binding domain, earning it the label of corepressor (Ahmad et al., 2003; Huynh et al., 2000).

BCL6 is a transcriptional repressor that interacts with the corepressors N-CoR, SMRT, and BCOR via its Pox virus and zinc finger/BR-C, ttk, and bab (POZ/BTB) domain (Dhordain et al., 1997; Ghetu et al., 2008; Huynh and Bardwell, 1998; Huynh et al., 2000). This domain is necessary for repression, and interaction with the aforementioned corepressors requires its homodimerization (Ahmad et al., 2003; Albagli et al., 1996; Chang et al., 1996; Deweindt et al., 1995; Ghetu et al., 2008; Huynh and Bardwell, 1998; Huynh et al., 2000; Seyfert et al., 1996). Crystallization of the BCOR BBD with the POZ domain of BCL has revealed that, like SMRT, BCOR binds to the lateral groove created by self-association of the POZ domain (Ahmad et al., 2003; Ghetu et al., 2008). Unlike N-CoR and SMRT, however, BCOR interacts only with BCL6 and not with other POZ domain-containing proteins (Huynh et al., 2000). BCOR and SMRT can simultaneously bind in symmetrical lateral grooves of the same POZ domain homodimer to form a ternary complex at promoters enriched for H3K27me3 to effect repression (Hatzi et al., 2013).

BCOR is similar to CBX8 in that it is also a member of a PRC1 complex, albeit a variant rather than the canonical form. These variants are identified by the polycomb

group RING finger (PCGF) protein that distinguishes the complex (Gao et al., 2012). BCOR interacts with RING1A/B, PCGF1/NSPC1, KDM2B, SKP1, RYBP, and RNF2 to form PRC1.1 (Farcas et al., 2012; Gao et al., 2012; Gearhart et al., 2006; Sánchez et al., 2007). Crystallization of the BCOR PUFD domain with the RING finger and WD40-associated ubiquitin like (RAWUL) domain of PCGF1 has demonstrated that steric interference allows BCOR to discriminate among PCGF proteins, permitting interaction with PCGF1 (and, in theory PCGF3), but not PCGF2 and PCGF4 (Junco et al., 2013). Interestingly, a recent study has shown that BCOR-containing PRC1.1 is recruited to bivalent promoters by the combined effort of CBX8 and BCL6 to promote the formation of germinal center (GC) B cells (Béguelin et al., 2016).

1.4.2 BCOR in disease

BCOR is an X-chromosomal gene whose mutation is causative in oculofaciocardiodental syndrome (OFCD), a form of microphthalmia characterized by multiple developmental abnormalities including congenital cataracts, pointed nose with separated cartilages, cleft palate, atrial and/or ventricular septal defect, radiculomegaly (particularly of the canines), oligodontia, and others (Gorlin et al., 1996; Marashi and Gorlin, 1990; Ng et al., 2004; Schulze et al., 1999; Wilkie et al., 1993). OFCD is inherited in an X-linked dominant pattern (Gorlin et al., 1996; Hedera and Gorski, 2003). As a result, those *BCOR* mutations which cause OFCD in females are hemizygous lethal in males (Ng et al., 2004). X chromosome inactivation in females leads to differential expression of wild type (WT) and mutated *BCOR*, which in turn leads to a wide range in OFCD symptom severity (Ng et al., 2004). Despite the ubiquitous expression of *BCOR*, a study over the course of mouse development reveals that *Bcor* is strongly expressed in extraembryonic tissue during gastrulation, but following embryonic turning expression in the embryo proper is markedly increased (Wamstad and Bardwell, 2007). Whole

mount and section *in situ* hybridization indicate significant levels of *Bcor* transcripts in the limb buds, branchial arches, neural tube, notochord, eye (retina, lens, and eyelid boundary), olfactory epithelium, and teeth primordium of developing mice, among other structures (Wamstad and Bardwell, 2007). The correlation of this pattern of expression with tissues affected in OFCD suggests that expression of mutant *Bcor* during embryogenesis may be responsible for the resultant OFCD phenotype. Studies of X chromosome inactivation in OFCD patients have revealed skewed methylation of the chromosome harboring the mutant *BCOR* allele in peripheral blood lymphocytes and leukocytes, indicating preferential survival of cells expressing WT *BCOR* (Hedera and Gorski, 2003; Ng et al., 2004). The requirement for the WT protein suggests that, like AF9, BCOR plays a role in hematopoiesis, as those peripheral blood cells expressing mutated *BCOR* generally do not persist.

BCOR mutations are also causative in Lenz microphthalmia syndrome (LMS), which like OFCD is characterized by microphthalmia and skeletal anomalies, but also results in mental retardation (Glanz et al., 1983; Ng et al., 2004). Like OFCD, LMS is in inherited in an X-linked manner, but in a recessive rather than dominant pattern (Ng, 2014; Ng et al., 2004). One form of *MAA2*-associated LMS has been associated with BCOR mutation, and *BCOR* remains one of only two genes whose mutation has been associated with LMS (Ng, 2014; Ng et al., 2004).

BCOR translocations have been identified in several tumor types. Fusion of *BCOR* to retinoic acid receptor alpha (*RARA*) in t(X;17)(p11;q12) occurs in a variant of acute promyelocytic leukemia (APL), the *BCOR-CCNB3* fusion has been found in a Ewing-like subtype of bone sarcoma, and translocations resulting in *ZC3H7B-BCOR* fusions have been discovered in endometrial stromal sarcomas (ESS) and ossifying fibromixoid tumors (OFMT) (Antonescu et al., 2014; Panagopoulos et al., 2013; Pierron

et al., 2012; Yamamoto et al., 2010). *ZC3H7B-BCOR* and *BCOR-MAML3* have been identified in undifferentiated small blue round cell sarcomas (Specht et al., 2016). Other mutations have been found in a range of cancers, including medulloblastoma, retinoblastoma, AML, and myelodysplastic syndrome (Damm et al., 2013; Grossmann et al., 2011; Pugh et al., 2012; Zhang et al., 2012). More recently, BCOR has been implicated in regulation of myeloid cell proliferation and differentiation and has been found necessary for MLL-AF9 leukemogenesis (van den Boom et al., 2016; Cao et al., 2016).

1.5 Protein disorder

1.5.1 Characterization of intrinsically disordered proteins

For some time, the view of proteins was that they adopted defined three-dimensional structures which were vitally connected to their functions. Though still the case for some proteins, particularly enzymes, this paradigm has become increasingly challenged. In fact, many genes encode proteins or segments of proteins that lack stable three-dimensional structure but are still functional, referred to as intrinsically disordered proteins (IDPs) (Dyson and Wright, 2005). One early view of IDPs was that they could be divided into two groups: intrinsic coils, which possessed no or nearly no secondary structure, and pre-molten globules, which displayed some residual secondary structure (Uversky, 2002b). A later view asserts that proteins exist on a continuum from the traditional tightly folded structures to highly extended, disordered states (Dyson and Wright, 2005). IDPs, found toward the latter end of the continuum, have also been described as ensembles of rapidly interconverting structures which are characterized by differences in backbone torsion angles (Tompa, 2005).

Compared to ordered domains, IDPs are characterized by high levels of charged and polar amino acids, such as E, K, P, Q, R and S, which promote disorder, and low

levels of hydrophobic amino acids such as F, I, L, M, V, W, Y, which promote order (Dunker et al., 2002; Romero et al., 2001; Vucetic et al., 2003). This results directly in their low mean hydrophobicity and high net charge (Uversky et al., 2000). The relative lack of hydrophobic residues may prevent the tight packing observed in globular, ordered proteins, thereby contributing to disorder (Romero et al., 2001). This biased composition allows for prediction of intrinsic disorder based only on primary amino acid sequence by PONDR, DISOmodePRED, and GLOBPLOT, among others (Bracken et al., 2004; Landing et al., 2003; Obradovic et al., 2003; Ward et al., 2004). Predictions indicate that some proteins may be entirely disordered, while others contain intrinsically disordered regions (IDRs) in addition to structured domains (Wright and Dyson, 2015). Roughly a third of eukaryotic proteins have been estimated to contain disordered regions longer than 30 residues, a testament to the frequency of IDP occurrence in the proteome (Ward et al., 2004).

IDPs may not be fully unstructured, and even in the most unfolded states show a propensity to form localized, pre-organized secondary structure which has been predicted to favor the binding process (Dyson and Wright, 2002a; Fuxreiter et al., 2004; Uversky et al., 2000). Some of these regions, called molecular recognition elements (MoREs), mediate binding events upon encountering a biological target (Oldfield et al., 2005). As a result of this interaction, the IDP adopts a defined structure in a process known as coupled folding and binding (Dyson and Wright, 2002b; Spolar and Record, 1994). A frequently-cited example of this phenomenon is the CREB kinase-inducible domain (KID), which is intrinsically disordered on its own and in the full-length protein, but folds to form a pair of helices upon binding its target domain in CBP (Radhakrishnan et al., 1997, 1998; Richards et al., 1996). Initiation of a binding event by short motifs allows IDPs to adopt different structures when binding different targets (Dyson and

Wright, 2005). One example of this also involves CBP, the nuclear coactivator-binding domain (NCBD) of which folds into very different structures when binding the activation domain of p160 nuclear receptor coactivators or interferon regulatory factor 3 (IRF3) (Radhakrishnan et al., 1997, 1998; Richards et al., 1996). Another example is a disordered region near the C-terminus of p53 which adopts extended, α-helical, or β-strand structures depending on binding partner (Dunker et al., 2008).

Stabilization of the secondary structure of these motifs may provide a small increase in binding affinity, but may also destabilize the complex, underscoring the importance of disorder for the function of these proteins (Baek et al., 2011; Bienkiewicz et al., 2002). Some disordered sequences, including MoREs, may gain stability in a cellular context due to a crowding effect resulting from the high concentration of macromolecules in that environment (Dedmon et al., 2002; Ellis, 2001; Sorenson et al., 2004). This highlights the importance of environment to protein structure, which has become a focus of recent studies (Hausrath and Kingston, 2017). For instance, ionic strength and identity modulate residual structure content in IDPs, affecting association and dissociation rates and shifting binding affinity up to six fold (Wicky et al., 2017).

1.5.2 Function and regulation of intrinsically disordered proteins

IDPs, so recently characterized, have been found to fulfill a wide variety of roles. They are involved in transcriptional regulation and gene expression, signal transduction, cell cycle regulation, and chaperone activity (Dunker et al., 2002; Galea et al., 2008; Iakoucheva et al., 2002; Tompa and Csermely, 2004; Uversky, 2002a; Ward et al., 2004; Wright and Dyson, 1999). They achieve these various functions through their ability to fluctuate over an ensemble of states or binding to partner molecules in a structurally adaptive process (Tompa, 2005).

IDPs often act as or interact with hubs for protein-protein interaction (Dunker et

al., 2005; Kim et al., 2008) and function integrally in signaling complex assembly through reversible protein-protein interactions and ultra-stable amyloid scaffolds (Li et al., 2012; Weatheritt et al., 2014). Their ability to transiently bind multiple partners contributes to the formation of regulatory networks with fine-tuned responses to cellular signals (Stein et al., 2009). For example, the cyclin-dependent kinase (Cdk) inhibitor $p21^{Cip1}$ can, depending on cellular conditions, interact with several Cyclin-Cdk complexes, Rho kinase, or apoptosis signal-regulating kinase 1 (Asada et al., 1999; Kriwacki et al., 1996; Tanaka et al., 2002). The modest affinity with which IDPs bind their partners allows for quick signal termination, and their fast association rates allow for signals to be quickly turned on (Dyson and Wright, 2005; Oldfield et al., 2005; Pontius, 1993; Wright and Dyson, 1999).

With so many vital functions, it is unsurprising that mutations in IDPs or changes in their cellular levels result in disease states (Babu et al., 2011; Gsponer et al., 2008; Vavouri et al., 2009). In addition, misfolding of the IDPs α-synuclein and tau cause Parkinson's Disease and Alzheimer's Disease, respectively (Breydo et al., 2012; Mukrasch et al., 2009). Database analysis shows that proteins involved in eukaryotic signal transduction or with ties to cancer are enriched for IDPs (Iakoucheva et al., 2002).

Intrinsic disorder provides accessibility for PTM, which can result in dense modification clusters (Pejaver et al., 2014). In fact, phosphorylation sites are predominantly located in disordered regions (Iakoucheva et al., 2004). Intrinsic disorder also correlates with ubiquitination sites and therefore proteasomal degradation (Cox et al., 2002; Prakash et al., 2004). These modifications often participate in regulation of binding, allowing IDPs to act as biological switches. The proteasome may degrade some IDPs and cleave others to release activated globular domains (David et al., 2002; Liu et al., 2003; Sheaff et al., 2000).

1.5.3 Fuzziness

At the time when discussion of IDPs was in its infancy, it was suggested that an important feature of these proteins was that they underwent a transition to order prior to or during their biological function (Uversky, 2002b). However, in more recent years an increasing number of complexes have been identified in which disorder is preserved and functionally necessary, a phenomenon termed fuzziness (Fuxreiter et al., 2011; Tompa and Fuxreiter, 2008). Fuzzy complexes are essentially the multi-protein equivalent of IDPs, a challenge to the idea that disorder disappears following binding events. Therefore, the functions, regulation, and advantages of fuzziness are similar to those of IDPs. The interactions in these complexes are often mediated by short motifs, which confer only weak binding and few structural constraints, allowing for the retention of significant flexibility (Neduva and Russell, 2005). Disordered regions are important in the formation of these complexes by stabilizing secondary structure in the contacting region, establishing long-range electrostatic interactions, or making transient physical contacts (Phylip et al., 2001; Yu et al., 1994). When connecting two bound ordered domains, their continued disorder provides entropic benefit to the complex (Bhattacharyya et al., 2006; van Leeuwen et al., 1997).

Fuzzy complexes have been classified into four models for based on the type and location of the disordered segment (Figure 1-2, taken from Fuxreiter, 2011 Figure 1), which will be presented in order of increasing disorder (Fuxreiter, 2011; Tompa and Fuxreiter, 2008). The simplest case is the polymorphic model, in which the disordered region adopts two different conformations (Figure 1-2A). This model is the only one of the four classified as static disorder; the other three are all forms of dynamic disorder. The clamp model involves retained disorder in a linker separating two folded regions which serve as the eponymous clamps. The freedom of the clamps is limited by the

disordered linker, which makes no direct contacts with the binding partner. In the flanking model, the primary interaction with the binding partner is made through a short motif with an adjacent disordered segment. This segment retains its disorder in the complex, but serves to increase binding affinity. For instance, the KID of CREB binds KIX domain of CBP with a 29 residue helix-turn-helix, but removal of the remainder of KID, which is unstructured, reduces K_D five fold (Zor et al., 2002). Finally, the random model encompasses the most dynamic of these complexes, in which all or most of the complex remains disordered in the bound state.

Because the effects of fuzziness occur at a distance from the disordered regions, it has been proposed that fuzziness be classified as a form of dynamic allostery (Fuxreiter, 2011). Dynamic allostery involves modulation of the dynamics at the binding interface by flexible but structured regulatory sites (Popovych et al., 2006; Tzeng and Kalodimos, 2009). Extending this definition to include unstructured regulatory sites would encompass fuzzy complexes. Like IDPs, fuzzy complexes are susceptible to regulation by PTMs. Additional regulation may be effected by alternative splicing, thereby including, excluding, or altering the length of fuzzy regions in context- or tissue-dependent manners (Buljan et al., 2012; Sharma et al., 2015).

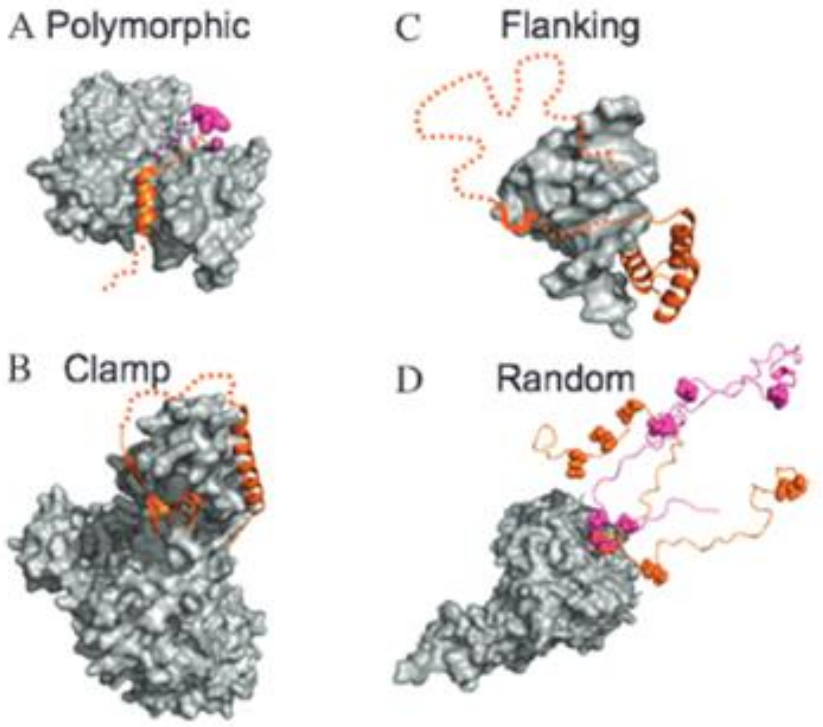


Figure 1-2. Structural categories of fuzzy complexes

Representative examples of structures of fuzzy complexes fitting each of the four structural categories. Taken from Figure 1, Fuxreiter, 2011.

(A) The polymorphic model, e.g. the WHD domain of Wiskott-Aldrich syndrome protein interacting with actin via an 18 residue segment (orange, PDB ID 2A3Z) or a 3 residue segment (purple, PDB ID 2FF3).

(B) The clamp model, e.g. the nonsense mediated decay factor UPF2 bound to UPF1 via two structured regions with an ambiguous linker (PDB ID 2WJV).

(C) The flanking model, e.g. Ultrabithorax homeodomain bound to the Extradenticle homeodomain, mediated by a short motif (bold) in a clamp-like region (PDB ID 1B8I).

(D) The random model, e.g. the Cdk inhibitor Sic1, whose nine phosphorylation sites interchange upon contact with Cdc4. Two of these, T45 and S76, are show in orange and purple, respectively. Phosphorylation sites are shown as spheres.

1.6 References for Chapter 1

Ahmad, K.F., Melnick, A., Lax, S., Bouchard, D., Liu, J., Kiang, C.-L., Mayer, S., Takahashi, S., Licht, J.D., and Privé, G.G. (2003). Mechanism of SMRT Corepressor Recruitment by the BCL6 BTB Domain. Mol. Cell *12*, 1551–1564.

Albagli, O., Dhordain, P., Bernardin, F., Quief, S., Kerkaert, J.P., and Leprince, D. (1996). Multiple domains participate in distance-independent LAZ3/BCL6-mediated transcriptional repression. Biochem. Biophys. Res. Commun. *220*, 911–915.

Antonescu, C.R., Sung, Y., Chen, C., Zhang, L., Chen, H., and Singer, S. (2014). Novel ZC3H7B-BCOR, MEAF6-PHF1, and EPC1-PHF1 Fusions in Ossifying Fibromyxoid Tumors — Molecular Characterization Shows Genetic Overlap with Endometrial Stromal Sarcoma. Cancer *193*, 183–193.

Asada, M., Yamada, T., Ichijo, H., Delia, D., Miyazono, K., Fukumuro, K., and Mizutani, S. (1999). Apoptosis inhibitory activity of cytoplasmic p21(Cip1/WAF1) in monocytic differentiation. EMBO J. *18*, 1223–1234.

El Ashkar, S., Schwaller, J., Pieters, T., Goossens, S., Demeulemeester, J., Christ, F., Van Belle, S., Juge, S., Boeckx, N., Engelman, A., et al. (2017). LEDGF/p75 is dispensible for hematopoiesis but essential for MLL-rearranged leukemogenesis. Blood.

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

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

Baek, S., Kutchukian, P., Verdine, G., Huber, R., Holak, T., Lee, K., and Popowicz, G. (2011). Structure of the stapled p53 peptide bound to Mdm2. J. Am. Chem. Soc. *134*, 103–106.

Bárdos, J.I., Saurin, A.J., Tissot, C., Duprez, E., and Freemont, P.S. (2000). HPC3 is a new human Polycomb orthologue that interacts and associates with RING1 and Bmi1 and has transcriptional repression properties. J. Biol. Chem. *275*, 28785–28792.

Béguelin, W., Teater, M., Gearhart, M.D., Calvo Fernández, M.T., Goldstein, R.L., Cárdenas, M.G., Hatzi, K., Rosen, M., Shen, H., Corcoran, C.M., et al. (2016). EZH2 and BCL6 Cooperate to Assemble CBX8-BCOR Complex to Repress Bivalent Promoters, Mediate Germinal Center Formation and Lymphomagenesis. Cancer Cell *30*, 197–213.

Bhattacharyya, R., Reményi, A., Good, M., Bashor, C., Falick, A., and Lim, W. (2006). The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. Science (80-.). *311*, 822–826.

Bienkiewicz, E., Adkins, J., and Lumb, K. (2002). Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27Kip1. Biochemistry *47*, 752–759.

Biondi, A., Cimino, G., Pieters, R., and Pui, C.-H. (2000). Biological and therapeutic

aspects of infant leukemia. Blood *96*, 24–33.

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

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

van den Boom, V., Maat, H., Geugien, M., Rodríguez López, A., Sotoca, A.M., Jaques, J., Brouwers-Vos, A.Z., Fusetti, F., Groen, R.W.J., Yuan, H., et al. (2016). Non-canonical PRC1.1 Targets Active Genes Independent of H3K27me3 and Is Essential for Leukemogenesis. Cell Rep. *14*, 332–346.

Borkin, D., He, S., Miao, H., Kempinska, K., Pollock, J., Chase, J., Purohit, T., Malik, B., Zhao, T., Wang, J., et al. (2015). Pharmacologic Inhibition of the Menin-MLL Interaction Blocks Progression of MLL Leukemia In Vivo Article Pharmacologic Inhibition of the Menin-MLL Interaction Blocks Progression of MLL Leukemia In Vivo. Cancer Cell *27*, 589–602.

Botbol, Y., Raghavendra, N.K., Rahman, S., Engelman, A., and Lavigne, M. (2008). Chromatinized templates reveal the requirement for the LEDGF/p75 PWWP domain during HIV-1 integration in vitro. Nucleic Acids Res. *36*, 1237–1246.

Bracken, C., Iakoucheva, L., Romero, P., and Dunker, A. (2004). Combining prediction, computation and experiment for the characterization of protein disorder. Curr. Opin. Struct. Biol. *14*, 570–576.

De Braekeleer, E., Meyer, C., Douet-Guilbert, N., Morel, F., Le Bris, M.-J., Berthou, C., Arnaud, B., Marschalek, R., Férec, C., and De Braekeleer, M. (2010). Complex and cryptic chromosomal rearrangements involving the MLL gene in acute leukemia: a study of 17 patients and review of the literature. Blood *44*, 268–274.

Breydo, L., Wu, J., and Uversky, V. (2012). α-Synuclein misfolding and Parkinson's disease. Biochim. Biophys. Acta *1822*, 261–285.

Buljan, M., Chalancon, G., Eustermann, S., Wagner, G., Fuxreiter, M., Bateman, A., and Babu, M. (2012). Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. Mol. Cell *46*, 871–883.

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

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

Cao, Q., Gearhart, M.D., Gery, S., Shojaee, S., Yang, H., Sun, H., Lin, D.-C., Bai, J.-W.,

Mead, M., Zhao, Z., et al. (2016). BCOR regulates myeloid cell proliferation and differentiation. Leukemia *30*, 1155–1165.

Chang, C., Ye, B.H., Chaganti, R.S.K., and Dalla-Favera, R. (1996). BCL-6, a POZ/zinc-finger protein, is a sequence-specific transcriptional repressor. PNAS *93*, 6947–6952.

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

Chen, J., Santillan, D.A., Koonce, M., Wei, W., Luo, R., Thirman, M.J., Zeleznik-Le, N.J., and Diaz, M.O. (2008). Loss of MLL PHD-finger-3 is necessary for MLL-ENL-induced hematopoietic stem cell immortalization. Cancer Res. *68*, 6199–6207.

Chen, Y., Anastassiadis, K., Kranz, A., Stewart, A.F., Arndt, K., Waskow, C., Yokoyama, A., Jones, K., Neff, T., Lee, Y., et al. (2017). MLL2, Not MLL1, Plays a Major Role in Sustaining MLL-Rearranged Acute Myeloid Leukemia. Cancer Cell *31*, 755–770.

Chou, D.M., Adamson, B., Dephoure, N.E., Tan, X., Nottke, A.C., Hurov, K.E., Gygi, S.P., Colaiácovo, M.P., and Elledge, S.J. (2010). A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. PNAS *107*, 18475–18480.

Chung, C.Y., Sun, Z., Mullokandov, G., Bosch, A., Qadeer, Z.A., Cihan, E., Rapp, Z.,

Parsons, R., Aguirre-Ghiso, J.A., Farias, E.F., et al. (2016). Cbx8 Acts Non-canonically with Wdr5 to Promote Mammary Tumorigenesis. Cell Rep. *16*, 472–486.

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

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

Collins, E.C., Appert, A., Ariza-McNaughton, L., Pannell, R., Yamada, Y., and Rabbitts, T.H. (2002). Mouse Af9 Is a Controller of Embryo Patterning, Like Mll, Whose Human Homologue Fuses with AF9 after Chromosomal Translocation in Leukemia. Mol. Cell. Biol. *22*, 7313–7324.

Cox, C., Dutta, K., Petri, E., Hwang, W., Lin, Y., Pascal, S., and Basavappa, R. (2002). The regions of securin and cyclin B proteins recognized by the ubiquitination machinery are natively unfolded. FEBS Lett. *527*, 303–308.

Creppe, C., Palau, A., Malinverni, R., Valero, V., and Buschbeck, M. (2014). A Cbx8-Containing Polycomb Complex Facilitates the Transition to Gene Activation during ES Cell Differentiation. PLoS Genet. *10*.

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

Damm, F., Chesnais, V., Nagata, Y., Yoshida, K., Scourzic, L., Okuno, Y., Itzykson, R., Sanada, M., Shiraishi, Y., Gelsi-Boyer, V., et al. (2013). BCOR and BCORL1 mutations in myelodysplastic syndromes and related disorders. Blood *122*, 3169–3177.

David, D., Layfield, R., Serpell, L., Narain, Y., Goedert, M., and Spillantini, M. (2002). Proteasomal degradation of tau protein. J. Neurochem. *83*, 301–310.

Dedmon, M., Patel, C., Young, G., and Pielak, G. (2002). FlgM gains structure in living cells. PNAS *99*, 12681–12684.

Deshpande, A.J., Deshpande, A., Sinha, A.U., Chen, L., Chang, J., Cihan, A., Fazio, M., Chen, C., Zhu, N., Koche, R., et al. (2014). AF10 Regulates Progressive H3K79 Methylation and HOX Gene Expression in Diverse AML Subtypes. Cancer Cell 896–908.

Deweindt, C., Albagli, O., Bernardin, F., Dhordain, P., Quief, S., Lantoine, D., Kerckaert, J.P., and Leprince, D. (1995). The LAZ3/BCL6 oncogene encodes a sequence-specific transcriptional inhibitor: a novel function for the BTB/POZ domain as an autonomous repressing domain. Cell Growth Differ. *6*, 1495–1503.

Dhordain, P., Albagli, O., Lin, R.J., Ansieau, S., Quief, S., Leutz, A., Kerckaert, J.-P.P.,

Evans, R.M., and Leprince, D. (1997). Corepressor SMRT binds the BTB/POZ repressing domain of the LAZ3/BCL6 oncoprotein. PNAS *94*, 10762–10767.

Dietrich, N., Bracken, A.P., Trinh, E., Schjerling, C.K., Koseki, H., Rappsilber, J., Helin, K., and Hansen, K.H. (2007). Bypass of senescence by the polycomb group protein CBX8 through direct binding to the INK4A-ARF locus. EMBO J. *26*, 1637–1648.

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

Dunker, A., Brown, C., Lawson, J., Iakoucheva, L., and Obradovic, Z. (2002). Intrinsic disorder and protein function. Biochemistry *41*, 6573–6582.

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

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

Dyson, H., and Wright, P. (2002a). Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. Adv Protein Chem *62*, 311–340.

Dyson, H., and Wright, P. (2002b). Coupling of folding and binding for unstructured proteins. Curr. Opin. Struct. Biol. *17*, 54–60.

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

Ellis, R. (2001). Macromolecular crowding: obvious but under- appreciated. Trends Biochem. Sci. *26*, 597–604.

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

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

Ernst, P., Wang, J., Huang, M., Goodman, R.H., and Korsmeyer, S.J. (2001). MLL and CREB Bind Cooperatively to the Nuclear Coactivator CREB-Binding Protein. Mol. Cell. Biol. *21*, 2249–2258.

Fair, K., Anderson, M., Bulanova, E., Mi, H., Tropschug, M., and Diaz, M.O. (2001). Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells. Mol. Cell. Biol. *21*, 3589–3597.

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., et al. (2012). KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands. Elife *2012*, 1–26.

Feng, Q., Wang, H., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002). Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Curr. Biol. *12*, 1052–1058.

Fuxreiter, M. (2011). Fuzziness: linking regulation to protein dynamics. Mol. Biosyst.

Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004). Preformed structural elements feature in partner recognition in intrinsically unstructured proteins. J. Mol. Biol. *338*, 1015–1026.

Fuxreiter, M., Simon, I., and Bondos, S. (2011). Dynamic protein–DNA recognition: beyond what can be seen. Trends Biochem. Sci. *36*, 415–423.

Galea, C., Wang, Y., Sivakolundu, S., and Kriwacki, R. (2008). Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits. Biochemistry *47*, 7598-609.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. Mol. Cell *45*, 344–356.

Gearhart, M.D., Corcoran, C.M., Wamstad, J.A., and Bardwell, V.J. (2006). Polycomb

group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. Mol. Cell. Biol. *26*, 6880–6889.

Ghetu, A.F., Corcoran, C.M., Cerchietti, L., Bardwell, V.J., Melnick, A., and Privé, G.G. (2008). Structure of a BCOR corepressor peptide in complex with the BCL6 BTB domain dimer. Mol. Cell *29*, 384–391.

Glanz, A., Forse, A., Polomeno, R.C., and Cole, D.E. (1983). Lenz microphthalmia: a malformation syndrome with variable expression of multiple congenital anomalies. Can. J. Ophthalmol. *18*, 41–44.

Gorlin, R.J., Marashi, a H., and Obwegeser, H.L. (1996). Oculo-facio-cardio-dental (OFCD) syndrome. Am. J. Med. Genet. *63*, 290–292.

Grossmann, V., Tiacci, E., Holmes, A.B., Kohlmann, A., Martelli, M.P., Kern, W., Spanhol-Rosseto, A., Klein, H.-U., Dugas, M., Schindela, S., et al. (2011). Whole-exome sequencing identifies mutations of BCOR in acute myeloid leukemia with normal karyotype. Blood *118*, 6153–6163.

Gsponer, J., Futschik, M., Teichmann, S., and Babu, M. (2008). Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science (80-.). *322*, 1365–1368.

Guenther, M.G., Jenner, R.G., Chevalier, B., Nakamura, T., Croce, C.M., Canaani, E., and Young, R. a (2005). Global and Hox-specific roles for the MLL1 methyltransferase.

PNAS *102*, 8603–8608.

Hatzi, K., Jiang, Y., Huang, C., Garrett-Bakelman, F., Gearhart, M.D., Giannopoulou, E.G., Zumbo, P., Kirouac, K., Bhaskara, S., Polo, J.M., et al. (2013). A hybrid mechanism of action for BCL6 in B-cells defined by formation of functionally distinct complexes at enhancers and promoters. Cell Rep. *4*, 578–588.

Hausrath, A.C., and Kingston, R.L. (2017). Conditionally disorderd proteins: bringing the environment back into the fold. Cell Mol. Life Sci. *74*, 3149–3162.

Hedera, P., and Gorski, J.L. (2003). Oculo-facio-cardio-dental syndrome: skewed X chromosome inactivation in mother and daughter suggest X-linked dominant Inheritance. Am. J. Med. Genet. A *123A*, 261–266.

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

Hsieh, J.J.-D., Cheng, E.H.-Y., and Korsmeyer, S.J. (2003a). Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. Cell *115*, 293–303.

Hsieh, J.J.-D., Ernst, P., Erdjument-bromage, H., Tempst, P., Korsmeyer, S.J., and Stanley, J. (2003b). Proteolytic Cleavage of MLL Generates a Complex of N- and C-Terminal Fragments That Confers Protein Stability and Subnuclear Localization. Mol.

Cell. Biol. *23*, 186–194.

Huang, H., Jiang, X., Li, Z., Li, Y., Song, C.-X., He, C., Sun, M., Chen, P., Gurbuxani, S., Wang, J., et al. (2013). TET1 plays an essential oncogenic role in MLL-rearranged leukemia. PNAS 1–6.

Huynh, K.D., and Bardwell, V.J. (1998). The BCL-6 POZ domain and other POZ domains interact with the co-repressors N-CoR and SMRT. Oncogene *17*, 2473–2484.

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

Iakoucheva, L., Brown, C., Lawson, J., Obradovic, Z., and Dunker, A. (2002). Intrinsic disorder in cell-signaling and cancer- associated proteins. J. Mol. Biol. *323*, 573–584.

Iakoucheva, L., Radivojac, P., Brown, C., O'Connor, T., Sikes, J., Obradovic, Z., and Dunker, A. (2004). The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. *32*, 1037–1049.

Iida, S., Seto, M., Yamamoto, K., Komatsu, H., Tojo, A., Asano, S., Kamada, N., Ariyoshi, Y., Takahashi, T., and Ueda, R. (1993). MLLT3 gene on 9p22 involved in t(9;11) leukemia encodes a serine/proline rich protein homologous to MLLT1 on 19p13. Oncogene *8*, 3085–3092.

Jones, B., Su, H., Bhat, A., Lei, H., Bajko, J., Hevi, S., Baltus, G.A., Kadam, S., Zhai, H., Valdez, R., et al. (2008). The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. PLoS Genet. *4*.

Junco, S.E., Wang, R., Gaipa, J.C., Taylor, A.B., Schirf, V., Gearhart, M.D., Bardwell, V.J., Demeler, B., Hart, P.J., and Kim, C. a (2013). Structure of the Polycomb Group Protein PCGF1 in Complex with BCOR Reveals Basis for Binding Selectivity of PCGF Homologs. Structure 1–7.

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

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

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

Kriwacki, R., Hengst, L., Tennant, L., Reed, S., and Wright, P. (1996). Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. PNAS *93*, 11504–11509.

Kuntimaddi, A., Achille, N.J., Thorpe, J., Lokken, A.A., Singh, R., Hemenway, C.S., Adli, M., Zeleznik-Le, N.J., and Bushweller, J.H. (2015). Degree of Recruitment of DOT1L to MLL-AF9 Defines Level of H3K79 Di- and Tri-methylation on Target Genes and Transformation Potential. Cell Rep. *11*, 808–820.

Landing, R., Russell, R., Neduva, V., and Gibson, T. (2003). GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res. *31*, 3701–3708.

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

van Leeuwen, H., Strating, M., Rensen, M., de Laat, W., and van der Vliet, P. (1997). Linker length and composition influence the flexibility of Oct-1 DNA binding. EMBO J. *16*, 2043–2053.

Li, J., McQuade, T., Siemer, A., Napetschnig, J., Moriwaki, K., Hsiao, Y., Damko, E., Moquin, D., Walz, T., McDermott, A., et al. (2012). The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis. Cell *150*, 339–350.

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

Li, Y., Sabari, B.R., Panchenko, T., Wen, H., Zhao, D., Guan, H., Wan, L., Huang, H., Tang, Z., Zhao, Y., et al. (2016). Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. Mol. Cell *62*, 181–193.

Liu, C., Corboy, M., DeMartino, G., and Thomas, P. (2003). Endoproteolytic activity of the proteasome. Science (80-.). *299*, 408–411.

Lokken, A.A., Achille, N.J., Chang, M., Lin, J.J., Kuntimaddi, A., Leach, B.I., Malik, B., Nesbit, J.B., Zhang, S., Bushweller, J.H., et al. (2014). Importance of a specific amino acid pairing for murine MLL leukemias driven by MLLT1/3 or AFF1/4. Leuk. Res. *38*, 1309–1315.

Maertens, G.N., El Messaoudi-Aubert, S., Racek, J.K., Nicholls, J., Rodriguez-Niedenführ, M., Gil, J., and Peters, G. (2009). Several distinct polycomb complexes regulate and co-localize on the INK4a tumor suppressor locus. PLoS One *4*.

Maethner, E., Garcia-Cuellar, M.-P., Breitinger, C., Takacova, S., Divoky, V., Hess, J.L., and Slany, R.K. (2013). MLL-ENL Inhibits Polycomb Repressive Complex 1 to Achieve Efficient Transformation of Hematopoietic Cells. Cell Rep. *1*, 1–14.

Marashi, A.H., and Gorlin, R.J. (1990). Radiculomegaly of canines and congenital cataracts--a syndrome? Oral Surg. Oral Med. Oral Pathol. *70*, 802–803.

Meyer, C., Hofmann, J., Burmeister, T., Gröger, D., Park, T.S., Emerenciano, M., Pombo de Oliveira, M., Renneville, A., Villarese, P., Macintyre, E., et al. (2017). The MLL recombinome of acute leukemias in 2017. Leukemia.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. (2002). MLL Targets SET Domain Methyltransferase Activity to Hox Gene Promoters.

Mol. Cell *10*, 1107–1117.

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

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

Mohan, M., Lin, C., Guest, E., and Shilatifard, A. (2010b). Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. Nat. Rev. Cancer *10*, 721–728.

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S.A., and Di Croce, L. (2012). Nonoverlapping functions of the polycomb group Cbx family of proteins in embryonic stem cells. Cell Stem Cell *10*, 47–62.

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

Muntean, A.G., and Hess, J.L. (2012). The Pathogenesis of Mixed-Lineage Leukemia. Annu. Rev. Pathol. *7*, 283–301.

Nakamura, T., Alder, H., Gu, Y., Prasad, R., Canaani, O., Kamada, N., Gale, R.P., Lange, B., Crist, W.M., Nowellii, P.C., et al. (1993). Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs. PNAS *90*, 4631–4635.

Neduva, V., and Russell, R. (2005). Linear motifs: evolutionary interaction switches. FEBS Lett. *579*, 3342–3345.

Ng, D. (2014). Lenz Microphthalmia Syndrome. In GeneReviews, (University of Washington, Seattle).

Ng, D., Thakker, N., Corcoran, C.M., Donnai, D., Perveen, R., Schneider, A., Hadley, D.W., Tifft, C., Zhang, L., Wilkie, A.O.M., et al. (2004). Oculofaciocardiodental and Lenz microphthalmia syndromes result from distinct classes of mutations in BCOR. Nat. Genet. *36*, 411–416.

O'Loghlen, A., Muñoz-Cabello, A.M., Gaspar-Maia, A., Wu, H.A., Banito, A., Kunowska, N., Racek, T., Pemberton, H.N., Beolchi, P., Lavial, F., et al. (2012). MicroRNA regulation of Cbx7 mediates a switch of polycomb orthologs during ESC differentiation. Cell Stem Cell *10*, 33–46.

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., and Dunker, A.K. (2003). Predicting intrinsic disorder from amino acid sequence. Proteins *53*, 566–572.

Okuda, H., Kawaguchi, M., Kanai, A., Matsui, H., Kawamura, T., Inaba, T., Kitabayashi,

I., and Yokoyama, A. (2014). MLL fusion proteins link transcriptional coactivators to previously active CpG-rich promoters. Nucleic Acids Res. *42*, 4241–4256.

Okuda, H., Kanai, A., Ito, S., Matsui, H., and Yokoyama, A. (2015). AF4 uses the SL1 components of RNAP1 machinery to initiate MLL fusion- and AEP-dependent transcription. Nat. Commun. *6*, 8869.

Okuda, H., Stanojevic, B., Kanai, A., Kawamura, T., Takahashi, S., Matsui, H., Takaori-Kondo, A., and Yokoyama, A. (2017). Cooperative gene activation by AF4 and DOT1L drives MLL-rearranged leukemia. J. Clin. Invest. *17*, 1–14.

Oldfield, C.J., Cheng, Y., Cortese, M.S., Romero, P., Uversky, V.N., and Dunker, A.K. (2005). Coupled folding and binding with α-helix-forming molecular recognition elements. Biochemistry *44*, 12454–12470.

Oza, J., Ganguly, B., Kulkarni, A., Ginjala, V., Yao, M., and Ganesan, S. (2016). A novel role of chromodomain protein CBX8 in DNA damage response. J. Biol. Chem. *291*, 22881–22893.

Panagopoulos, I., Thorsen, J., Gorunova, L., Haugom, L., and Bjerkehagen, B. (2013). Fusion of the ZC3H7B and BCOR Genes in Endometrial Stromal Sarcomas Carrying an X ; 22-Translocation. Genes. Chromosomes Cancer *52*, 610–618.

Park, S., Osmers, U., Raman, G., Schwantes, R.H., Diaz, M.O., and Bushweller, J.H. (2010). The PHD3 domain of MLL Acts as a CYP33-regulated switch between MLL-

mediated activation and repression. Biochemistry *49*, 6576–6586.

Pejaver, V., Hsu, W., Xin, F., Dunker, A., Uversky, V., and Radivojac, P. (2014). The structural and functional signatures of proteins that undergo multiple events of post-translational modification. Protein Sci. *23*, 1077–1093.

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

Phylip, L., Lees, W., Brownsey, B., Bur, D., Dunn, B., Winther, J., Gustchina, A., Li, M., Copeland, T., Wlodawer, A., et al. (2001). The potency and specificity of the interaction between the IA3 inhibitor and its target aspartic proteinase from Saccharomyces cerevisiae. J. Biol. Chem. *276*, 2023–2030.

Pierron, G., Tirode, F., Lucchesi, C., Reynaud, S., Ballet, S., Cohen-Gogo, S., Perrin, V., Coindre, J.-M., and Delattre, O. (2012). A new subtype of bone sarcoma defined by BCOR-CCNB3 gene fusion. Nat. Genet. *44*, 461–466.

Pina, C., May, G., Soneji, S., Hong, D., and Enver, T. (2008). MLLT3 regulates early human erythroid and megakaryocytic cell fate. Cell Stem Cell *2*, 264–273.

Pontius, B. (1993). Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. Trends Biochem. Sci. *18*, 181–186.

Popovych, N., Sun, S., Ebright, R., and Kalodimos, C. (2006). Dynamically driven protein allostery. Nat. Struct. Mol. Biol. *13*, 831–838.

Prakash, S., Tian, L., Ratliff, K., Lehotzky, R., and Matouschek, A. (2004). An unstructured initiation site is required for efficient proteasome-mediated degradation. Nat. Struct. Mol. Biol. *11*, 830–837.

Pugh, T.J., Weeraratne, S.D., Archer, T.C., Pomeranz Krummel, D. a., Auclair, D., Bochicchio, J., Carneiro, M.O., Carter, S.L., Cibulskis, K., Erlich, R.L., et al. (2012). Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature *488*, 4–8.

Radhakrishnan, I., Péerez-Alvarado, G.C., Parker, D., Dyson, H.J., Montminy, M.R., and Wright, P.E. (1997). Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. Cell *91*, 741–752.

Radhakrishnan, I., Pérez-Alvarado, G., Dyson, H., and Wright, P. (1998). Conformational preferences in the Ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. FEBS Lett. *430*, 317–322.

Richards, J., Bächinger, H., Goodman, R., and Brennan, R. (1996). Analysis of the structural properties of cAMP-responsive element-binding protein (CREB) and phosphorylated CREB. J. Biol. Chem. *271*, 13716–13723.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., and Dunker, A.K. (2001). Sequence complexity of disordered protein. Proteins Struct. Funct. Genet. *42*, 38–48.

Rowley, J.D. (1992). The der(11) chromosome contains the critical breakpoint junction in the 4;11, 9;11, and 11;19 translocations in acute leukemia. Genes. Chromosomes Cancer *5*, 264–266.

Sánchez, C., Sánchez, I., Demmers, J.A.A., Rodriguez, P., Strouboulis, J., and Vidal, M. (2007). Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. Mol. Cell. Proteomics *6*, 820–834.

Schulze, B.R., Horn, D., Kobelt, A., Tariverdian, G., and Stellzig, A. (1999). Rare dental abnormalities seen in oculo-facio-cardio-dental (OFCD) syndrome: three new cases and review of nine patients. Am. J. Med. Genet. *82*, 429–435.

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

Seyfert, V.L., Allman, D., He, Y., Staudt, L.M., and Seyfert, L. (1996). Transcriptional repression by the proto-oncogene BCL-6. Oncogene *12*, 2331–2342.

Sharma, R., Raduly, Z., Miskei, M., and Fuxreiter, M. (2015). Fuzzy complexes: Specific binding without complete folding. FEBS Lett. *589*, 2533–2542.

Sheaff, R., Singer, J., Swanger, J., Smitherman, M., Roberts, J., and Clurman, B. (2000). Proteasomal turnover of p21Cip1 does not require p21Cip1 ubiquitination. Mol. Cell *5*, 403–410.

Slany, R.K. (2009). The molecular biology of mixed lineage leukemia. Haematologica *94*, 984–993.

Sorenson, M., Ray, S., and Darst, S. (2004). Crystal structure of the flagellar sigma/anti-sigma complex sigma(28)/FlgM reveals an intact sigma factor in an inactive conformation. Mol. Cell *14*, 127–138.

Specht, K., Zhang, L., Sung, Y., Nucci, M., Dry, S., Vaiyapuri, S., Richter, G.H.S., Fletcher, C.D.M., and Antonescu, C.R. (2016). Novel BCOR-MAML3 and ZC3H7B-BCOR Gene Fusions in Undifferentiated Small Blue Round Cell Sarcomas. Am. J. Surg. Pathol. *40*, 433–442.

Spolar, R., and Record, M. (1994). Coupling of local folding to site-specific binding of proteins to DNA. Science (80-.). *263*, 777–784.

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

Steger, D.J., Lefterova, M.I., Ying, L., Stonestrom, A.J., Schupp, M., Zhuo, D., Vakoc,

A.L., Kim, J.-E., Chen, J., Lazar, M. a, et al. (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Mol. Cell. Biol. *28*, 2825–2839.

Stein, A., Pache, R., Bernadó, P., Pons, M., and Aloy, P. (2009). Dynamic interactions of proteins in complex networks: a more structured view. FEBS J. *276*, 5390–5405.

Tanaka, H., Yamashita, T., Asada, M., Mizutani, S., Yoshikawa, H., and Tohyama, M. (2002). Cytoplasmic p21Cip1/WAF1 regu- lates neurite remodeling by inhibiting Rho-kinase activity. J. Cell Biol. *158*, 321–329.

Tang, J., Wang, G., Zhang, M., Li, F., Sang, Y., Wang, B., Hu, K., Wu, Y., Luo, R., Liao, D., et al. (2014). Paradoxical role of CBX8 in proliferation and metastasis of colorectal cancer. Oncotarget *5*, 10778–10790.

Thiel, A.T., Blessington, P., Zou, T., Feather, D., Wu, X., Yan, J., Zhang, H., Liu, Z., Ernst, P., Koretzky, G.A., et al. (2010). MLL-AF9-Induced Leukemogenesis Requires Co-Expression of the Wild Type Mll Allele. Cancer Cell *17*, 148–159.

Tkachuk, D.C., Kohler, S., and Cleary, M.L. (1992). Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. Cell *71*, 691–700.

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

Tompa, P., and Csermely, P. (2004). Intrinsic disorder in cell-signaling and cancer-associated proteins. FASEB J. *18*, 1169–1175.

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

Tzeng, S., and Kalodimos, C. (2009). Dynamic activation of an allosteric regulatory protein. Nature *462*, 368–372.

Uversky, V. (2002a). What does it mean to be natively unfolded? Eur. J. Biochem. *269*, 2–12.

Uversky, V.N. (2002b). Natively unfolded proteins: A point where biology waits for physics. Protein Sci. *11*, 739–756.

Uversky, V., Gillespie, J., and Fink, A. (2000). Why are natively unfolded proteins unstructured under physiologic conditions? Proteins *41*, 415–427.

Vavouri, T., Semple, J., Garcia-Verdugo, R., and Lehner, B. (2009). Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. Cell *138*, 198–208.

Vucetic, S., Brown, C., Dunker, A., and Obradovic, Z. (2003). Flavors of protein disorder. Proteins *52*, 573–584.

Wamstad, J.A., and Bardwell, V.J. (2007). Characterization of Bcor expression in mouse development. Gene Expr. Patterns *7*, 550–557.

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

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. J. Mol. Biol. *337*, 635–645.

Weatheritt, R., Gibson, T., and Babu, M. (2014). Asymmetric mRNA localization contributes to fidelity and sensitivity of spatially localized systems. Nat. Struct. Mol. Biol. *21*, 833–839.

Wicky, B.I.M., Shammas, S.L., and Clarke, J. (2017). Affinity of iDPs to their targets is modulated by ion-specific changes in kinetics and residual structure. PNAS *114*, 9882–9887.

Wilkie, A.O., Taylor, D., Scambler, P.J., and Baraitser, M. (1993). Congenital cataract, microphthalmia and septal heart defect in two generations: a new syndrome? Clin. Dysmorphol. *2*, 114–119.

Wong, P., Iwasaki, M., Somervaille, T.C.P., So, C.W.E., So, C.W.E., and Cleary, M.L.

(2007). Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential. Genes Dev. *21*, 2762–2774.

Wright, P., and Dyson, H. (1999). Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. J. Mol. Biol. *293*, 321–331.

Wright, P.E., and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. Nat. Publ. Gr. *16*, 18–29.

Wright, R.L., and Vaughan, A.T.M. (2014). A systematic description of MLL fusion gene formation. Crit. Rev. Oncol. Hematol. 1–9.

Xiao, W., Ou, C., Qin, J., Xing, F., Sun, Y., Li, Z., and Qiu, J. (2014). CBX8, a novel DNA repair protein, promotes tumorigenesis in human esophageal carcinoma. Int. J. Clin. Exp. Pathol. *7*, 4817.

Yamamoto, Y., Tsuzuki, S., Tsuzuki, M., Handa, K., Inaguma, Y., and Emi, N. (2010). BCOR as a novel fusion partner of retinoic acid receptor alpha in a t(X;17)(p11;q12) variant of acute promyelocytic leukemia. Blood *116*, 4274–4283.

Yano, T., Nakamura, T., Blechman, J., Sorio, C., Dang, C. V, Geiger, B., and Canaani, E. (1997). Nuclear punctate distribution of ALL-1 is conferred by distinct elements at the N terminus of the protein. PNAS *94*, 7286–7291.

Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D.J., Kitabayashi, I., Herr,

W., and Cleary, M.L. (2004). Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. Mol. Cell. Biol. *24*, 5639–5649.

Yokoyama, A., Somervaille, T.C.P., Smith, K.S., Rozenblatt-Rosen, O., Meyerson, M., and Cleary, M.L. (2005). The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. Cell *123*, 207–218.

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

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

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

Yu, H., Chen, J., Feng, S., Dalgarno, D., Brauer, A., and Schreiber, S. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. Cell *76*, 933–945.

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

Zeisig, B.B., Arteaga, M.-F., Thirant, C., and So, C.W.E. (2011). Collaboration between PcG proteins and MLL fusions in Leukemogenesis: an emerging paradigm. Cancer Cell *20*, 551–553.

Zeleznik-Le, N.J., Harden, a M., and Rowley, J.D. (1994). 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. PNAS *91*, 10610–4.

Zhang, J., Benavente, C.A., McEvoy, J., Flores-Otero, J., Ding, L., Chen, X., Ulyanov, A., Wu, G., Wilson, M., Wang, J., et al. (2012). A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature *481*, 329–334.

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

Zhang, W., Xia, X., Reisenauer, M.R., Rieg, T., Lang, F., Kuhl, D., Vallon, V., and Kone, B.C. (2007). Aldosterone-induced Sgk1 relieves Dot1a-Af9 – mediated transcriptional repression of epithelial Na + channel α. J. Clin. Invest. *117*, 773–783.

Zhu, L., Li, Q., Wong, S.H.K., Huang, M., Klein, B.J., Shen, J., Ikenouye, L., Onishi, M., Schneidawind, D., Buechele, C., et al. (2016). ASH1L links histone H3 lysine 36 dimethylation to MLL Leukemia. Cancer Discov. *6*, 770–783.

Zor, T., Mayr, B., Dyson, H., Montminy, M., and Wright, P. (2002). Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. J. Biol. Chem. *277*, 42241–42248.

Chapter 2. Structural and biophysical characterization of the BCOR-AF9 complex

2.1 Introduction

BCOR is alternatively spliced, yielding four isoforms with variable ability to interact with AF9 (Srinivasan et al., 2003; Wamstad and Bardwell, 2007). Following this discovery, little further investigation into the role of BCOR in MLL leukemias occurred. No study has addressed the role of direct BCOR-AF9 interaction. Our goal was to identify the minimal AF9 interaction site of BCOR, solve the structure of this site in complex with AF9, and characterize the affinity of the interaction.

Coexpression of an initial complex provided spectra with good peak quality and dispersion, but without all backbone resonances assignable. A shorter BCOR peptide in complex with AF9 remedied the situation, and structure solution proceeded using this complex. Intriguingly, the two BCOR peptides described above bind AF9 with vastly differing affinities, modulated by the residues unassignable in the former construct.

2.2 Defining the BCOR binding motif for interaction with AF9

Alternative splicing leads to the expression of four BCOR isoforms in mice and humans (Srinivasan et al., 2003; Wamstad and Bardwell, 2007). When the BCOR-AF9 interaction was first discovered in a yeast two-hybrid assay it was noted that not all BCOR isoforms were capable of AF9 binding (Srinivasan et al., 2003). The two isoforms which were unable to bind AF9 made use of an alternative splice site at the 5’ end of exon 8 (Srinivasan et al., 2003; Wamstad and Bardwell, 2007). The portion of exon 8 located between the two splice acceptor sites codes for 34 amino acids that were therefore presumed to include residues required for AF9 binding. In fact, contained within this region is a 7 amino acid stretch of alternating hydrophobic residues, consistent with the AF9 binding motifs found in AF4 and DOT1L (Kuntimaddi et al., 2015; Leach et al., 2013) (Figure 2-1). To identify the minimally interacting region of

BCOR, we developed several BCOR constructs of varying lengths, all of which included the hydrophobic motif, and coexpressed them with the C-terminal 70 AF9 AHD residues previously used for structure determination of the AF4- and DOT1L-AF9 complexes (Kuntimaddi et al., 2015; Leach et al., 2013). ^{15}N HSQC spectra were acquired for each construct and evaluated for peak quality and dispersion. A complex consisting of BCOR residues 1176-1228 bound to AF9 ($BCOR_{long}$-AF9) produced good quality spectra, but upon assignment of backbone resonances it was discovered that roughly 20 amino acids near the C-terminus were not visible by NMR. However, a shorter BCOR peptide containing residues 1176-1207 in complex with AF9 ($BCOR_{short}$-AF9) was completely assignable, including residues not visible in the $BCOR_{long}$-AF9 spectra (Figure 2-2). As peak quality and dispersion were good and all residues were assignable, this construct was chosen for use in structure determination. Its poor solubility compared to its longer counterpart was rectified by the addition of a thioredoxin affinity tag preceding the His tag, both of which were removed by tobacco etch virus (TEV) protease digestion before collecting spectra.

```
AF4    LMVKITLDLL
DOT1L  LPVSIPLASV
BCoR   LKVCIELTGL
CBX8   LIARIPVARI
```

Figure 2-1. AF9 binding motif sequence

Sequence alignment of the AF9 binding motif in each of the four binding partners. Conserved hydrophobic residues are in red.

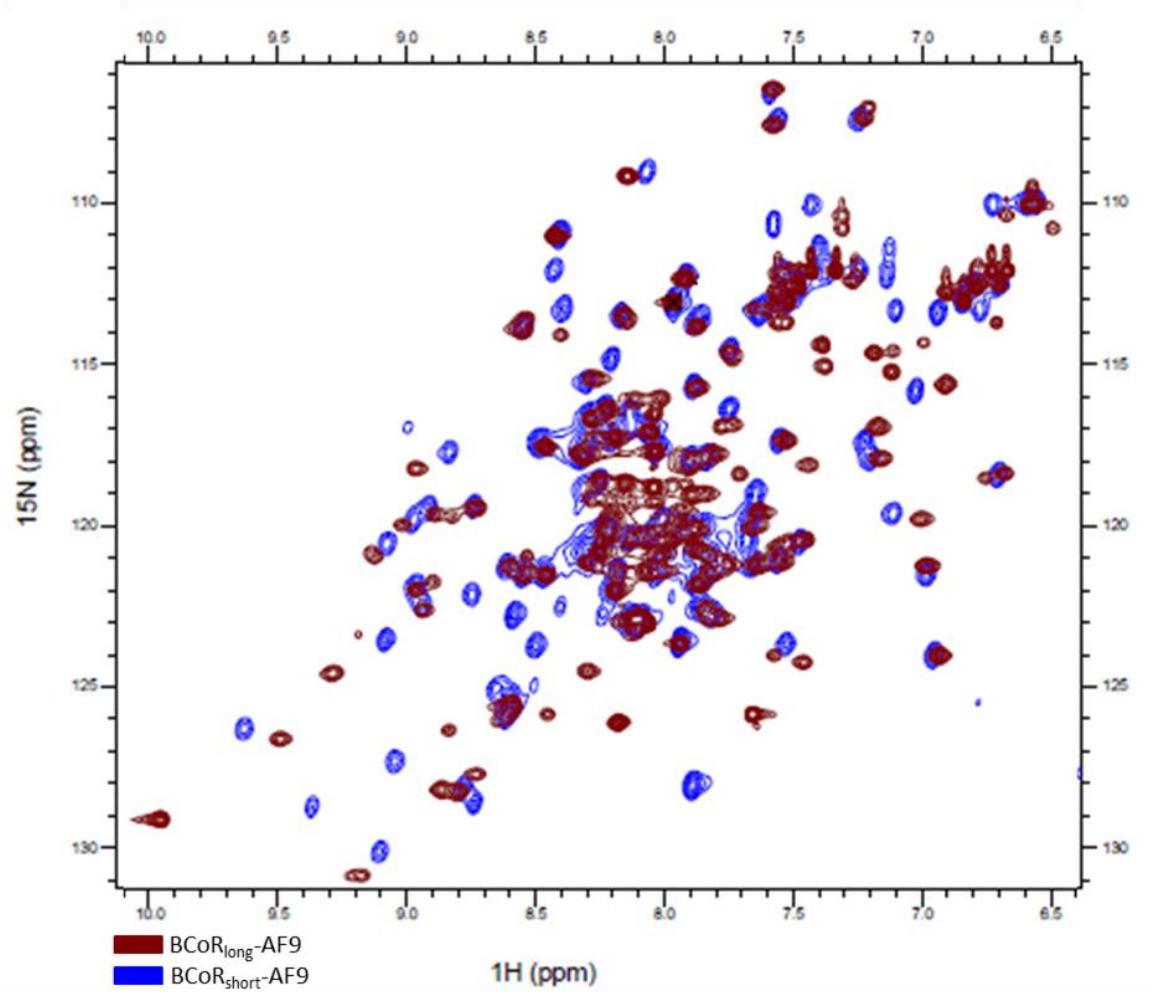


Figure 2-2. Overlay of BCOR$_{long}$- and BCOR$_{short}$-AF9 ^{15}N-^{1}H HSQC spectra.

An overlay of the ^{15}N-^{1}H HSQC spectra of the BCOR$_{long}$-AF9 (maroon) and BCOR$_{short}$-AF9 (blue) complexes shows peaks visible in the latter that are absent in the former.

2.3 Structure of the BCOR-AF9 AHD complex

The solution NMR structure of the $BCOR_{short}$-AF9 complex (PDB ID 6B7G) was solved using NOE and dihedral angle restraints without significant violations (Table 2-1). An ensemble of the 10 lowest energy structures shows a short interaction site involving a β strand of alternating hydrophobic and hydrophilic BCOR residues (Figure 2-3A). This strand creates an antiparallel β sheet with a β hairpin in AF9. The remaining AF9 residues form three helices which fold around the binding site, making additional contacts with the BCOR peptide and burying the hydrophobic side chains of the β sheet. This structure is similar to those of the other AF9 complexes, as demonstrated by aligning the $BCOR_{short}$- and DOT1L-AF9 complexes (Figure 2-3B). The backbone RMSDs for structured AF9 residues between the AF4-, DOT1L-, and CBX8-AF9 complexes and $BCOR_{short}$-AF9 are 1.9, 1.4, and 1.6 Å, respectively (Table 2-2). $\{^{15}N\}$-^{1}H heteronuclear NOE (hetNOE) measurements indicated some regions of significant flexibility, particularly at the BCOR N-terminus and AF9 C-terminus (Figure 2-4A). R_1 and R_2 data were recorded at 600 MHz to gain further insight into the dynamics of the $BCOR_{short}$-AF9 complex. Values of R_1R_2 for most of the structured residues of the AF9 and BCOR backbones were elevated above the expected value for the complex at this magnetic field strength in the absence of chemical exchange with S^2=0.87, the average of TALOS+ predicted values for structured residues in this complex (Figure 2-4B-C). This is indicative of a significant amount of conformational exchange and prevented the use of residual dipolar coupling (RDC) restraints for further structure refinement.

Table 2-1. Statistics for the NMR Solution structure of the $BCOR_{short}$-AF9 complex	
NMR Distance and Dihedral Constraints	Value
Distance Constraints	
Total NOE	1744
Intraresidue	1002
Interresidue	742
Sequential (|i-j| = 1)	350
Medium range (|i-j| <= 4)	175
Long range (|i-j| > 5)	217
Intermolecular	111
Total dihedral angle restraints	132
Φ	66
Ψ	66
Structure Statistics	
Violations (mean and SD)	
Distance constraints (Å)	0.059 ± 0.001
Dihedral angle constraints (°)	0.379 ± 0.042
Maximum dihedral angle violation (°)	7.167
Maximum distance constraint violation (Å)	0.805
Deviations from idealized geometry	
Bond lengths (Å)	0.003 ± 0.000
Bond angles (°)	0.467 ± 0.006
Impropers (°)	0.295 ± 0.013
Average pairwise rmsd (Å)	
Heavy	0.954
Backbone	0.347

Table 2-2. Backbone RMSDs for AF9 complexes compared to $BCOR_{short}$-AF9	
Complex	**RMSD (Å)**
AF4-AF9	1.6
CBX8-AF9	1.5
DOT1L-AF9	1.3

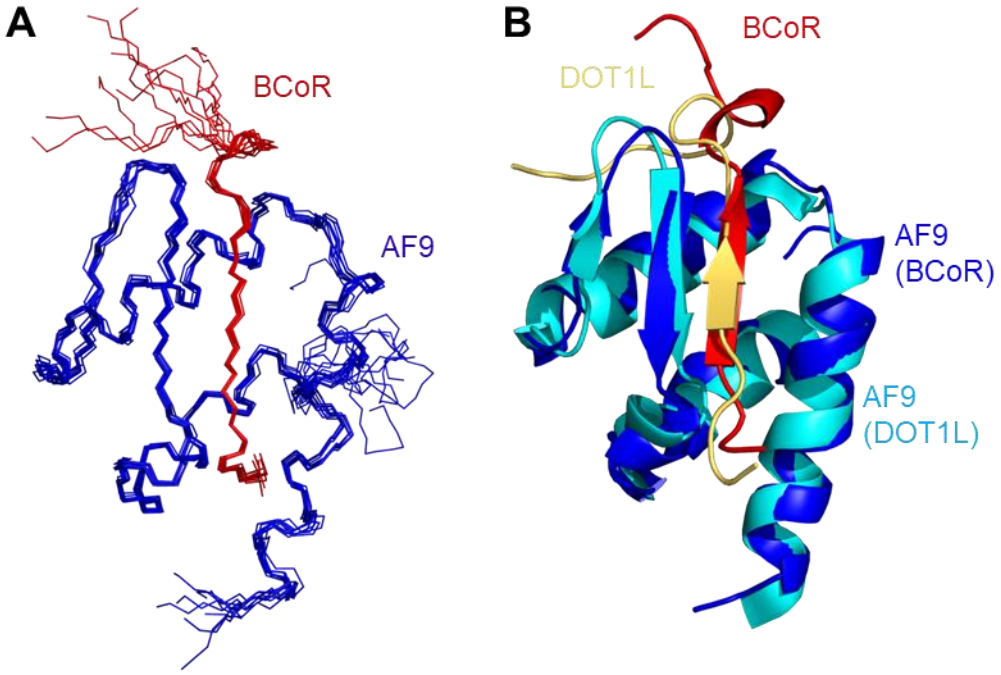


Figure 2-3. NMR solution structure of the $BCOR_{short}$-AF9 complex

(A) Ensemble of the ten lowest energy conformers $BCOR_{short}$-AF9 complex, with BCOR in red and AF9 in blue. Unstructured BCOR residues 1175-1191 have been omitted for clarity.

(B) Alignment of the BCOR-AF9 and DOT1L-AF9 structures. BCOR is shown in red and its AF9 in blue; DOT1L is shown in gold and its AF9 in cyan. Backbone RMSD for AF9 residues 502-562 between BCOR-AF9 and the other AF9 complexes is shown in Table 2-2. Unstructured BCOR residues 1175-1191 and DOT1L residues 893-900 have been omitted for clarity.

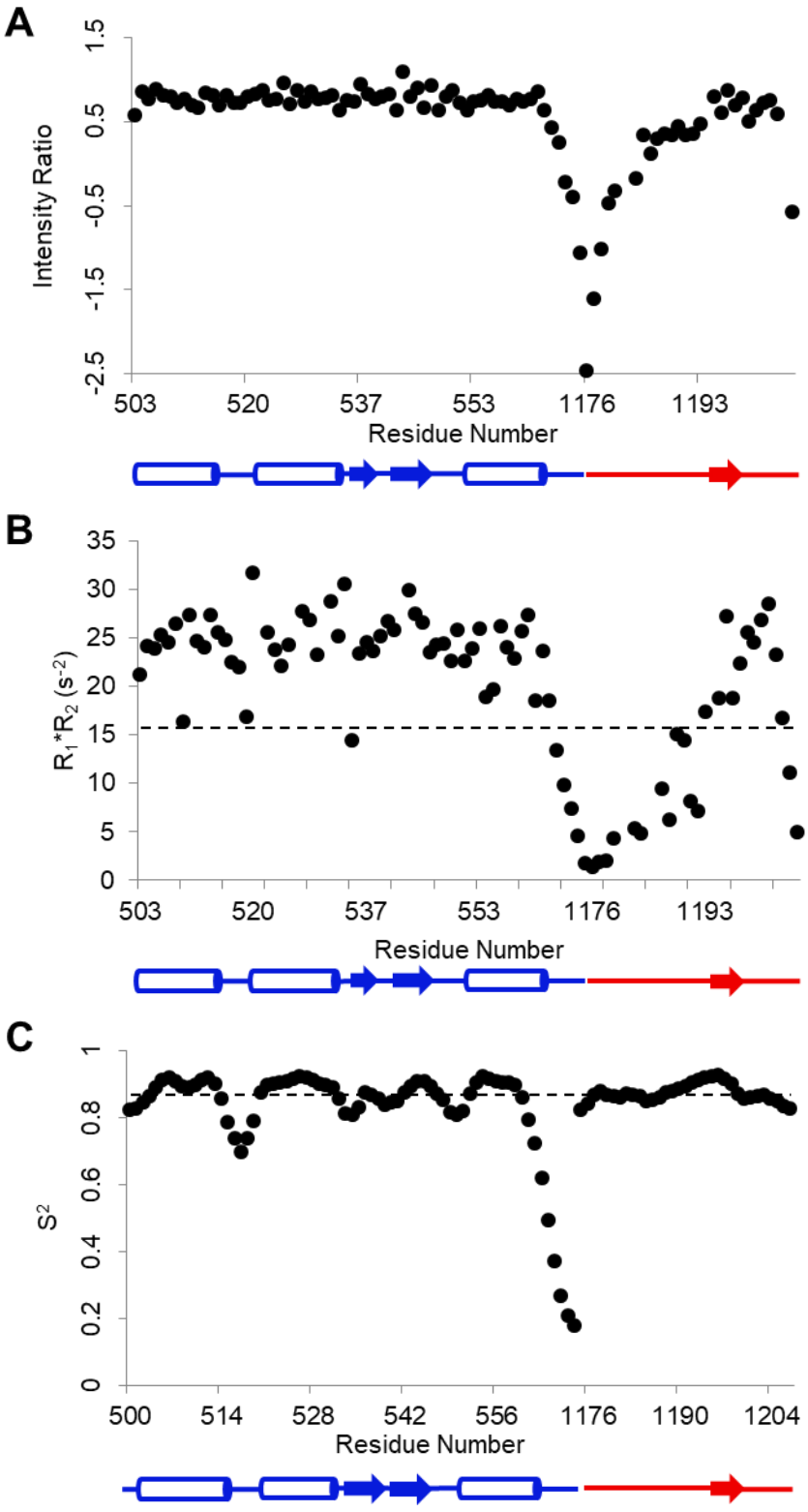


Figure 2-4. Backbone dynamics of the $BCOR_{short}$-AF9 complex

(A) {^{15}N}-^{1}H heteronuclear NOE intensity ratios for the BCOR-AF9 backbone. Secondary structure is shown below. Cylinders indicate alpha helices and arrows indicate beta strands, with AF9 in blue and BCOR in red.

(B) ^{15}N R_1*R_2 for the $BCOR_{short}$-AF9 backbone. Dashed line represents the expected value at 14 T in the absence of chemical exchange using S^2=0.87 (Kneller et al., 2002). Secondary structure is shown below. Secondary structure is shown below as in (A).

(C) S^2 values for the BCOR-AF9 backbone as predicted by TALOS+. Dashed line represents the average of structured residues. Secondary structure is shown below as in (A).

2.4 $BCOR_{short}$ and $BCOR_{long}$ differ in their affinity for AF9

As our lab has previously described, the intrinsically disordered nature of AF9 likely allows it to act as a hub able to exchange between its four binding partners in a context-dependent manner (Leach et al., 2013). In fluorescence anisotropy (FA) assays, a fluoresceinated BCOR peptide corresponding to that used in structure determination ($BCOR_{short}$-FITC) binds AF9 tagged with maltose binding protein (MBP-AF9) too weakly to be saturated (Figure 5A). As AF4 and DOT1L both bind AF9 tightly, it seemed unlikely that BCOR, with such a low affinity, would be able to compete for AF9 binding. However, a longer BCOR peptide extended by 19 residues at the C-terminus and corresponding to $BCOR_{long}$ ($BCOR_{long}$-FITC) binds much more tightly, with a K_D of 18 ± 2 nM (Figure 2-5).

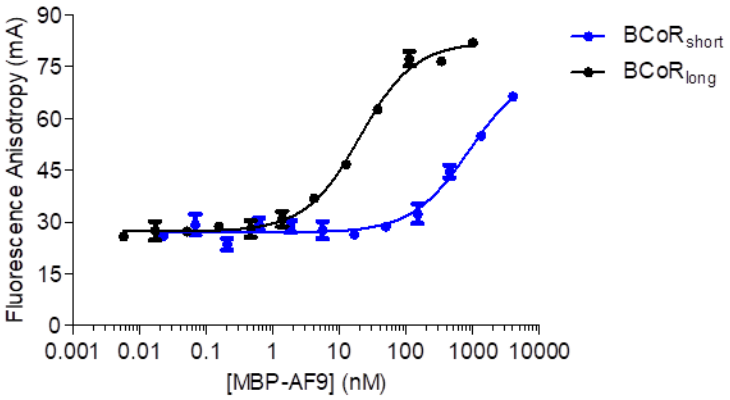


Figure 2-5. Characterization of $BCOR_{short}$- and $BCOR_{long}$-AF9 binding

Results of fluorescence anisotropy assays for binding of MBP-tagged AF9 to fluoresceinated $BCOR_{short}$ (blue) and $BCOR_{long}$ (black). Error bars indicate standard error of the mean (SEM) for three replicates.

2.5 Discussion

Though the requirement for the other AF9 binding partners in MLL-AF9 leukemia has been previously described, only recently has a study described the vital role BCOR plays in MLL-AF9 leukemia (van den Boom et al., 2016; Kuntimaddi et al., 2015; Leach et al., 2013; Okuda et al., 2017; Tan et al., 2011). However, the role of direct recruitment of BCOR by the MLL-AF9 fusion has not been characterized. We show that the BCOR-AF9 complex adopts a mixed alpha-beta structure highly similar to the AF4- and DOT1L-AF9 complexes we described previously (Kuntimaddi et al., 2015; Leach et al., 2013). This serves as further confirmation that the interactions of the AF9 AHD are mutually exclusive, as described for the AF4 and DOT1L complexes and suggested for the CBX8 and BCOR complexes (Biswas et al., 2011; Leach et al., 2013; Yokoyama et al., 2010). The BCOR-AF9 structure differs most from the previously published structures at the loop joining the two AF9 β strands (Figure 2-3B). This loop was also the most significant difference between the AF4- and DOT1L-AF9 structures, and is among the most dynamic regions in the complexes (Figure 2-4). It is notable that residues in this loop, particularly N540, T541, and F543, differ significantly in their chemical shift depending on binding partner, as we have previously described (Leach et al., 2013). BCOR is unique amongst the AF9 binding partners in two ways: 1) its ability to interact with AF9 is isoform dependent (Srinivasan et al., 2003), suggesting a potential mechanism for tissue-specific regulation of the BCOR-AF9 interaction and 2) BCOR affinity for AF9 is greatly increased by the addition of 19 residues C-terminal to the identified LXVXIXLXXL AF9-binding motif. This region unfortunately appears to be in intermediate exchange, and therefore not visible in the suite of experiments typically used to make backbone assignments. However, differences in the ^{15}N-^{1}H HSQC spectra for the two BCOR constructs (Figure 2-2) may make it possible to derive some indirect information about

the missing residues.

2.6 Experimental procedures

2.6.1 Protein expression and purification

BCOR residues 1176-1207 ($BCOR_{short}$) were coexpressed with AF9 residues 499-568 from pETDuet1 (EMD Millipore) modified to include a TEV cleavage site between the His tag and BCOR and a thioredoxin tag from pET 32a (Novagen) N-terminal to the His tag. The complex was expressed using Rosetta 2(DE3) cells (Novagen) in European Molecular Biology Laboratory medium (www.embl.de/pepcore/pepcore_services/protein_expression/ecoli/n15_c13) with the following modifications: 5 g/L ^{12}C glucose or 1 g/L ^{13}C glucose, 1 g/L $(NH_4)_2SO_4$, 8 g/L Na_2HPO_4, 2 g/L KH_2PO_4, 0.3 mM Na_2SO_4 rather than 1 mM $MgSO_4$, and supplemented with 5 mL/L Bioexpress (Cambridge Isotopes) rather than trace elements. The $BCOR_{long}$-AF9 complex was coexpressed similarly, but without the added thioredoxin tag, using BCOR residues 1176-1228.

The BCOR-AF9 complexes were purified using Ni-NTA agarose. The His and thioredoxin-His tags were removed by digestion with TEV protease. The complexes were passed through Ni-NTA resin to remove the tags and purified by size exclusion chromatography using a Superdex 75 column (GE Life Sciences).

2.6.2 $BCOR_{short}$-AF9 resonance assignments

All NMR experiments for BCOR-AF9 structure determination were conducted with a 750 µM sample of $BCOR_{short}$-AF9 in 25 mM Bis-Tris/MES, 100 mM NaCl, 1 mM DTT, pH 6.0 at 25° C using either a Varian Inova 600 MHz, Bruker 600 MHz, or Bruker 800 MHz magnet equipped with cryogenically cooled probes. Heteronuclear NOE experiments were conducted using a relaxation delay of 3 seconds. Residues identified as flexible based on decreased heteronuclear NOE intensity ratios were removed from

the constructs, with the exception of BCOR residues required for efficient TEV cleavage. Resonance assignments were determined from HNHA, HNCO, HNCACB, HNCA (CBX8-AF9 only), CBCA(CO)NH, CC(CO)NH-TOCSY, HCCH-TOCSY, 3D ^{15}N-edited NOESY (100 ms mixing time), 3D ^{13}C-edited aliphatic NOESY (100 ms mixing time), and 3D ^{13}C-edited aromatic NOESY (100 ms mixing time) spectra. Stereospecific assignments for the methyl groups of valine and leucine in both complexes were made using a 10% ^{13}C sample (Neri et al., 1989). All NMR data were processed using NMRPipe (Delaglio et al., 1995) and assigned using CcpNmr Analysis.

2.6.3 NMR relaxation experiments

T_1 and T_2 relaxation experiments were acquired using delays of 10, 180, 300, 500, 900, 1300, 1800, and 2300 ms for T_1 and 10, 30, 50, 90, 130, 170, and 230 ms for T_2 on the Bruker 600 MHz magnet. Data were fitted using CcpNmr Analysis (Vranken et al., 2005).

2.6.4 Structure calculation and refinement

Dihedral angle restraints were derived from TALOS+ based on N, C', C^{α}, C^{β}, H^{α}, and H^{N} chemical shifts with the range around the restrained angles set to no less than ±10° (Shen et al., 2009). CBX8-AF9 NOE assignments were made manually. BCOR-AF9 NOE assignments were made using CYANA (Güntert et al., 1997) and were refined in CcpNmr Analysis. Distance restraints were generated based on NOE cross-peak intensities and separated into four categories: 1.8-2.7, 1.8-3.3, 1.8-4.2, and 1.8-5.5 Å. Structure calculations were conducted using the XPLOR-NIH (Schwieters et al., 2003) simulated annealing protocol with dihedral angle and NOE-derived distance restraints, cooling in 25° C increments. The lowest energy structure was selected for refinement in XPLOR-NIH with the temperature increment reduced to 4° C. NMR structures were

aligned using MOLMOL (Koradi et al., 1996) and visualized using PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC).

2.6.5 Fluorescence anisotropy assays

For fluorescence anisotropy experiments, AF9 AHD residues 475-568 were expressed from pMAL-c2 (NEB) using Rosetta 2(DE3) cells (Novagen) in LB medium. MBP-AF9 was purified using amylose resin (NEB) and a Superdex 75 column (GE Life Sciences) to remove aggregates. Purified MBP-AF9 was titrated into a solution of 5 nM fluorescent peptide ($BCOR_{short}$-FITC or $BCOR_{long}$-FITC) in 50 mM Tris, 150 mM KCl, 1 mM DTT, pH 7.5. Samples were serially diluted and incubated for 1 hour at room temperature. Anisotropy was measured using a PHERAstar plate reader (BMG Labtech) and curves fitted to a single site binding model accounting for ligand depletion (Veiksina et al., 2010). Fluorescent constructs used were ordered from Biosyn and are as follows:

$BCOR_{short}$-FITC: TNSSSNHLEDPHYSELTNLKVCIELTGLHPK-Ahx-FITC

$BCOR_{long}$-FITC:

TNSSSNHLEDPHYSELTNLKVCIELTGLHPKKQRHLLHLRERWEQQVSAAK-Ahx-FITC

2.7 References for Chapter 2

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

van den Boom, V., Maat, H., Geugien, M., Rodríguez López, A., Sotoca, A.M., Jaques, J., Brouwers-Vos, A.Z., Fusetti, F., Groen, R.W.J., Yuan, H., et al. (2016). Non-canonical PRC1.1 Targets Active Genes Independent of H3K27me3 and Is Essential for Leukemogenesis. Cell Rep. *14*, 332–346.

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

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

Kneller, J.M., Lu, M., and Bracken, C. (2002). An effective method for the discrimination of motional anisotropy and chemical exchange. J. Am. Chem. Soc. *124*, 1852–1853.

Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph *14*, 29–32.

Kuntimaddi, A., Achille, N.J., Thorpe, J., Lokken, A.A., Singh, R., Hemenway, C.S., Adli, M., Zeleznik-Le, N.J., and Bushweller, J.H. (2015). Degree of Recruitment of DOT1L to MLL-AF9 Defines Level of H3K79 Di- and Tri-methylation on Target Genes and Transformation Potential. Cell Rep. *11*, 808–820.

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

Neri, D., Szyperski, T., Otting, G., Senn, H., and Wüthrich, K. (1989). Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional13C labeling. Biochemistry *28*, 7510–7516.

Okuda, H., Stanojevic, B., Kanai, A., Kawamura, T., Takahashi, S., Matsui, H., Takaori-Kondo, A., and Yokoyama, A. (2017). Cooperative gene activation by AF4 and DOT1L drives MLL-rearranged leukemia. J. Clin. Invest. *17*, 1–14.

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

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

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

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

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

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

Wamstad, J.A., and Bardwell, V.J. (2007). Characterization of Bcor expression in mouse development. Gene Expr. Patterns *7*, 550–557.

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

Chapter 3. BCOR-AF9 binding is required for MLL-AF9 leukemogenesis

3.1 Introduction

A mutation to selectively disrupt the BCOR-AF9 interaction would be a robust tool for studying the role this interaction plays in MLL-AF9 leukemia. Previous work in our lab has demonstrated that AF4, DOT1L, CBX8, and BCOR share a common binding site on AF9, which has been confirmed in Chapter 2 (Kuntimaddi et al., 2015; Leach et al., 2013). This makes identification of a mutant selective for BCOR very difficult, as mutations at or near the binding site are likely to affect the other partners, as well. BCOR, however, is distinct in the modulation of its binding by a region C-terminal to the AF9 binding motif. Targeting this region allowed for selective abrogation of the BCOR-AF9 interaction.

Though mutations affecting AF4 and DOT1L binding lead to decreases in colony forming ability (Kuntimaddi et al., 2015; Lokken et al., 2014), the AF9 E531R mutation led to a unique phenotype affecting a set of genes distinct from the other mutations. Loss of the BCOR-AF9 interaction also led to increased survival of mice with MLL-AF9 leukemia.

3.2 Structure-based E531R mutation specifically disrupts BCOR binding

As we have previously described, the intrinsically disordered nature of AF9 likely allows it to act as a hub able to exchange between its four binding partners in a context-dependent manner (Leach et al., 2013). In fluorescence anisotropy assays, $BCOR_{short}$-FITC binds MBP-AF9 too weakly to be saturated (Figure 2-5). As AF4 and DOT1L both bind AF9 tightly, it seemed unlikely that BCOR, with such a low affinity, would be able to compete for AF9 binding. However, a longer $BCOR_{long}$-FITC binds much more tightly, with a K_D of 18 ± 2 nM (Figure 2-5, Table 3-1).

To probe the functional role of the BCOR-AF9 interaction, we sought a mutation

in AF9 that would abrogate binding to BCOR without affecting interactions with the other binding partners. Finding a selective mutation is difficult because AF9 adopts the same structure with all four binding partners, each of which binds at the same site. Mutation near the interface thus risks affecting all of the AF9 interactions. However, the tighter binding of $BCOR_{long}$-FITC makes the C-terminal region of that construct and attractive target for selective disruption of the BCOR-AF9 interaction. To identify candidate AF9 residues, ^{15}N HSQC spectra of $BCOR_{short}$- and $BCOR_{long}$-AF9 were compared. The chemical shift perturbations of residues in $BCOR_{long}$-AF9 compared to $BCOR_{short}$-AF9 were quantified according to a published equation (Yuan et al., 2002) (Figure 3-1). The largest difference between the two occurred at residue E531, suggesting this as a site of interaction between the C-terminal extension of BCOR and AF9. Accordingly, this residue was targeted for mutation. Because the C-terminus of $BCOR_{long}$ contains several positively charged residues, we hypothesized that AF9 E531 may participate in an electrostatic interaction with one of them. Therefore, in order to effect the greatest reduction in binding affinity, we developed a charge reversal mutation, AF9 E531R, and assayed it as above for its effects on BCOR. The binding of $BCOR_{long}$ was greatly weakened, recapitulating the binding of $BCOR_{short}$ (Figure 3-2A). Importantly, the binding of AF4, DOT1L sites 2 and 3 (the two highest affinity DOT1L AF9 binding sites), and CBX8 A335V, a CBX8 mutant designed to bind tightly enough to measure in our assay, saw little to no change (Figure 3-2B-E).

Table 3-1. Relative affinities of AF9 binding partners for AF9 WT and E531R			
Construct	WT K_D (nM)	E531R K_D (nM)	Fold Change
$BCOR_{long}$	18 ± 2	> 3000	> 167
$BCOR_{short}$	> 4000	> 3000	N/A
AF4	0.10 ± 0.07	0.73 ± 0.23	7.3
DOT1L site 3	4 ± 1	14 ± 2	3.5
DOT1L site 2	33 ± 2	50 ± 3	1.5
CBX8 A335V	12 ± 1	6 ± 1	0.5

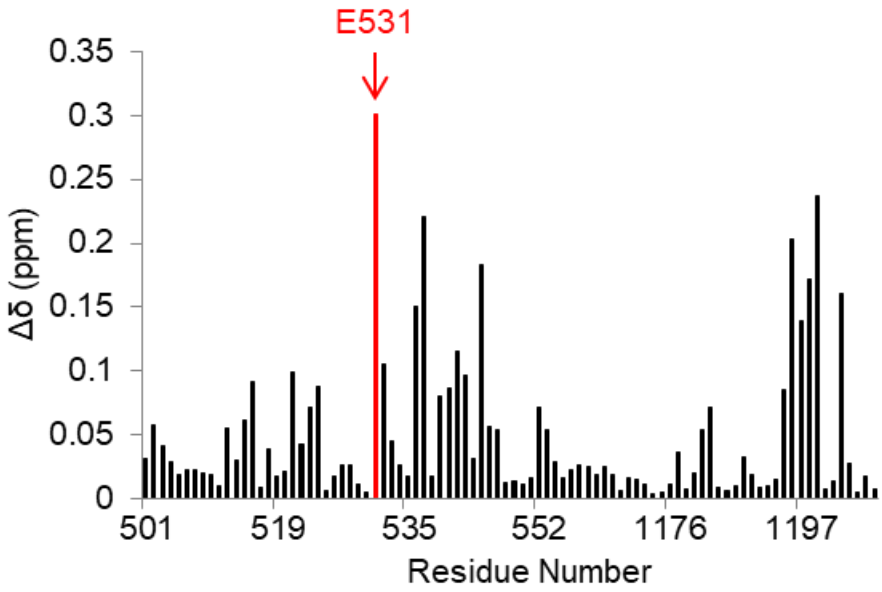
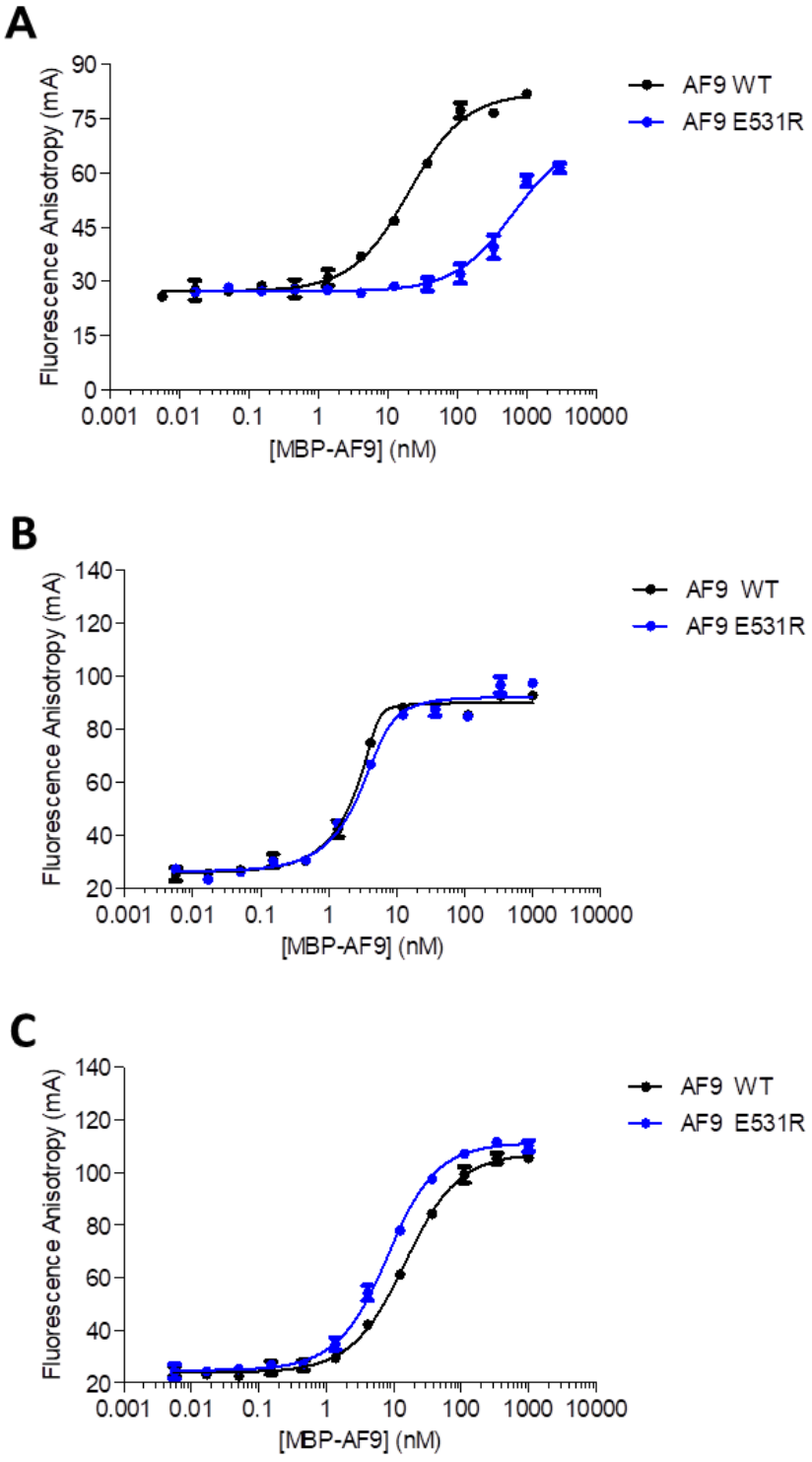


Figure 3-1. Chemical shift perturbations between $BCOR_{long}$- and $BCOR_{short}$-AF9 complexes

Weighted chemical shift differences for AF9 residues in $BCOR_{long}$-AF9 versus $BCOR_{short}$-AF9 complexes calculated using $|\Delta\delta^{15}N|/4.69 + |\Delta\delta^{1}H_N|$ (Yuan et al., 2002). AF9 E531 is shown in red.


A
Fluorescence Anisotropy (mA)
90
75
60
45
30
15
0
0.001 0.01 0.1 1 10 100 1000 10000
[MBP-AF9] (nM)
AF9 WT
AF9 E531R
B
Fluorescence Anisotropy (mA)
140
120
100
80
60
40
20
0.001 0.01 0.1 1 10 100 1000 10000
[MBP-AF9] (nM)
AF9 WT
AF9 E531R
C
Fluorescence Anisotropy (mA)
140
120
100
80
60
40
20
0.001 0.01 0.1 1 10 100 1000 10000
[MBP-AF9] (nM)
AF9 WT
AF9 E531R

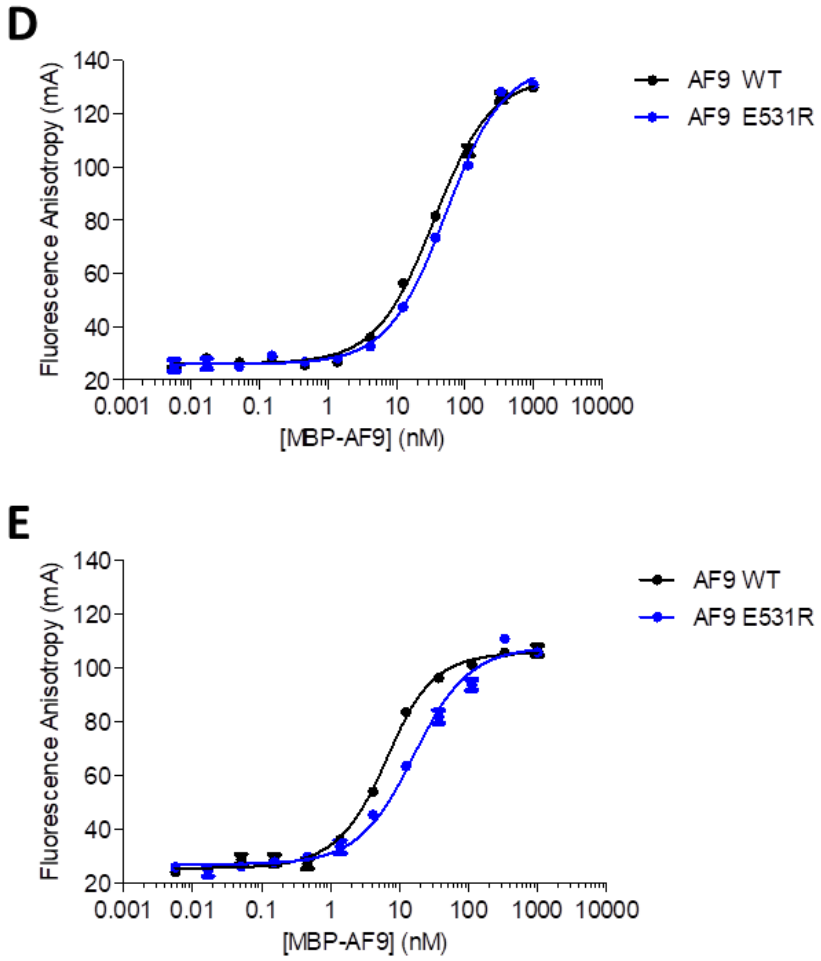


Figure 3-2. Characterization of MBP-AF9 E513R Binding

Results of fluorescence anisotropy assays for binding of MBP-tagged AF9 E531R to fluoresceinated peptides of $BCOR_{long}$ (A), AF4 (B), CBX8 A335V (C), DOT1L site 2 (D), and DOT1L site 3 (E).

3.3 MLL-AF9 E531R impacts proliferation, not colony formation[1]

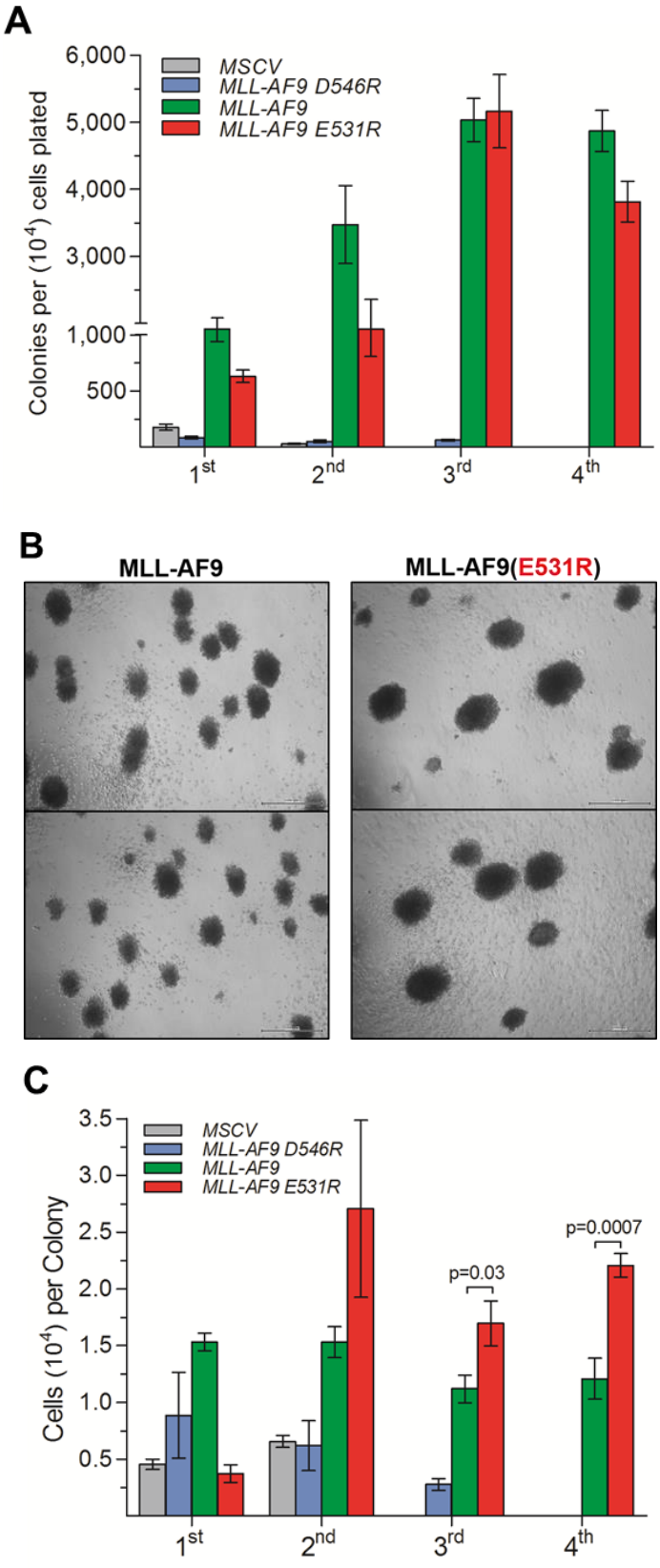
The robust reduction in BCOR affinity for AF9 E531R makes that mutation an excellent tool for probing the effects of this interaction in cells. To that end, WT MLL-AF9, MLL-AF9 E531R, and MLL-AF9 D546R (which abrogates DOT1L binding) (Kuntimaddi et al., 2015), or empty vector (MSCV) were transduced into mouse bone marrow (BM) stem/progenitor cells and were replated weekly in methylcellulose to measure colony formation ability. While the D546R mutant severely inhibited colony formation as we have previously reported (Kuntimaddi et al., 2015), the E531R mutant had no effect on the number of colonies (Figure 3-3A). However, the colonies were larger (Figure 3-3B) and had a greater number of cells per colony at weeks 2-4, with an increase in mean Feret diameter compared to WT colonies (Figure 3-3C-D).

3.4 The BCOR-AF9 interaction is required for MLL-AF9 *in vivo* leukemogenesis[2]

To evaluate whether the disruption of the BCOR-AF9 interaction will affect leukemogenesis *in vivo*, we transplanted retrovirally transformed BM cells into sublethally irradiated mice. Mice expressing WT MLL-AF9 developed AML as expected under these conditions (small number of cells transplanted, sublethal irradiation), whereas none of the MLL-AF9 E531R-expressing mice developed leukemia (Figure 3-4, Table 3-2).

[1] Colony formation data in this section were provided by Nicholas Achille and Dr. Nancy Zeleznik-Le, Oncology Research Institute, Loyola University Chicago, Maywood, IL. Dr. Zeleznik-Le also edited this section.

[2] Mouse data in this section were provided by Nicholas Achille, Shubin Zhang, and Dr. Nancy Zeleznik-Le, Oncology Research Institute, Loyola University Chicago, Maywood, IL. Dr. Zeleznik-Le also provided the section itself.


A
6,000
5,000
4,000
3,000
1,000
500
Colonies per (10^4) cells plated
MSCV
MLL-AF9 D546R
MLL-AF9
MLL-AF9 E531R
1st
2nd
3rd
4th
B
MLL-AF9
MLL-AF9(E531R)
C
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Cells (10^4) per Colony
MSCV
MLL-AF9 D546R
MLL-AF9
MLL-AF9 E531R
p=0.03
p=0.0007
1st
2nd
3rd
4th

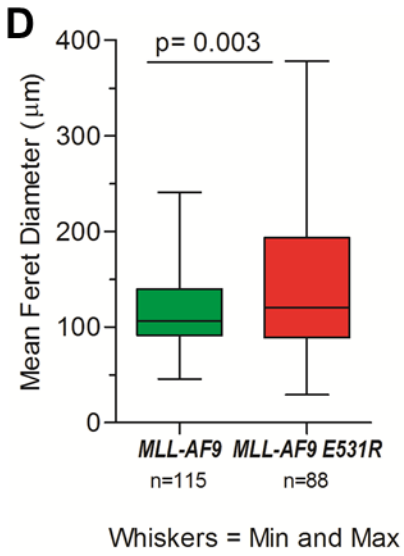


Figure 3-3. AF9 E531R Affects Cell Proliferation

Serial replating assays on methylcellulose for MSCV, MLL-AF9 D546R, MLL-AF9 WT, and MLL-AF9 E531R.

(A) Quantification of colony formation reported as colonies per 10^4 cells plated. Error bars show standard deviations (SDs) of replicate measurements.

(B) Bright-field images of MLL-AF9 WT and E531R colonies. Bars are equal to 500 µm.

(C) Quantification of colony formation reported as cells per colony. Error bars show SDs of replicate measurements.

(D) Feret diameter plots for MLL-AF9 WT and MLL-AF9 E531R cells. Whiskers are used for ranges as indicated.

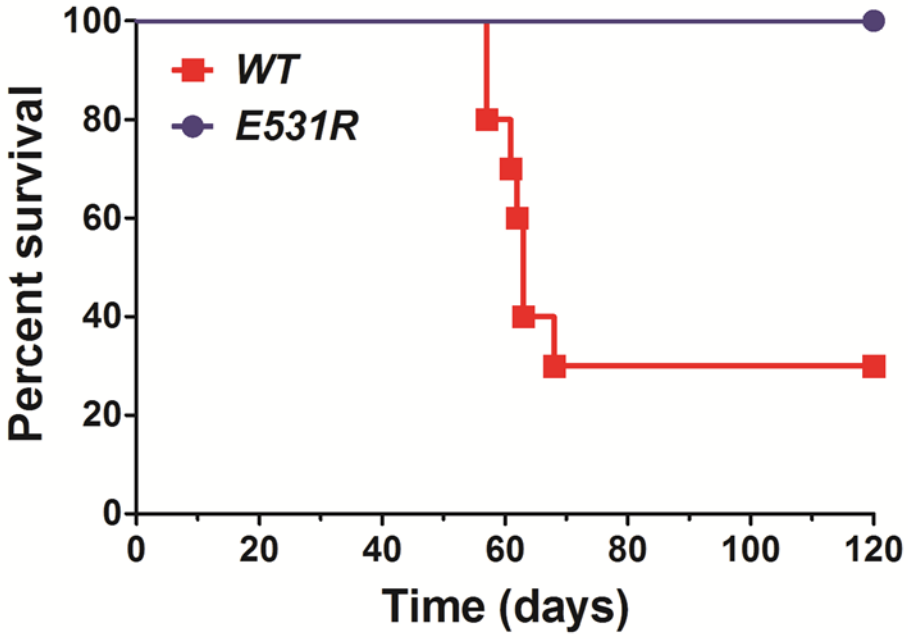


Figure 3-4. Effect of the MLL-AF9 E531R mutation on mouse survival

Survival curves for sublethally irradiated mouse recipients of BM cells retrovirally transformed with MLL-AF9 WT (red) or MLL-AF9 E531R (blue).

Table 3-2. Comparison of survival curves for mouse recipients of BM cells retrovirally transformed with MLL-AF9 WT or E531R	
Log-rank (Mantel Cox) Test	
Chi square	10.52
df	1
P value	0.0012
P value summary	**
Are the survival curves sig different?	Yes
Gehan-Breslow-Wilcoxon Test	
Chi square	9.970
df	1
P value	0.0016
P value summary	**
Are the survival curves sig different?	Yes
Median survival	
WT	63.00
E531R	Undefined
Hazard Ratio	
Ratio	0.07726
95% CI of ratio	0.01655 to 0.3631

3.5 The BCOR-AF9 interaction targets a unique set of genes[3]

In order to further investigate the importance of the BCOR-AF9 interaction with MLL-AF9, RNAseq was performed on mouse bone marrow cells transformed with either MLL-AF9 or MLL-AF9 E531R. Abrogation of BCOR-AF9 binding resulted in the misregulation of 1002 genes by greater than twofold with a false discovery rate (FDR) < 0.05 (Figure 3-5A-B). Of these, 516 were upregulated and 486 were downregulated.

Of 139 identified MLL-AF9 ChIP targets (Bernt et al., 2011), only a small number were affected by the E531R mutation (Figure 3-5C). Notably, only four were downregulated by greater than twofold, while none were upregulated at this level. We have shown previously that a subset of these MLL-AF9 targets exhibit a loss of H3K79 methylation upon loss of DOT1L-AF9 binding (Kuntimaddi et al., 2015). Genes affected by loss of BCOR-AF9 binding show very little overlap with this subset, with only 2 of 37 genes misregulated by greater than twofold. There is greater overlap, however, with PRC1.1 targets in MLL-AF9 cells (van den Boom et al., 2016). A total of 186 genes (18.6%) altered by the E531R mutation are located within 5 kb of PRC1.1 complex binding or binding of both canonical PRC1 and PRC1.1 (Figure 3-5D). Analysis of KEGG pathway-associated terms showed changes in genes involved in the hematopoietic cell lineage, but also showed significant change in cell adhesion molecules (CAMs) (Figure 3-5E). Gene ontology analysis indicated significant changes in genes associated with immune response and inflammation, in line with known BCL6 function (Dent et al., 1997). Additionally, affected genes are associated with positive regulation of ERK1 and ERK2 cascade, which regulate multiple cell processes, including proliferation, differentiation, and survival (Wortzel and Seger, 2011).

---

[3] Cells used for RNAseq in this section were transformed and grown by Nicholas Achille and Dr. Nancy Zeleznik-Le, Oncology Research Institute, Loyola University Chicago, Maywood, IL. Dr. Zeleznik-Le also provided the portion of this section dedicated to qPCR.

To directly compare the effect of specifically modulating binding of either BCOR, AF4, DOT1L, or CBX8 to MLL-AF9 via direct AF9 interaction, gene expression of select direct MLL-AF9 target genes was assessed by quantitative RT-PCR (qPCR) in bone marrow transduced with the different MLL-AF9 point mutants (E531R, D544R, or D546, respectively) or transduced with CBX8 A335V following *Cbx8* knockdown. Interestingly, target genes were differentially affected by the various mutants (Figure 3-6). For example, decreased binding of DOT1L (Figure 3-6C) showed significantly decreased expression of *Hoxa7*, *Hoxa9*, and *Hoxa10*, whereas abrogated BCOR binding (Figure 3-6A) showed increased *Hoxa7*, no effect on *Hoxa9*, and decreased *Hoxa10* expression. Higher binding affinity CBX8 (Figure 3-6D) significantly increased *Hoxa10* expression, but decreased *Meis1* and *Myb* expression, unaffected and decreased by the E531R mutation, respectively. In contrast, all mutations caused decreased expression of *Eya1*.

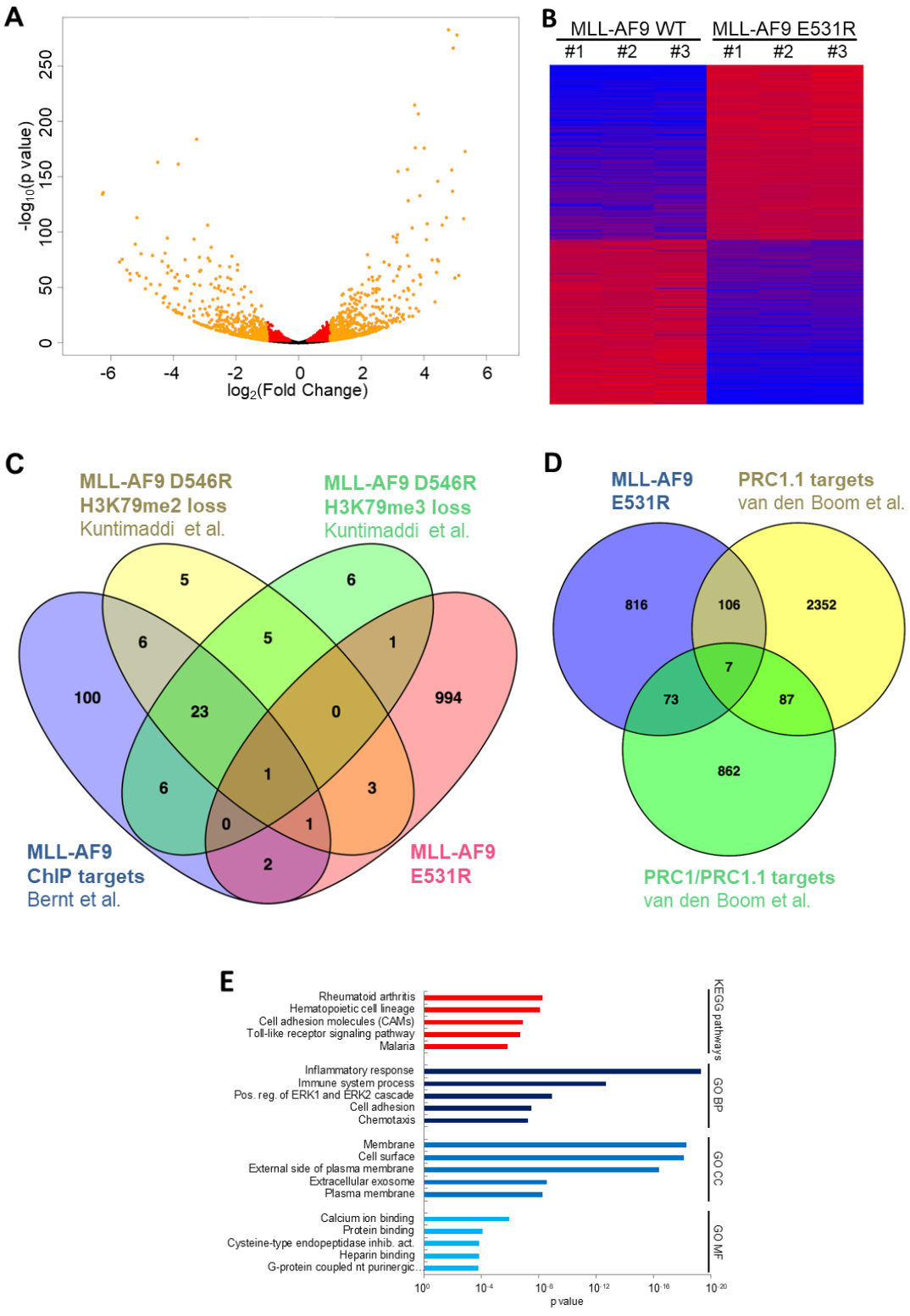


Figure 3-5. MLL-AF9 E531R RNAseq.

(A) Volcano plot showing differential expression in MLL-AF9 E531R samples compared

to wild type. Genes with FDR < 0.05 are show in red. Those with FDR < 0.05 and altered > two fold are shown in orange.

(B) Heat map of genes with greater than two-fold expression change between MLL-AF9 WT and E531R, with increase expression in red and decreased in blue.

(C) Venn diagram showing overlap of MLL-AF9 E531R altered genes (pink) with MLL-AF9 ChIP targets (blue) (Bernt et al., 2011), genes showing loss of H3K79me2 (yellow), and H3K79me3 (green) in MLL-AF9 D546R samples (Kuntimaddi et al., 2015).

(D) Venn diagram showing overlap of MLL-AF9 E531R misregulated genes (blue) with genes within 5 kb of PRC1.1 targets (yellow) or PRC1/PRC1.1 targets (green) (van den Boom et al., 2016).

(E) KEGG pathway and gene ontology (GO) analysis for genes altered by greater than two-fold in MLL-AF9 E531R samples. KEGG pathway results are in red, GO in blue, including biological process (BP, dark blue), cellular component (CC, medium blue), and molecular function (MF, light blue). The five most significant results from each category are shown.

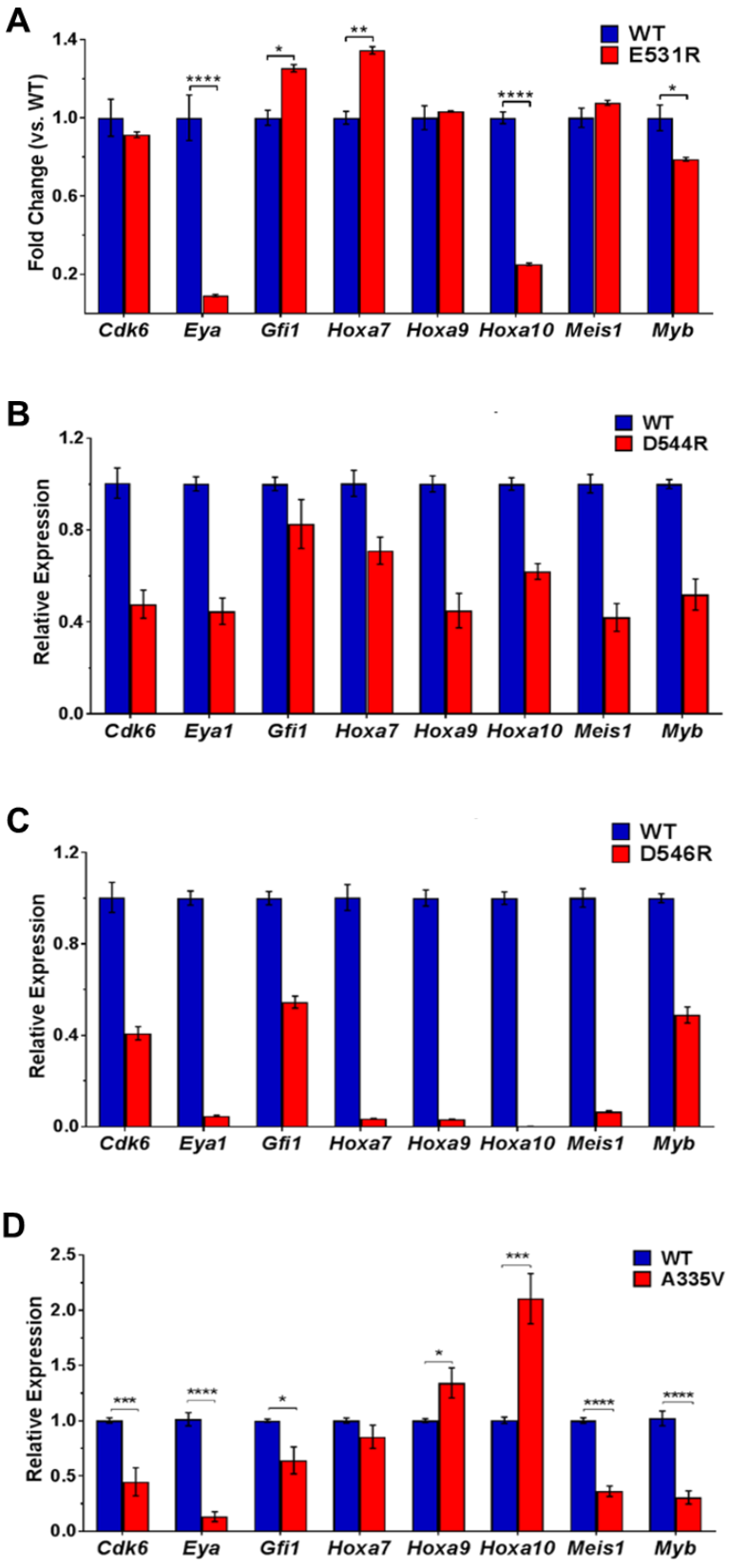

A
WT
E531R
Fold Change (vs. WT)
1.4
1.0
0.6
0.2
****
*
**
****
*
Cdk6
Eya
Gfi1
Hoxa7
Hoxa9
Hoxa10
Meis1
Myb
B
WT
D544R
Relative Expression
1.2
0.8
0.4
0.0
Cdk6
Eya1
Gfi1
Hoxa7
Hoxa9
Hoxa10
Meis1
Myb
C
WT
D546R
Relative Expression
1.2
0.8
0.4
0.0
Cdk6
Eya1
Gfi1
Hoxa7
Hoxa9
Hoxa10
Meis1
Myb
D
WT
A335V
Relative Expression
2.5
2.0
1.5
1.0
0.5
0.0
***
****
*
*
***
****
****
Cdk6
Eya
Gfi1
Hoxa7
Hoxa9
Hoxa10
Meis1
Myb

Figure 3-6. Quantitative PCR of genes affected by MLL-AF9 mutants

Relative expression measured by quantitative PCR for eight genes misregulated by MLL-AF9 mutations. Results are given in fold change relative to expression in MLL-AF9 WT-transformed cells for MLL-AF9 E531R (A), MLL-AF9 D544R (B), MLL-AF9 D546R (C), and MLL-AF9 WT with *Cbx8* knockdown and CBX8 A335V addback (D), with significance indicated by asterisks.

3.6 Discussion

Our structural studies made it possible to identify an AF9 point mutation which selectively disrupts BCOR binding, thus avoiding the challenges associated with effects of mutations on other binding partners (Kuntimaddi et al., 2015). The AF9 E531R mutant drastically reduces BCOR binding affinity while having little effect on AF4, DOT1L, and CBX8 (Table 3-1, Figure 3-2).

Previous work has indicated that AF9 mutations affecting AF4 (D544R) and DOT1L (D546R) both decrease the colony forming ability of MLL-AF9 transformed cells (Kuntimaddi et al., 2015; Lokken et al., 2014). This is concomitant with a loss of expression of *Hoxa9* and *Meis1*, genes which are both necessary to drive MLL leukemogenesis (Ayton and Cleary, 2003; Zeisig et al., 2004). Knowing that MLL leukemias require AF4 and DOT1L to activate and maintain transcription of MLL targets, it is not surprising that loss of their recruitment would result in loss of transforming ability and target expression. It would therefore be logical to expect that loss of the AF9 interaction with BCOR, whose canonical function is repressive, would have the opposite effect and that the E531R mutation would increase MLL-AF9 transformed cell proliferation. However, a recent study explored the role of BCOR in hematopoiesis by creating a mutant unable to interact with PRC1.1 and an overexpression system (Cao et al., 2016). Loss of the interaction with PRC1.1 resulted in an increase in proliferation and differentiation, a loss of the H2AK119ub mark at the promoters of *Hoxa* genes, and an increase in *Hoxa* gene expression. In addition, genes involved in myeloid cell differentiation were disproportionately upregulated by this loss-of-function mutation. *Bcor* overexpression had the opposite effect on proliferation and *Hoxa* expression. Other studies have observed that BCOR-containing PRC1.1 is recruited to target promoters by the CpG binding activity of KDM2B, and knockdown of KDM2B in human and mouse

ESCs induces differentiation (Farcas et al., 2012; He et al., 2013; Wu et al., 2013). Hematopoietic stem and progenitor cells (HSPCs) expressing a *Bcor* mutant unable to bind BCL6 saw increased proliferation rates under myeloid conditions when compared to WT, but showed no enhancement of colony formation (Tanaka et al., 2017). Consistent with this, our results demonstrate that loss of the BCOR-AF9 interaction leads to increased proliferation, without a change in colony forming ability (Figure 3-3). However, knockdown of *BCOR* in MLL-AF9 cells led to a loss of colony formation, indicating some function of BCOR (but not AF9 binding) is required for colony forming ability (van den Boom et al., 2016). In addition, loss of the BCOR-AF9 interaction dramatically improves survival in mice (Figure 3-4). Taken together, this is indicative that BCOR recruitment by MLL-AF9 is a vital step in downregulation of proliferative genes, likely via PRC1.1-mediated H2AK119 ubiquitination, to maintain the self-renewal capacity of hematopoietic progenitors. This would be consistent with the MLL-AF9 mutant that cannot recruit BCOR being unable to cause leukemia *in vivo* (Figure 3-4).

Loss of BCOR-AF9 binding affects the expression of 1002 genes greater than two fold (Figure 3-5A-B). The genes thus altered overlap only minimally with those that experience loss of H3K79me2/3 upon abrogation of DOT1L-AF9 binding, indicating that BCOR recruitment to AF9 targets a different subset of genes (Kuntimaddi et al., 2015) (Figure 3-5C). Moreover, qPCR confirms the RNAseq results and reaffirms that BCOR influences a subset of genes distinct from AF4 and DOT1L (Figure 3-6). The BCOR-dependent gene set overlaps more significantly with genes located near PRC1.1 and PRC1/PRC1.1 binding sites in MLL-AF9 cells, in agreement with its inclusion in the PRC1.1 complex (van den Boom et al., 2016) (Figure 3-5D). Of the 18.6% of altered genes in this subset, only one is a demonstrated MLL-AF9 target based on ChIP (Bernt et al., 2011), suggesting that the transcriptional effects exerted by the BCOR-AF9

interaction happen distally rather than at the point of recruitment. The genes affected by loss of this interaction are associated with several pathways and processes, including hematopoiesis, cell adhesion, and the ERK cascade (Figure 3-5E). Amongst the variety of cellular processes governed by the ERK cascade are differentiation and proliferation (Wortzel and Seger, 2011). This, combined with the negligible overlap with MLL-AF9 D546-affected genes, provides some explanation for the unique phenotype seen upon loss of BCOR binding when compared to loss of AF4 or DOT1L binding (Kuntimaddi et al., 2015; Lokken et al., 2014).

3.7 Experimental procedures

Sections **3.7.2**, **3.73**, and **3.75** were kindly provided by Dr. Nancy Zeleznik-Le, Oncology Research Institute, Loyola University Chicago, Maywood, IL

3.7.1 Fluorescence anisotropy assays

For fluorescence anisotropy experiments, AF9 AHD residues 475-568 were expressed from pMAL-c2 (NEB) using Rosetta 2(DE3) cells (Novagen) in LB medium. MBP-AF9 was purified using amylose resin (NEB) and a Superdex 75 column (GE Life Sciences) to remove aggregates. Purified MBP-AF9 was titrated into a solution of 5 nM fluorescent peptide ($BCOR_{short}$-FITC or $BCOR_{long}$-FITC, FITC-AF4, FITC-CBX8 A335V, FITC-DOT1L site 2, FITC-DOT1L site3) in 50 mM Tris, 150 mM KCl, 1 mM DTT, pH 7.5. Samples were serially diluted and incubated for 1 hour at room temperature. Anisotropy was measured using a PHERAstar plate reader (BMG Labtech) and curves fitted to a single site binding model accounting for ligand depletion (Veiksina et al., 2010). Fluorescent constructs used were ordered from Biosyn and are as follows:

$BCOR_{short}$-FITC: TNSSSNHLEDPHYSELTNLKVCIELTGLHPK-Ahx-FITC

$BCOR_{long}$-FITC:

TNSSSNHLEDPHYSELTNLKVCIELTGLHPKKQRHLLHLRERWEQQVSAAK-Ahx-FITC

FITC-AF4: FITC-Ahx-QSLMVKIDLDLLSRIPQPPGK

FITC-CBX8 A335V: FITC-Ahx-RPSLIVRIPVARILGDPEEE

FITC-DOT1L site 2: FITC-Ahx-TSLPISIPLSTVQPNK

FITC-DOT1L site 3: FITC-Ahx-NKLPVSIPLASVVLPSRAERARST

3.7.2 Serial replating assays

All studies involving mice were approved by Loyola University's Institutional Animal Care and Use Committee, according to standards set forth in the National Institutes of Health Guidelines. MLL-AF9 wild-type and MLL-AF9 mutant serial replating assays were conducted with murine bone marrow c-kit+ cells transduced with MSCVneo, MSCVneo-MLL-AF9 (WT), MSCVneo-MLL-AF9 (D546R), or MSCVneo-MLL-AF9 (E531R) retroviruses. Cells were plated in methylcellulose medium with cytokines and G418 as we have previously described (Cierpicki et al., 2010). Colonies and cells were enumerated and cells were serially replated after 7 days for each of four weeks. Colony size was quantified using the Fiji distribution of Image J analysis software (Schindelin et al., 2012, 2015). The arithmetic mean of the individual mean Feret diameters was then calculated for each experimental group. Statistical significance was determined using an unpaired *t-test* with a confidence interval of 95%.Colony assays were conducted in duplicate and repeated n=6 for all of the constructs.

3.7.3 *In vivo* transplantation

Transformed bone marrow cells (25,000) were transplanted into sublethally irradiated (450cGy, Radsource, RS-2000) mice vial tail vein injection. Mice were prophylactically treated with Baytril (Enrofloxacin, Sigma, #17849) at 0.1mg/mL in water *ad.lib*. for 14 days post irradiation. Mice were monitored for leukemia development by observing for lethargy, significant weight loss or ruffled fur. Leukemia was verified by peripheral blood

CBC analysis (Hemavet 950FS), and after sacrifice by examining the bone marrow, peripheral blood and spleen.

3.7.4 RNAseq

RNA was isolated from three MLL-AF9 WT and three MLL-AF9 E531R samples by HudsonAlpha Genome Services Lab using the miRNeasy mini kit (Qiagen) and the standard protocol with on-column DNase treatment. Isolated RNA was quantified by Qubit (Invitrogen) and underwent quality control with a 2100 Bioanalyzer (Agilent). Samples were normalized 500 ng of input in 50 µL and mRNAs were enriched by poly(A) selection. Resultant RNA was used for directional first-strand synthesis and cDNA library synthesis. Libraries were sequenced using a 100PE lane on a HiSeq 2500 (Illumina). Quality of data in FASTQ format was determined with FastQC (Babraham Bioinformatics). Adapter sequences were determined using BBMerge (Joint Genome Institute). Cutadapt (Martin, 2011) was used for adapter removal and quality trimming. Reads were mapped using HISAT2 (Kim et al., 2015) and sorted using SAMtools (Li et al., 2009). Read counts were obtained using featureCounts (Liao et al., 2014). Differential expression analysis was performed using DESeq2 (Love et al., 2014). Figures were generated using gplots (The Comprehensive R Archive Network).

3.7.5 Quantitative RT-PCR

Cells were harvested from methylcellulose colony assays after 4 weeks, RNA was isolated, and cDNA was synthesized. Quantitative real-time PCR was performed using TaqMan probes for *Cdk6*, *Eya1*, *Gfi1*, *Hoxa7*, and *Myb* (IDT, primer/probe sequences below), or for *Hoxa9*, *Hoxa10*, *Meis1* and *Myb* (Applied Biosystems), and data were analyzed using the $2^{-\Delta\Delta Ct}$ method. Expression was normalized to *Polr2a* expression (IDT, primer/probe sequences below) and was performed in duplicate on three independent samples.

Primer sequences:

Cdk6

Probe: /56-FAM/AAG ATG CAA /ZEN/CCG ACA CTC CAG AGG /3IABkFQ/

Primer 1: GAT CCA CGT CTG AAC TTC CAC

Primer 2: AAG TCC TGC TCC AGT CCA

Eya1

Probe: /56-FAM/ACA GAC CCC /ZEN/ACA GCA GAG TAC AGT A/3IABkFQ/

Primer 1: CTT CCC ATC TGA ACC TCG AC

Primer 2: AAT GCC ACT TAC CAA CTC CAG

Gfi1

Probe: /56-FAM/CAG AGA GCG /ZEN/GCA CAG TGA CTT CT/3IABkFQ/

Primer 1: CGA GTT CGA GGA CTT TTG GAG

Primer 2: TTG AAA GGC AGC GTG TAG G

Hoxa10

Probe: /56-FAM/AGT TGG CTG /ZEN/ CAT TTT CGC CTT TGG /3IABkFQ/

Primer 1: CGT CTG GTG CTT CGT GTA AG

Primer 2: GCA GTG ATT TCT GAA ATG AGT CA

Myb

Probe: /56-FAM/TCC GGT TGG /ZEN/GCA GAT AAT TGG CA/3IABkFQ

Primer 1: ATG AGT TCA GGG TTC AGC AC

Primer 2: TGG AAC AGA ACG GAA CAG AC

Polr2r

Probe: /5HEX/CCA CCA CCT /ZEN/CTT CCT CCT CTT GC/3IABkFQ/

Primer 1: CAG GGT CAT ATC TGT CAG CAT G

Primer 2: GGT CCT TCG AAT CCG CAT C

3.8 References for Chapter 3

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

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

van den Boom, V., Maat, H., Geugien, M., Rodríguez López, A., Sotoca, A.M., Jaques, J., Brouwers-Vos, A.Z., Fusetti, F., Groen, R.W.J., Yuan, H., et al. (2016). Non-canonical PRC1.1 Targets Active Genes Independent of H3K27me3 and Is Essential for Leukemogenesis. Cell Rep. *14*, 332–346.

Cao, Q., Gearhart, M.D., Gery, S., Shojaee, S., Yang, H., Sun, H., Lin, D.-C., Bai, J.-W., Mead, M., Zhao, Z., et al. (2016). BCOR regulates myeloid cell proliferation and differentiation. Leukemia *30*, 1155–1165.

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

Dent, A., Shaffer, A., Yu, X., Allman, D., and Staudt, L. (1997). Control of inflammation, cytokine expression, and germinal center formation by BCL-6. Science (80-.). *276*, 589–592.

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., et al. (2012). KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands. Elife *2012*, 1–26.

He, J., Shen, L., Wan, M., Taranova, O., Wu, H., and Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat. Cell Biol. *15*, 373–384.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kuntimaddi, A., Achille, N.J., Thorpe, J., Lokken, A.A., Singh, R., Hemenway, C.S., Adli, M., Zeleznik-Le, N.J., and Bushweller, J.H. (2015). Degree of Recruitment of DOT1L to MLL-AF9 Defines Level of H3K79 Di- and Tri-methylation on Target Genes and Transformation Potential. Cell Rep. *11*, 808–820.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lokken, A.A., Achille, N.J., Chang, M., Lin, J.J., Kuntimaddi, A., Leach, B.I., Malik, B., Nesbit, J.B., Zhang, S., Bushweller, J.H., et al. (2014). Importance of a specific amino acid pairing for murine MLL leukemias driven by MLLT1/3 or AFF1/4. Leuk. Res. *38*, 1309–1315.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10.

Schindelin, J., Arganda-Carreras, E., Frise E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676-82.

Schindelin, J., Rueden, M., Hiner, MC, and Elicieri, KW. (2015). The ImageJ ecosystem An open platform for biomedical image analysis. Mol. Reprod. Dev. *82*, 518-29.

Tanaka, T., Nakajima-takagi, Y., Aoyama, K., Tara, S., Oshima, M., Saraya, A., Koide, S., Si, S., Manabe, I., Sanada, M., et al. (2017). Internal deletion of BCOR reveals a tumor suppressor function for BCOR in T lymphocyte malignancies. 1–13.

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

Wortzel, I., and Seger, R. (2011). The ERK Cascade: Distinct Functions within Various Subcellular Organelles. Genes Cancer *2*, 195–209.

Wu, X., Johansen, J.V., and Helin, K. (2013). Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. Mol. Cell *49*, 1134–1146.

Yuan, X., Davydova, N., Conte, M.R., Curry, S., and Matthews, S. (2002). Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein. Nucleic Acids Res. *30*, 456–462.

Zeisig, B.B., Milne, T., Garcia-Cuellar, M.-P., Schreiner, S., Martin, M., Fuchs, U., Borkhardt, A., Chanda, S.K., Walker, J., Soden, R., et al. (2004). Hoxa9 and Meis1 Are Key Targets for MLL-ENL-Mediated Cellular Immortalization. Mol. Cell. Biol. *24*, 617–628.

Chapter 4. $BCOR_{long}$-AF9 is a fuzzy complex

4.1 Introduction

As discussed briefly in Chapter 2, BCOR residues 1208-1228 are important for high affinity AF9 binding, but unassignable in all NMR spectra collected. Though these residues have proven difficult to study directly, it is still possible to derive some information about their structure, however transient, indirectly. Predictions based on available chemical shifts and CD indicate the C-terminal regions of BCOR may adopt a helical conformation. In addition, mutations of residues in the putative helix to proline in an attempt to disrupt secondary structure have a distinct effect on NMR spectra. We present the case for classifying $BCOR_{long}$-AF9 as a flanking model of fuzzy complexes (Fuxreiter, 2011; Tompa and Fuxreiter, 2008), in which the BCOR C-terminal helix makes transient connections to AF9 to increase binding affinity, while resulting in the intermediate exchange observed by NMR.

4.2 The C-terminal region of $BCOR_{long}$ shows evidence of structure[4]

The text of Chapter 2 briefly discussed the absence of peaks corresponding to a number of residues in the $BCOR_{long}$ peptide. Specifically, all residues from 1176 to 1207 in the $BCOR_{short}$ peptide (with the exception of prolines) are represented in its ^{15}N-^{1}H HSQC, but only residues 1176-1200, 1203, and 1216-1228 are found in the $BCOR_{long}$-AF9 spectrum. In addition, peaks corresponding to several AF9 residues are significantly broadened in the context of $BCOR_{long}$ binding. These amino acids constitute the loop of the AF9 β hairpin that forms part of the binding interface. The broadening of NMR signal into the baseline is indicative of a type of chemical exchange known as intermediate exchange. Chemical exchange occurs when residues undergo conformational exchange between two states of differing chemical shift. When the rate of exchange between the

[4] The hetNOE experiments referenced in this section were performed by Aravinda Kuntimaddi.

two conformations is slow compared to the difference in their shifts, two distinct peaks are observed in the spectrum. When the rate of exchange is much faster than the difference in shifts, a single averaged peak is observed. However, when the rate of exchange is similar to the difference in shifts, an intermediate state occurs in which the peak becomes significantly broadened and disappears into the baseline.

Despite the lack of direct information on BCOR residues 1204-1215, it is still possible to learn something about residues 1216-1228. HetNOE measurements for the $BCOR_{long}$-AF9 backbone indicate that the residues on either side of the region in exchange are structured (Figure 4-1). These experiments, however, do not indicate the type(s) of secondary structure that may be present.

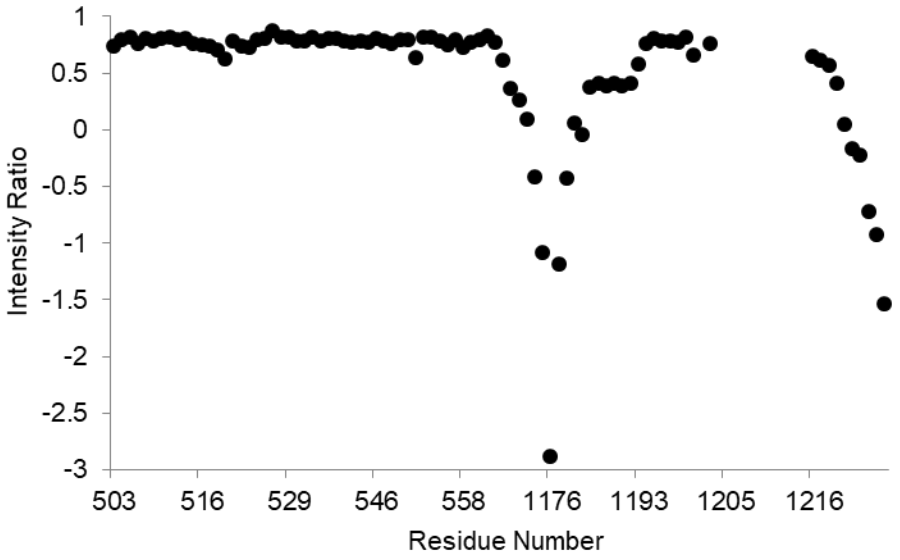


Figure 4-1. Backbone dynamics of the $BCOR_{long}$-AF9 complex

$\{^{15}N\}$-^{1}H heteronuclear NOE intensity ratios for the BCOR-AF9 backbone. The gap represents unassignable residues 1204-1215.

4.3 The C-terminal region of $BCOR_{long}$ has helical character

In order to gain insight into possible secondary structural elements of the BCOR C-terminus we first used Jpred, a web server that predicts secondary structure from primary sequence using the Jnet neural network (Cuff and Barton, 1999, 2000). The result of Jpred predictions for BCOR correctly identifies the β strand at the AF9 binding motif (Figure 4-2). Additionally, it predicts an α helix from residues 1208-1218.

The available chemical shifts for the $BCOR_{long}$ backbone resonances were used to generate TALOS+ secondary structure predictions (Shen et al., 2009) (Figure 4-3). Predictions for residues included in the $BCOR_{short}$-AF9 construct matched well to their actual secondary structure as presented in the NMR solution structure in Chapter 2. Residues 1215-1219 were predicted to be helical, with the helical probability trailing off and being overtaken by probability of random coil in subsequent residues, in line with hetNOE data.

In order to obtain secondary structural information pertinent to the C-terminal BCOR residues, we performed circular dichroism (CD) experiments. Collecting CD spectra of both BCOR-AF9 complexes at the same concentration and subtracting them yields a spectrum representative of the secondary structural elements present in BCOR residues 1208-1228 (Figure 4-4A). Although the distances between the measured spectra of $BCOR_{short}$- and $BCOR_{long}$-AF9 and the nearest matches in the K2D3 database are large, analysis of both the spectra indicates a mixture of α helical and β strand character, with a significant emphasis on the former (Figure 4-4B-C). Allowing for the large error, this is a match for the structure of $BCOR_{short}$-AF9 presented in Chapter 2. Though still eliciting a warning in the analysis, the difference spectrum is visually a much better match to the database (Figure 4-4D). Like the full constructs, predicted β strand character is very low, at 0.34%. The α helical character is predicted to account for

73.23% of the spectrum, less than the full constructs. This would represent roughly 15 of the 21 residues that account for the difference in the BCOR peptide. Taken with the TALOS+-predicted secondary structure, this would suggest an α helix from residue 1205-1219. As residue 1205 is a proline, it is unlikely that it would begin a helix, though the successive lysines at positions 1206-1207 may be able to do so.

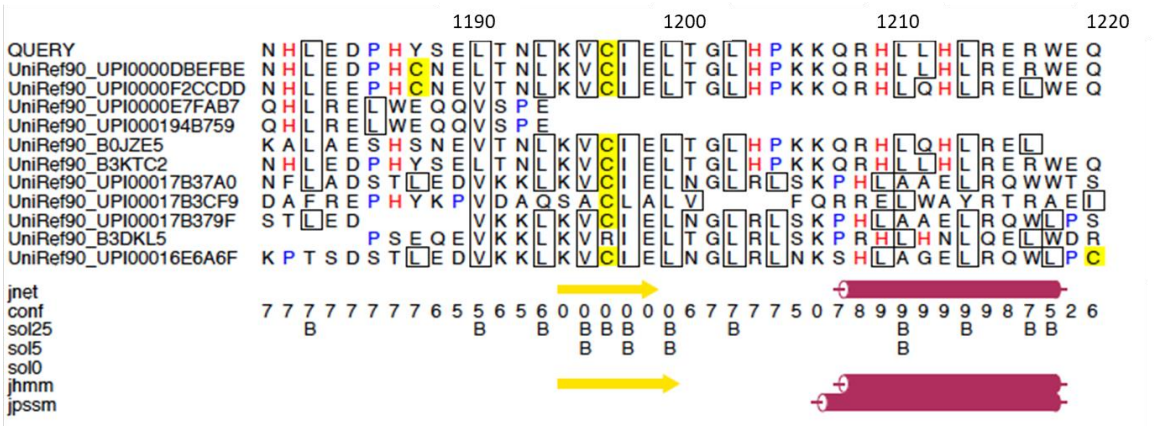


Figure 4-2. $BCOR_{long}$ secondary structure prediction

JPred secondary structure prediction for BCOR residues 1181-1220. Yellow arrows indicate a β strand prediction, while maroon cylinders indicate an α helix prediction.

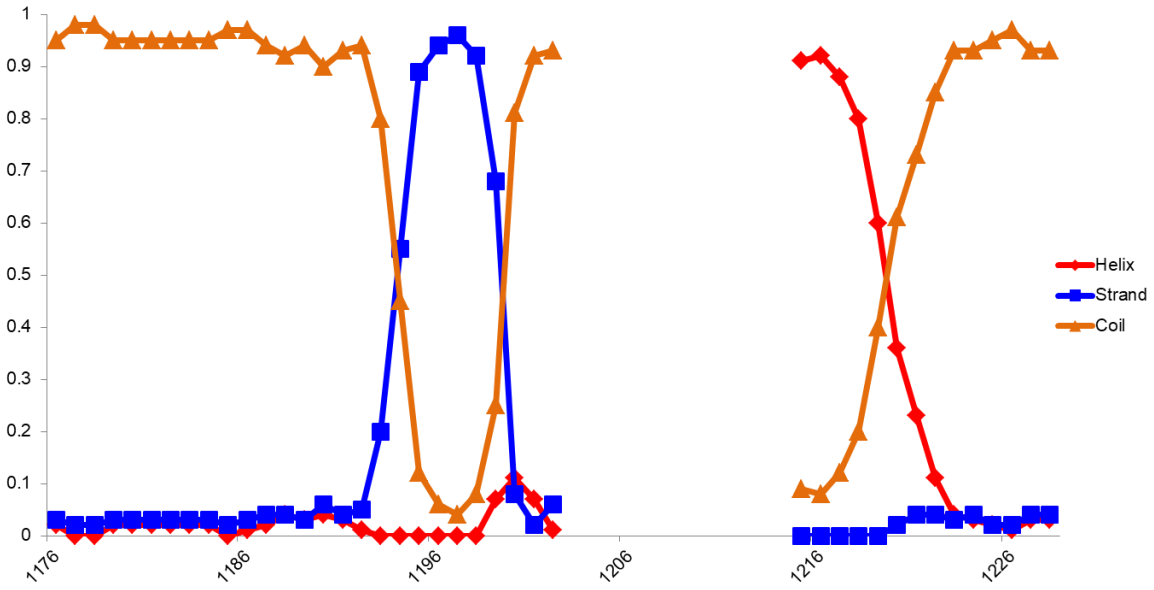
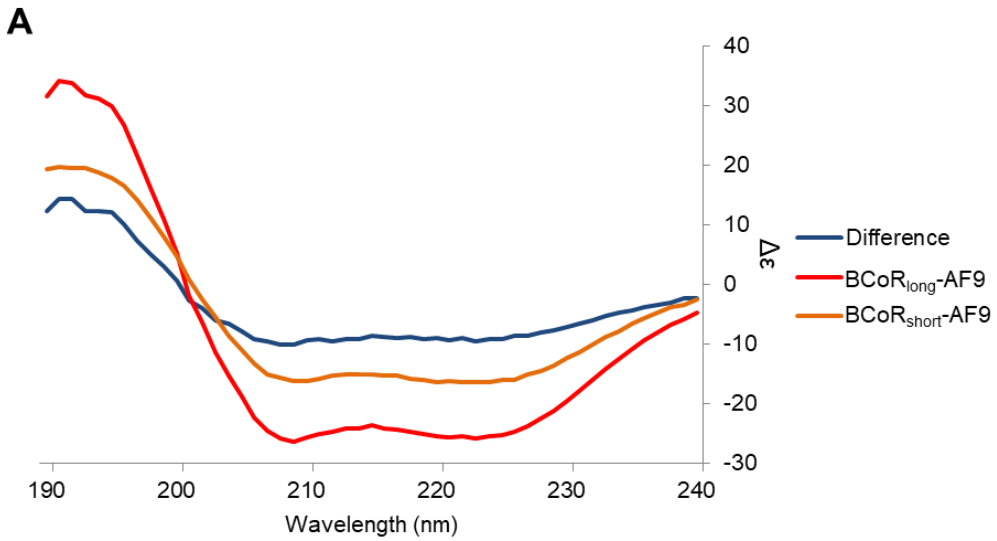
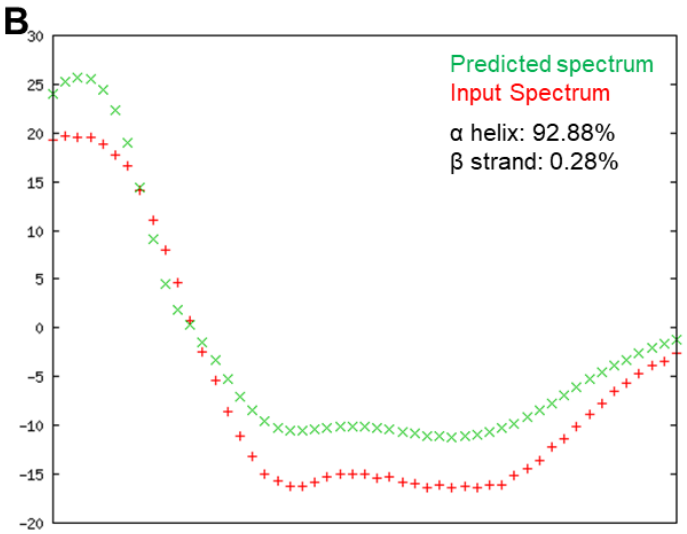
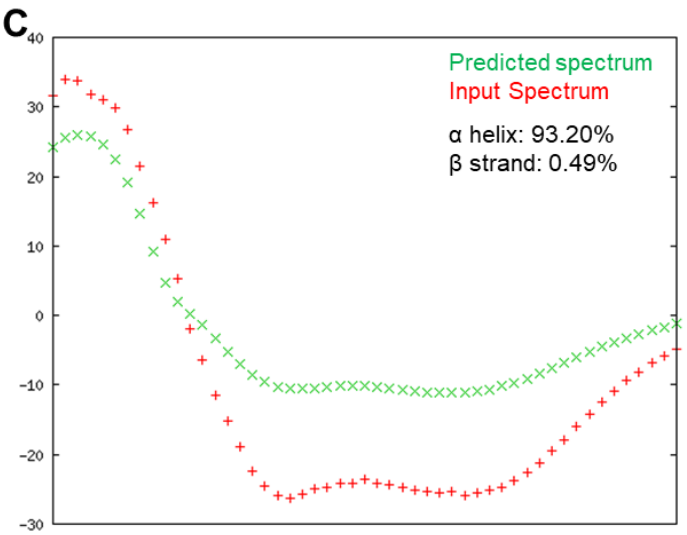


Figure 4-3 TALOS+-predicted secondary structure for $BCOR_{long}$

Probability that each residue in $BCOR_{long}$ is part of an α helix (red), β strand (blue), or random coil (orange). The gap represents unassignable residues 1204-1215.


A
40
30
20
10
0
-10
-20
-30
Δε
190
200
210
220
230
240
Wavelength (nm)
Difference
BCoR$_{long}$-AF9
BCoR$_{short}$-AF9


B
Predicted spectrum
Input Spectrum
α helix: 92.88%
β strand: 0.28%


C
Predicted spectrum
Input Spectrum
α helix: 93.20%
β strand: 0.49%

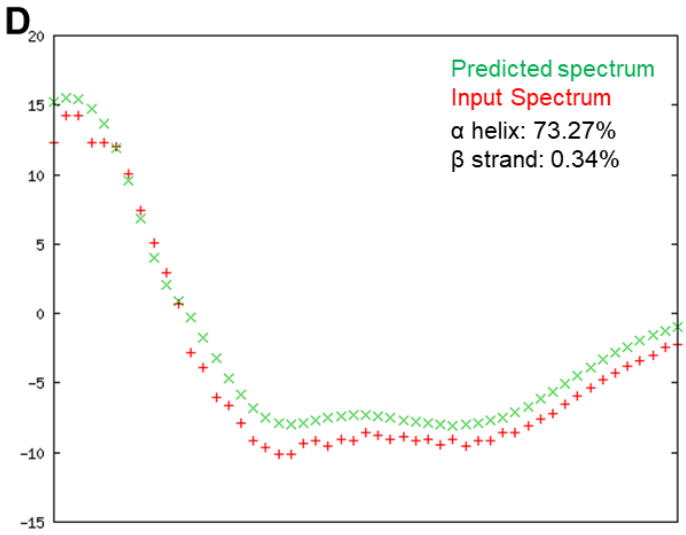


Figure 4-4. Circular dichroism spectra of the BCOR-AF9 complexes

(A) Circular dichroism spectra of $BCOR_{long}$-AF9 (red), $BCOR_{short}$-AF9 (orange), and the difference between the two (blue). Spectra are the average of 8 scans collected at 1 nm intervals.

(B-D) K2D3 predicted secondary structure from the CD spectra shown in A for $BCOR_{short}$-AF9 (B), $BCOR_{long}$-AF9 (C), and the difference between them (D). Input spectra are shown in red and predicted spectra in green. Alpha helix and beta strand percent predictions are given at top right.

4.4 Proline mutants recapitulate $BCOR_{short}$-AF9

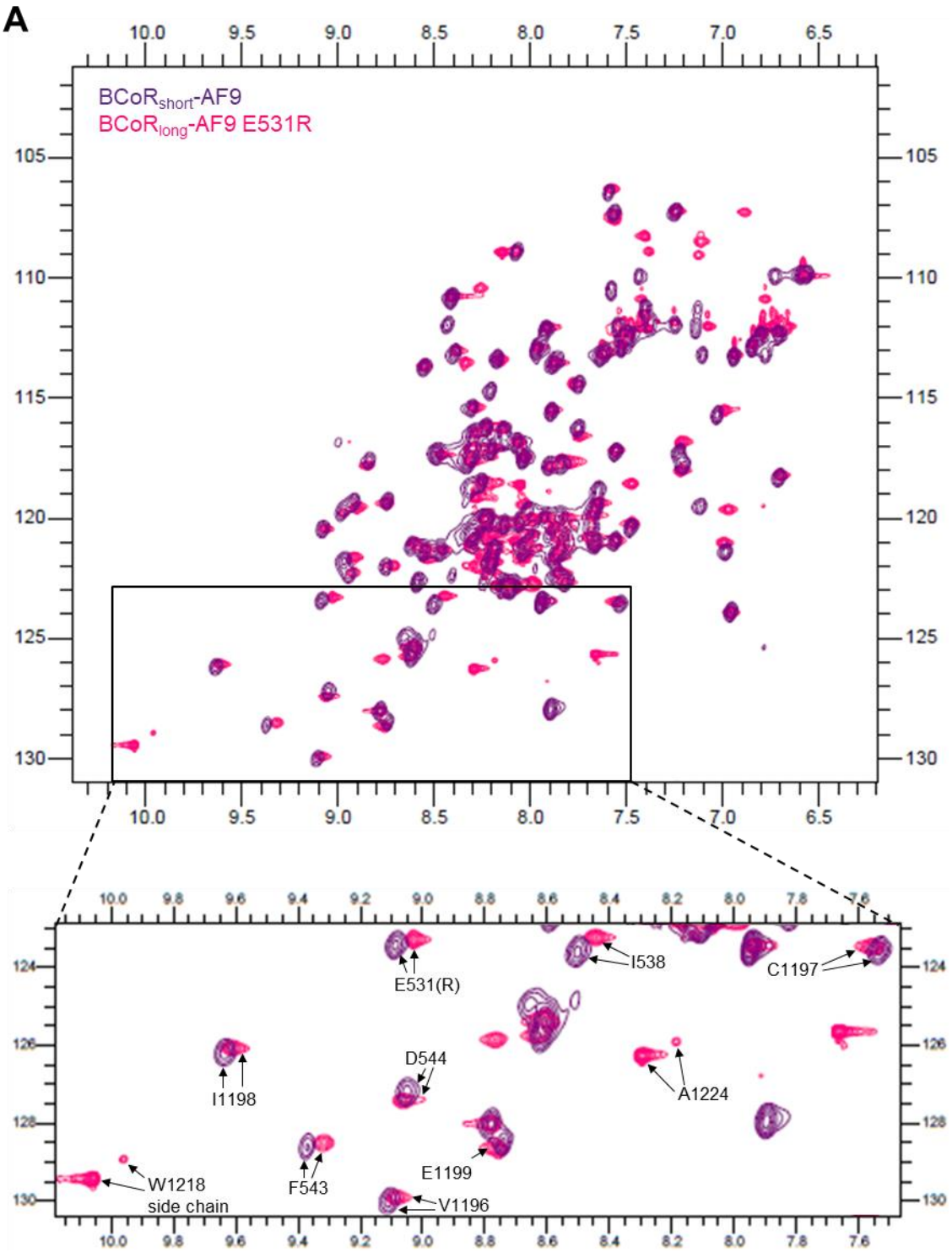
The previous sections provide indications, both theoretical and empirical, that the C-terminal region of $BCOR_{long}$ forms an α helix. If this is the case, mutation of residues in this region to proline, a residue which will interrupt secondary structure, should prevent formation of the putative helix. Without a direct readout of the chemical environment in this region, we endeavored to indirectly derive information about changes as a result of these mutations. Comparison of ^{15}N-^{1}H HSQC spectra for $BCOR_{long}$- and $BCOR_{short}$-AF9 indicates a significant difference in chemical shift for some AF9 and BCOR residues, particularly near the binding interface, as addressed in Figure 3-1. As these chemical shift perturbations (CSPs) are due to the additional residues in $BCOR_{long}$ and are presumably dependent upon the interaction of those residues with AF9, interruption of the α helix by mutation to proline should recapitulate the $BCOR_{short}$-AF9 spectrum, thus providing clear readout for this experiment.

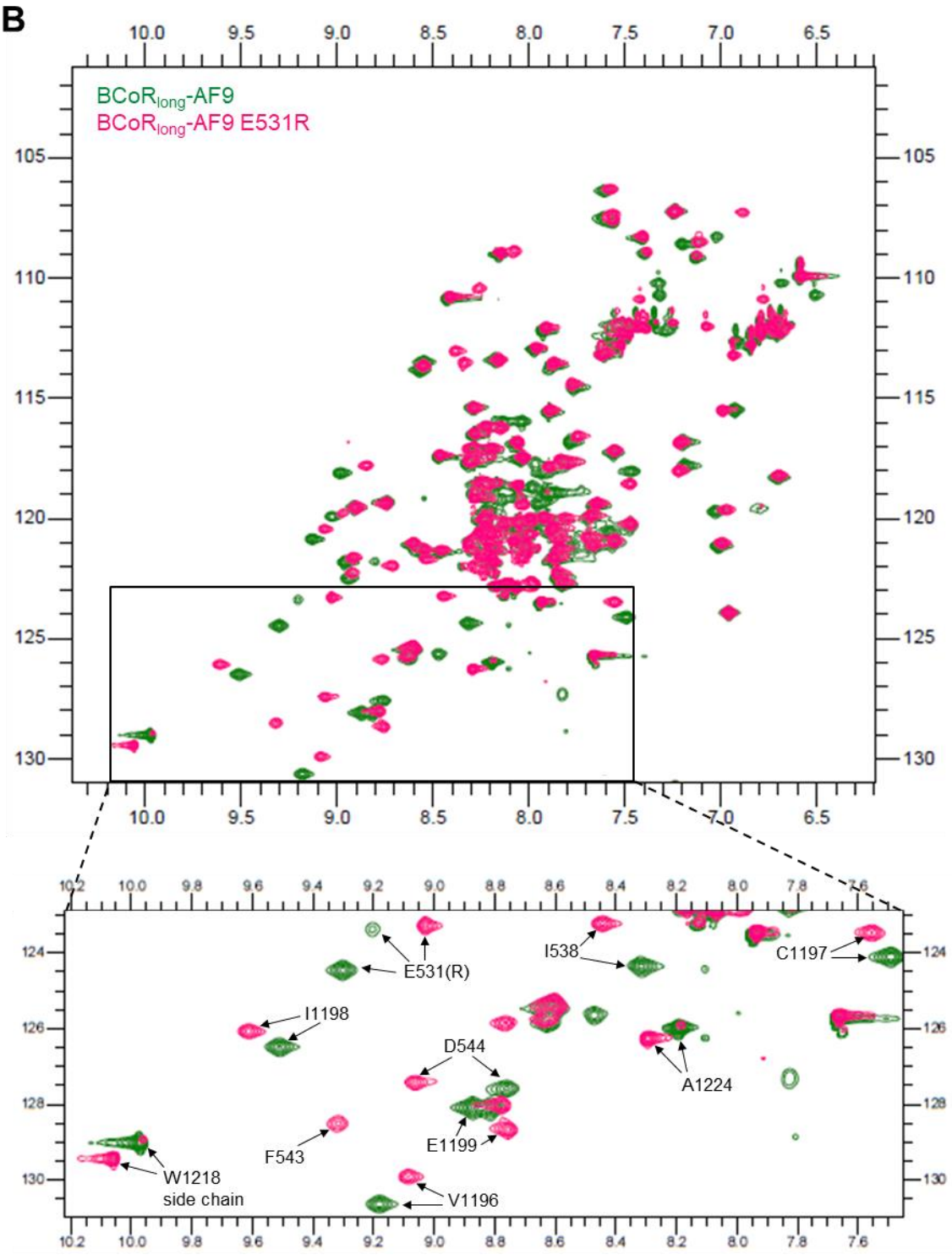
As proof of concept for this, the AF9 E531R mutation was introduced into the $BCOR_{long}$-AF9 expression vector. Its ^{15}N-^{1}H HSQC spectrum was collected and compared to the spectra of $BCOR_{long}$- and $BCOR_{short}$-AF9 (Figure 4-5). It is readily visually apparent that the E531R spectrum is nearly identical to that of $BCOR_{short}$-AF9 (Figure 4-5A). Naturally, not all $BCOR_{long}$-AF9 E531R peaks have corresponding peaks in the $BCOR_{short}$-AF9 spectrum as the former construct is 21 residues longer. However, the peaks present in both differ only slightly in chemical shift with the sole exception of T1201 (Figure 4-5C). It is possible that this residue, located near the C-terminal end of structured residues in the $BCOR_{short}$-AF9 complex, experiences a different chemical environment in the context of additional C-terminal residues that is not dependent upon their interaction with AF9. In contrast, the $BCOR_{long}$-AF9 E531R spectrum differs in significant ways from its WT counterpart (Figure 4-5B). As with the differences between

$BCOR_{long}$- and $BCOR_{short}$-AF9 shown in Figure 3-1, the largest differences between $BCOR_{long}$-AF9 WT and E531R are localized to the binding interface and the mutated residue, with a smaller secondary effect at AF9 helix 2 (Figures 4-5D-E). These residues provide a robust readout for changes in the interaction between AF9 and the C-terminal residues of $BCOR_{long}$.

Based on the assumption of an alpha helix from BCOR residue 1208 to 1219, two residues near its middle were selected for mutation to proline. ^{15}N-^{1}H HSQC spectra were collected for $BCOR_{long}$ L1212P- and $BCOR_{long}$ H1213P-AF9. Because these two mutants were nearly identical with negligible differences in chemical shift between them, further analysis was done using only the L1212P mutant (Figure 4-6).

The L1212P mutation (and by extension, the H1213P mutation), like the E531R mutation, yields a spectrum resembling that of $BCOR_{short}$-AF9 (Figure 4-7A). Again like E531R, the shift differences are negligible with the exception of T1201 (Figure 4-7C). Accordingly, the spectrum differs significantly from the $BCOR_{long}$-AF9 spectrum in ways similar to E531R (Figure 4-7B), with CSPs that also localize to the binding interface and AF9 helix 2 (Figure 4-7D-E). These comparisons suggest that mutation of L1212 to P interrupts the presumed helical secondary structure of $BCOR_{long}$, preventing its interaction with AF9. It should be noted, however, that comparison of the CSPs due to L1212P to those due to the E531R reveals observable differences indicating incomplete interruption of binding (Figure 4-7F). Here, as well, the largest shifts localize to the binding interface and AF9 helix 2, but it is important to point out that the shift differences between these two complexes are much smaller than those of either mutant compared to $BCOR_{long}$-AF9 WT. Comparing the spectra of both mutants and the WT, it becomes evident that for some residues, L1212P peaks seem to represent an intermediate state between the other two, but resemble the E531R peaks more closely (Figure 4-7G).


A
BCoR$_{short}$-AF9
BCoR$_{long}$-AF9 E531R
E531(R)
I538
C1197
D544
I1198
A1224
E1199
F543
W1218
side chain
V1196


B
BCoR$_{long}$-AF9
BCoR$_{long}$-AF9 E531R
E531(R)
I538
C1197
I1198
D544
A1224
F543
E1199
W1218
side chain
V1196

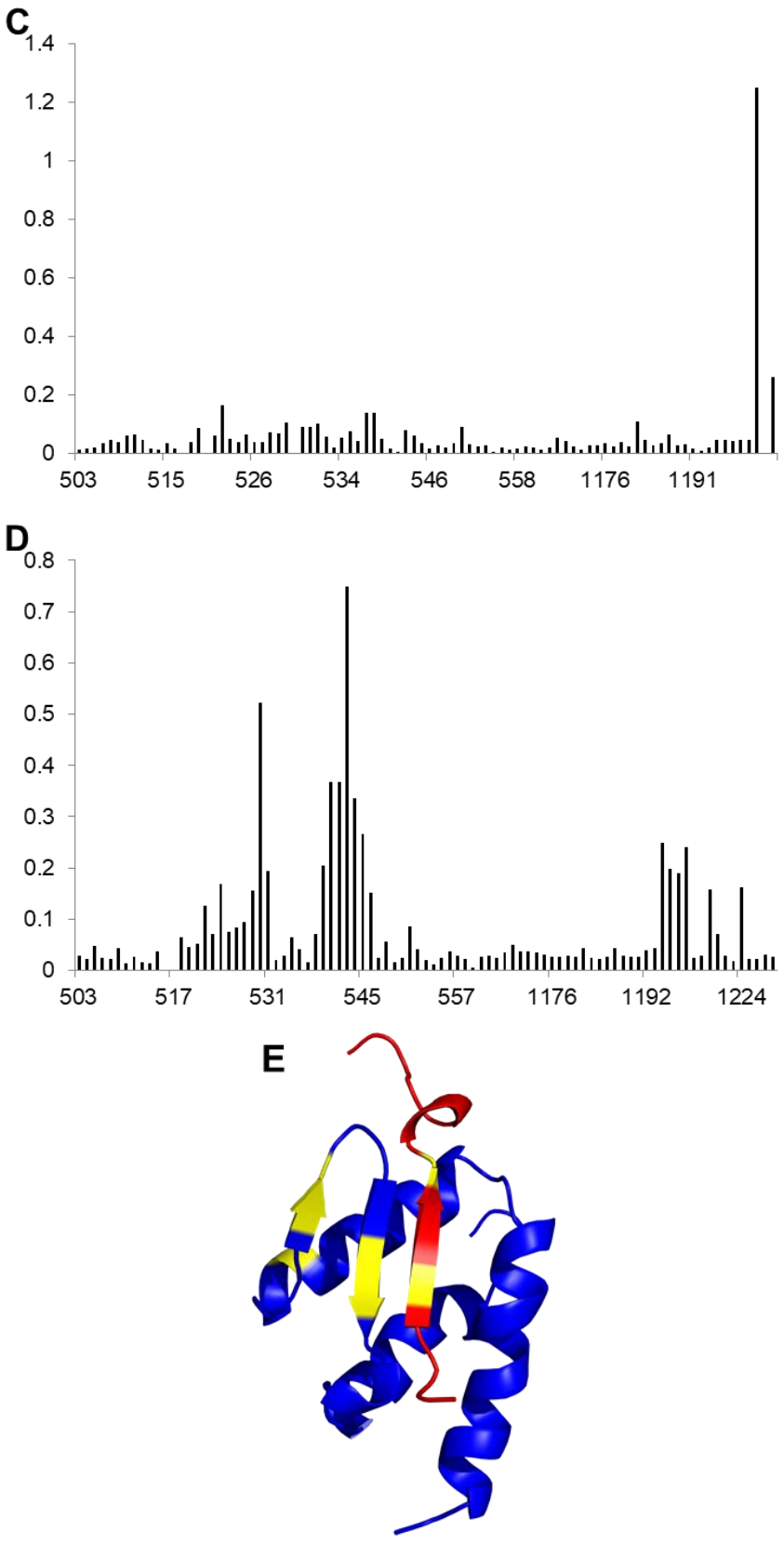

C
1.4
1.2
1
0.8
0.6
0.4
0.2
0
503
515
526
534
546
558
1176
1191
D
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
503
517
531
545
557
1176
1192
1224
E

Figure 4-5. BCOR-AF9 mutants affect the structure of the complex

(A) Overlay of the ^{15}N-^{1}H HSQC spectra of the $BCOR_{short}$-AF9 (purple) and $BCOR_{long}$-AF9 E531R (pink) complexes. The solid box in the full spectrum represents the region showed magnified below.

(B) Overlay of the ^{15}N-^{1}H HSQC spectra of the $BCOR_{long}$-AF9 (green) and $BCOR_{long}$-AF9 E531R (pink) complexes. The solid box in the full spectrum represents the region showed magnified below.

(C) Weighted chemical shift differences for $BCOR_{long}$-AF9 E531R versus $BCOR_{short}$-AF9 complexes calculated using $|\Delta\delta^{15}N|/4.69 + |\Delta\delta^{1}H_N|$ (Yuan et al., 2002).

(D) Weighted chemical shift differences for $BCOR_{long}$-AF9 E531R versus $BCOR_{long}$-AF9 complexes calculated as in (C).

(E) The largest shifts from (D) mapped in yellow on the lowest energy conformer of the $BCOR_{short}$-AF9 complex.

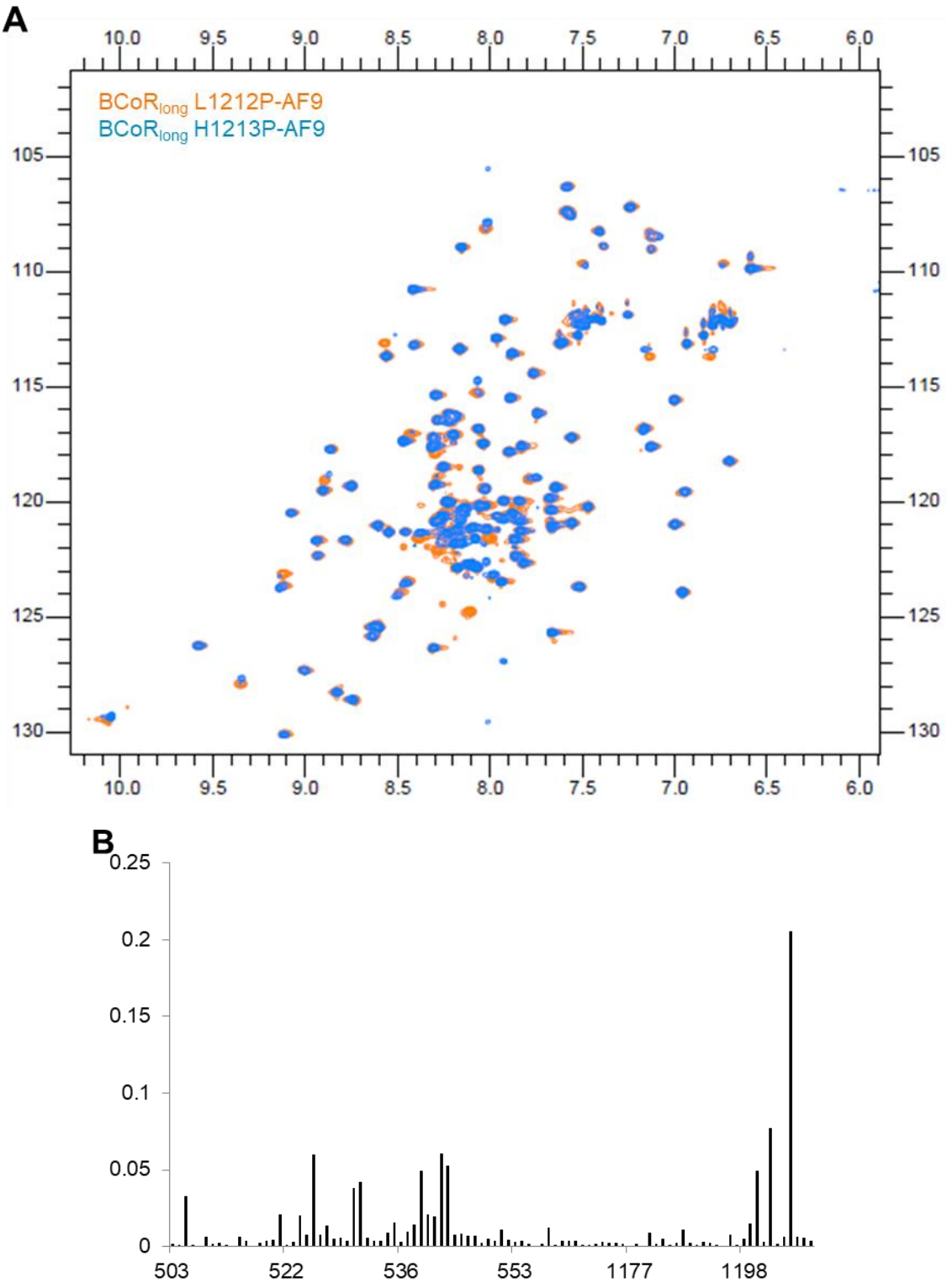
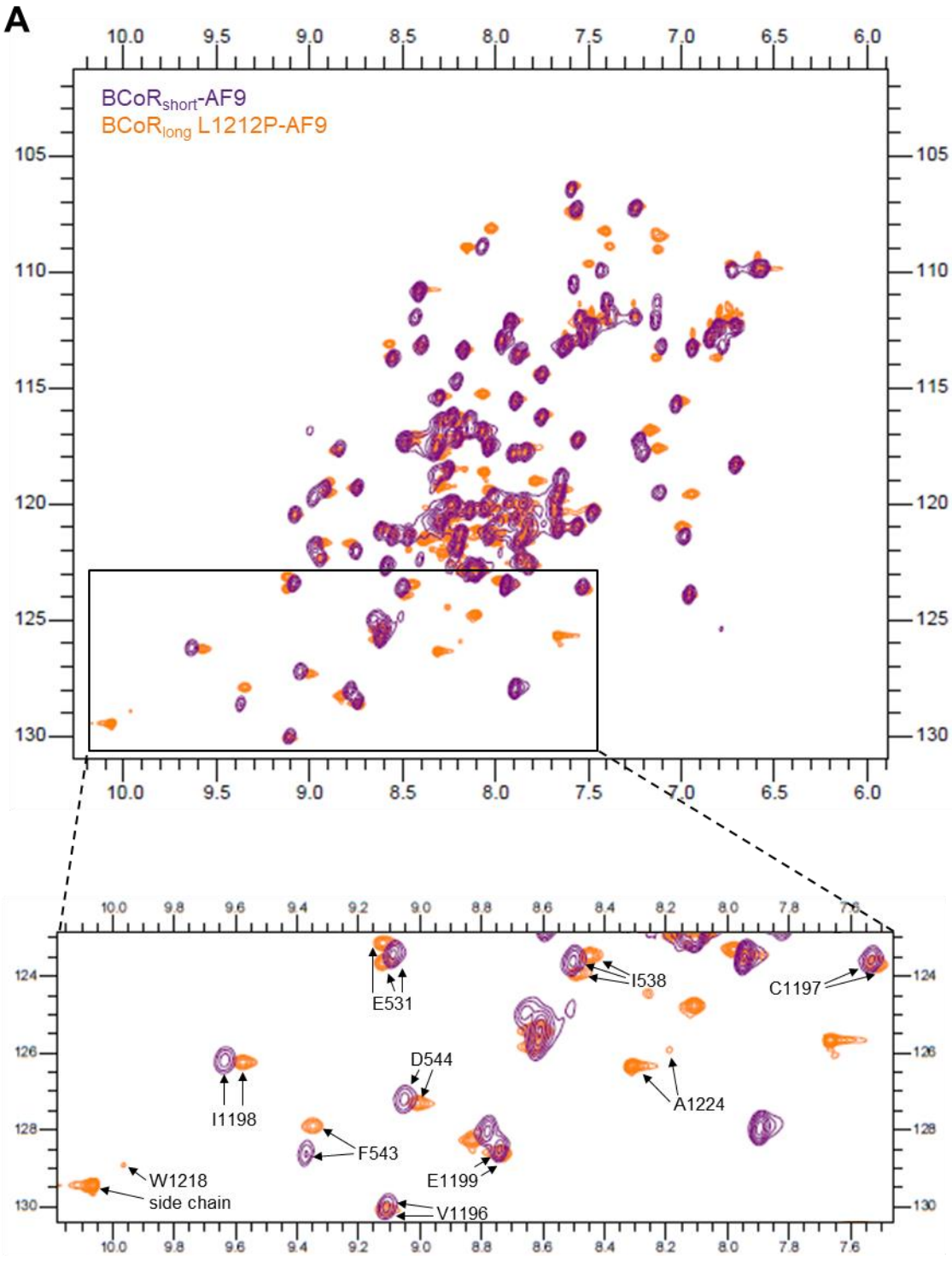
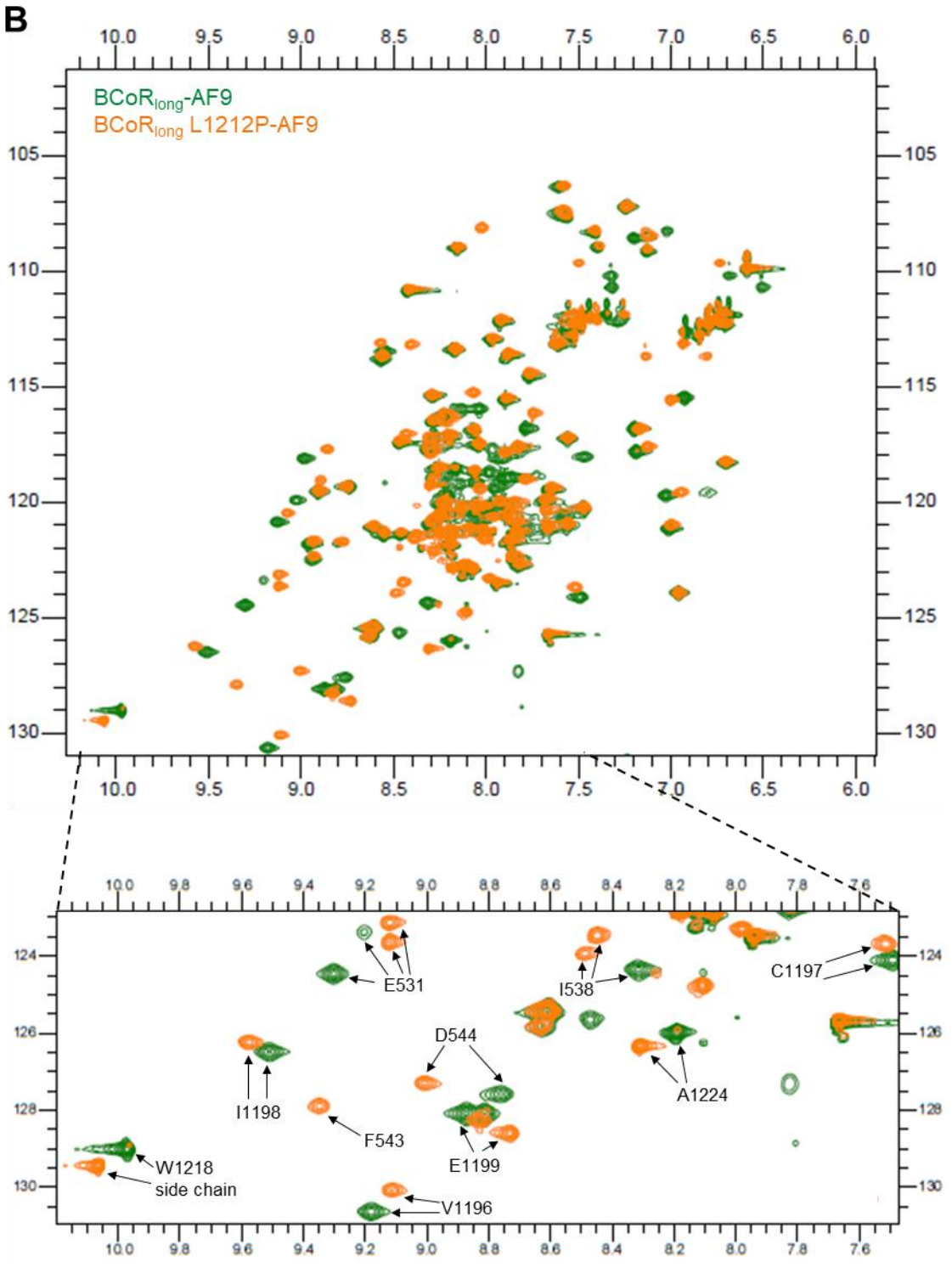


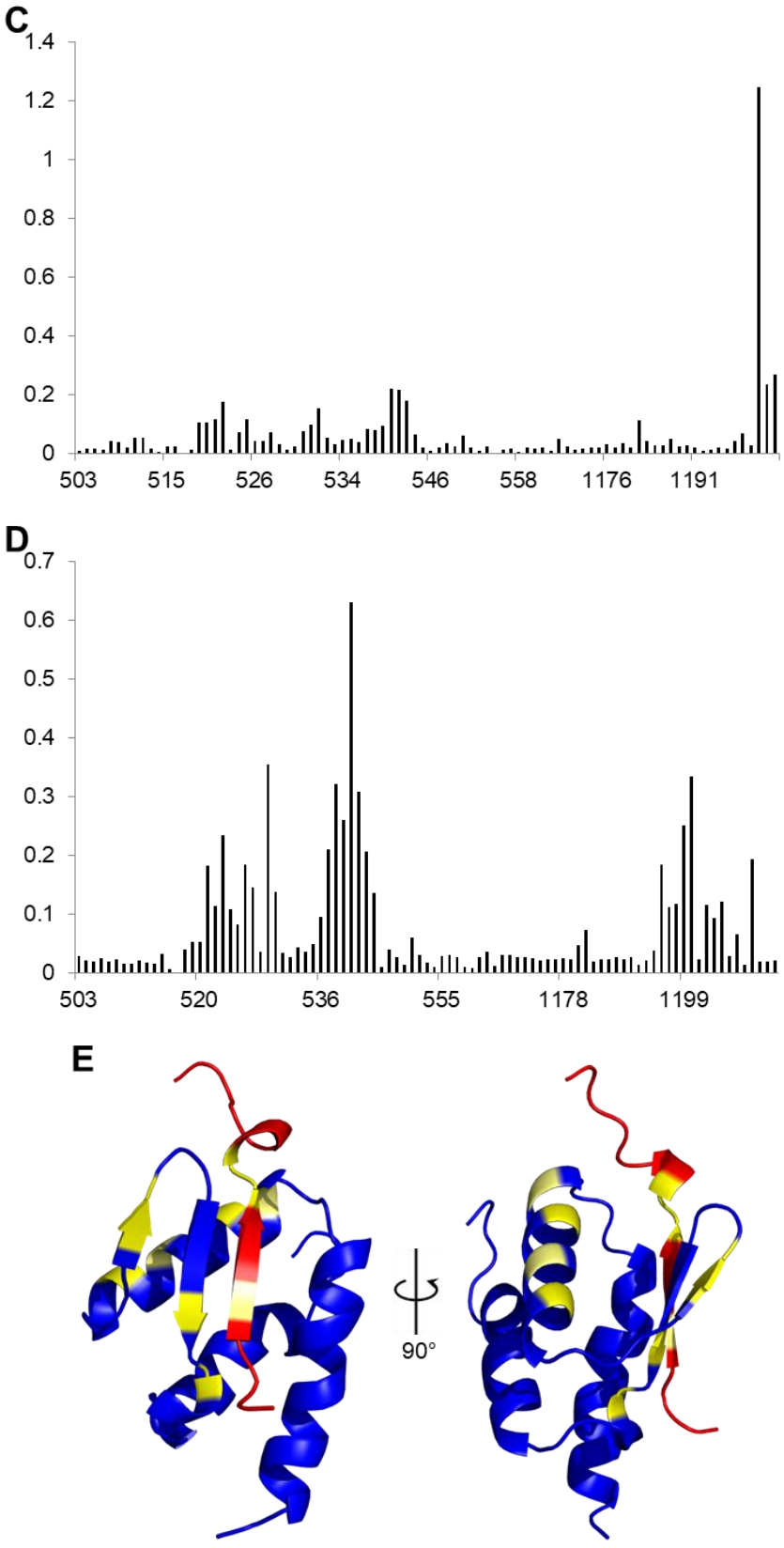
Figure 4-6. Comparison of BCOR$_{long}$ L1212P- and H1213P-AF9 mutants

(A) Overlay of the ^{15}N-^{1}H HSQC spectra of the $BCOR_{long}$ L1212P-AF9 (orange) and $BCOR_{long}$ H1213P-AF9 (blue) complexes.

(B) Weighted chemical shift differences for $BCOR_{long}$-AF9 E531R versus $BCOR_{short}$-AF9 complexes calculated using $|\Delta\delta^{15}N|/4.69 + |\Delta\delta^{1}H_N|$ (Yuan et al., 2002).


A
BCoR$_{short}$-AF9
BCoR$_{long}$ L1212P-AF9
E531
I538
C1197
D544
I1198
A1224
F543
E1199
W1218
side chain
V1196


B
BCoR$_{long}$-AF9
BCoR$_{long}$ L1212P-AF9
E531
I538
C1197
D544
A1224
I1198
F543
E1199
W1218
side chain
V1196


C
1.4
1.2
1
0.8
0.6
0.4
0.2
0
503
515
526
534
546
558
1176
1191
D
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
503
520
536
555
1178
1199
E
90°

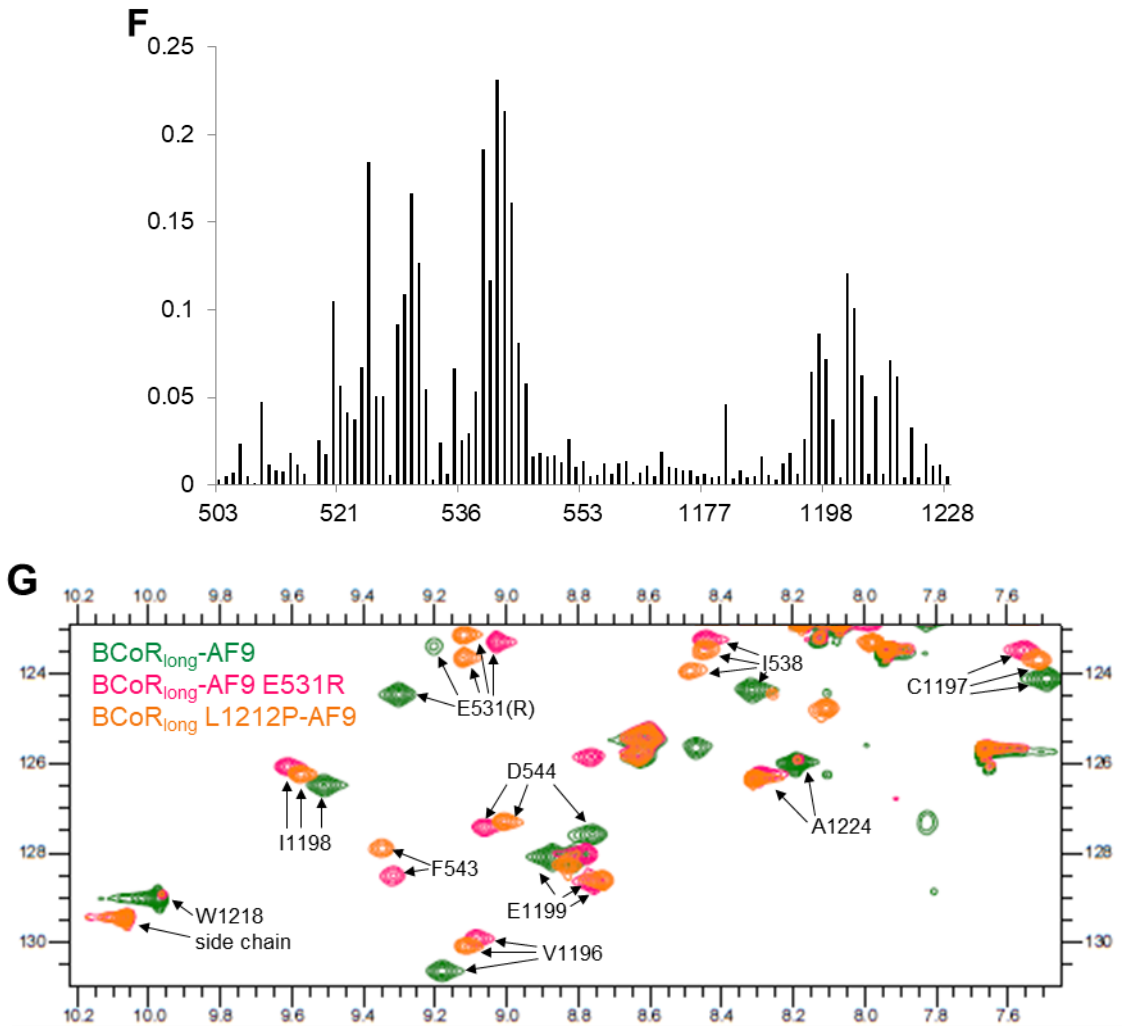


Figure 4-7. Proline mutations interrupt the secondary structure of BCOR$_{long}$-AF9

(A) Overlay of the ^{15}N-^{1}H HSQC spectra of the BCOR$_{short}$-AF9 (purple) and BCOR$_{long}$ L1212P-AF9 (orange) complexes. The solid box in the full spectrum represents the region showed magnified below.

(B) Overlay of the ^{15}N-^{1}H HSQC spectra of the BCOR$_{long}$-AF9 (green) and BCOR$_{long}$ L1212P-AF9 (orange) complexes. The solid box in the full spectrum represents the region showed magnified below.

(C) Weighted chemical shift differences for BCOR$_{long}$ L1212P-AF9 versus BCOR$_{short}$-AF9 complexes calculated using $|\Delta\delta^{15}N|/4.69 + |\Delta\delta^{1}H_N|$ (Yuan et al., 2002).

(D) Weighted chemical shift differences for $BCOR_{long}$ L1212P-AF9 versus $BCOR_{long}$-AF9 complexes calculated as in (C).

(E) The largest shifts from (D) mapped in yellow on the lowest energy conformer of the $BCOR_{short}$-AF9 complex.

(F) Weighted chemical shift differences for $BCOR_{long}$ L1212P-AF9 versus $BCOR_{long}$-AF9 E531R complexes calculated as in (C).

(G) Overlay of a region of the ^{15}N-^{1}H HSQC spectra of the $BCOR_{long}$-AF9 (green), $BCOR_{long}$-AF9 E531R (pink), and $BCOR_{long}$ L1212P-AF9 (orange) complexes.

4.5 Discussion

Despite the lack of readily available structural information on BCOR residues 1204-1215 by NMR, predictions and indirect observations make a compelling argument for the presence of an alpha helix that makes contact with AF9 to modulate binding affinity. HetNOE measurements are indicative of structure in neighboring assignable residues (Figure 4-1), and predictions based on primary amino acid sequence and chemical shift of assignable residues indicate the possible presence of an α helix (Figures 4-2, 4-3). Subtraction of CD spectra of $BCOR_{long}$-AF9 and $BCOR_{short}$-AF9 indicate the presence of alpha helical character equivalent to approximately 15 residues (Figure 4-4). Mutation of L1212 and H1213 to P, which would interrupt any secondary structure present, resulted in ^{15}N-^{1}H HSQC spectra reminiscent of the $BCOR_{short}$-AF9 complex (Figure 4-7A). These data indicate the likely presence of an α helix at the $BCOR_{long}$ C-terminus that is vital for tight BCOR-AF9 binding.

Based on chemical shift, TALOS+ predicts an α helix ending around BCOR E1219 (Figure 4-3). This combined with the rough assumption of a 15 residue helix based on K2D3 predictions from CD data (Figure 4-4D) would suggest a helix spanning residues 1204-1219. This is unlikely due to the proline at position 1205. It is possible, however, that the helix would begin with residue 1206. Residues 1206 and 1207 are disordered in the NMR solution structure of $BCOR_{short}$-AF9, but this is because they are the last two residues at the C-terminus of the construct. In the context of a longer BCOR peptide they may be more likely to adopt secondary structure. On the other hand, Jpred prediction based on primary sequence suggests the helix may start at the succeeding residue, Q1208 (Figure 4-2). A helical wheel projection (from a script produced by Don Armstrong and Raphael Zidovetzki) for residues 1208-1219 shows an amphipathic helix (Figure 4-8). The polar side of the helix contains a large number of charged residues,

outnumbering the hydrophobic residues on the opposite side. Enrichment of charged amino acids and depletion of hydrophobic amino acids is typical of IDRs and in this case likely contributes to the transient nature of the interaction with AF9, as there are fewer hydrophobic side chains to drive the packing of the helix against the rest of the BCOR-AF9 complex (Dunker et al., 2002; Romero et al., 2001; Vucetic et al., 2003). Amphipathic elements such as this one are so common among functional sites within predicted disordered regions that they have been proposed as a means of identifying such sites bioinformatically (Dyson and Wright, 2005).

Changes in chemical shift due to the L1212P mutation or the E531R mutation which disrupt the interaction with $BCOR_{long}$ can be mapped onto the surface of the $BCOR_{short}$-AF9 structure to give an indication of where the putative helix may interact (Figures 4-5E, 4-7E). The shifted residues are predominantly found on the first AF9 beta strand, the C-terminal margin of BCOR's AF9 binding motif, and along one side of AF9 helix 2. The largest shifts are all found on the same surface of the complex, indicating the C-terminal helix makes contacts in a groove formed by AF9 helix 2 and the beta hairpin (Figure 4-9).

Spectra of the proline mutants do not completely recapitulate the $BCOR_{short}$-AF9 spectrum in the way that the AF9 E531R spectrum does (Figure 4-7F). In fact, closer inspection of the mutant spectra compared to WT $BCOR_{long}$-AF9 shows that the BCOR L1212P mutation seems to occupy an intermediate state between the WT and the unbound state (Figure 4-7G). This may be the result of continued but less frequent interaction of the BCOR C-terminus with AF9. Even in the context of the proline mutation, the remaining residues may continue to fold into a shorter and/or less stable helix, demonstrating the significant propensity of IDRs to form short pre-organized structural elements (Fuxreiter et al., 2004).

The primary interaction of $BCOR_{long}$ with AF9 occurs via a short LXVXIXLXXL motif. Additional BCOR residues C-terminal to this motif make transient contacts with AF9 and significantly increase binding affinity. This description fits well within the definition of the flanking model of fuzzy complexes (Fuxreiter, 2011; Tompa and Fuxreiter, 2008).

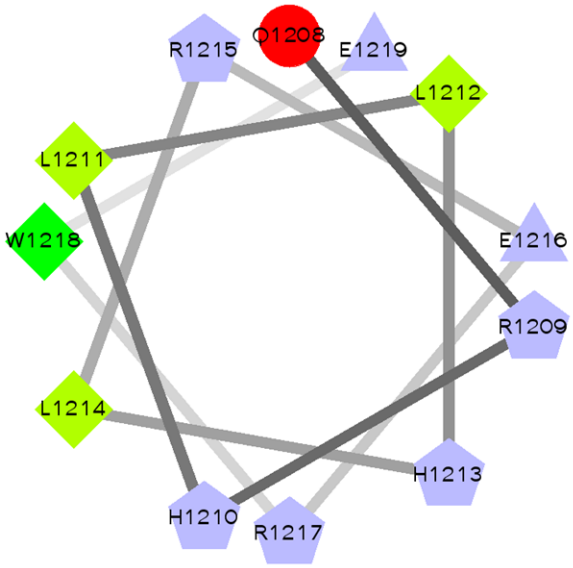


Figure 4-8. Helical wheel projection of the BCOR C-terminus

Helical wheel projection of BCOR residues 1208-1219 with charged residues in gray, polar residues in red, hydrophobic residues in yellow, and aromatic residues in green.

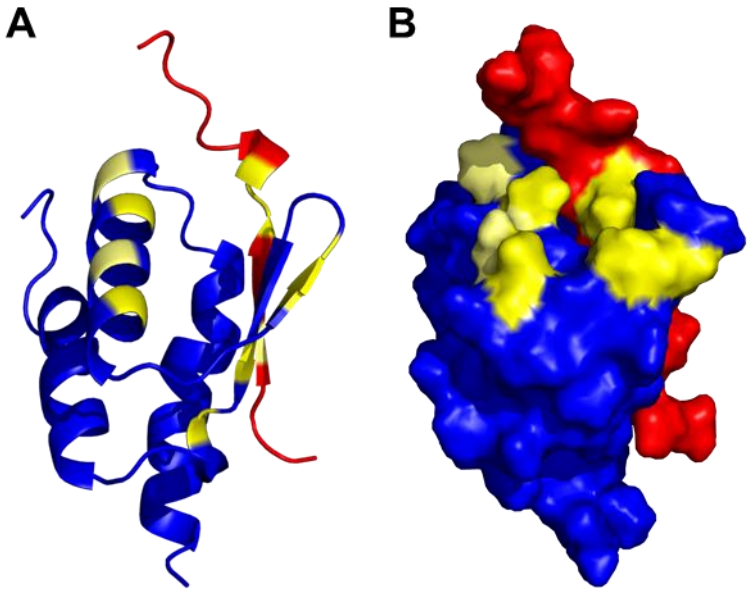


Figure 4-9. Mapping the interaction of $BCOR_{long}$ C-terminal helix

(A) Solution structure of the $BCOR_{short}$-AF9 complex with residues experiencing the greatest CSPs in yellow.

(B) Spacefill model of (A) showing the surface of the complex colored as above.

4.6 Experimental procedures

4.6.1 Protein expression and purification

For circular dichroism studies, BCOR residues 1176-1207 ($BCOR_{short}$) or 1176-1228 ($BCOR_{long}$) were coexpressed with AF9 residues 499-568 from pETDuet1 (EMD Millipore) modified in the case of $BCOR_{short}$-AF9 as described in Chapter 2. The complexes were expressed from Rosetta 2(DE3) cells (Novagen) in LB medium. The His and thioredoxin-His tags were removed by digestion with TEV protease and the complexes were purified by size exclusion chromatography using a Superdex 75 column (GE Life Sciences).

For NMR studies, these complexes or $BCOR_{long}$-AF9 mutants were expressed and purified similarly, but using European Molecular Biology Laboratory medium (www.embl.de/pepcore/pepcore_services/protein_expression/ecoli/n15_c13) with the modifications indicated in Chapter 2 and supplemented with 5 mL/L Bioexpress (Cambridge Isotopes) rather than trace elements.

4.6.2 Secondary structure prediction

JPred secondary structure predictions were made using the JPred 4 server with Jnet (http://www.compbio.dundee.ac.uk/jpred/) (Drozdetskiy et al., 2015). Secondary structure probabilities were derived from TALOS+ based on N, C', C^{α}, C^{β}, H^{α}, and H^{N} chemical shifts (Shen et al., 2009).

4.6.3 Circular dichroism

$BCOR_{short}$ and $BCOR_{long}$-AF9 were expressed as described in above. The complexes were desalted into 10 mM sodium phosphate pH 6.05 using PD-10 desalting columns (GE Life Sciences). CD experiments were conducted on 13 µM samples at 1 nm intervals with 0.5 s averaging time on an Aviv 410 spectropolarimeter. Spectra shown are an average of 8 scans. Secondary structure predictions were made with

K2D3 (Louis-Jeune et al., 2012).

4.6.4 NMR experiments

^{15}N-enriched BCORI_{ong}-AF9 WT, E531R, L1212P, and H1213P samples were prepared for NMR at a concentration of 750 µM in 25 mM Bis-Tris/MES, 100 mM NaCl, 1 mM DTT, pH 6.0 with 5% D_2O ^{15}N-^{1}H HSQC spectra were recorded on a Bruker 600 MHz magnet equipped with cryogenically cooled probes.

4.7 References for Chapter 4

Cuff, J., and Barton, G. (1999). Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. Proteins *34*, 508–519.

Cuff, J., and Barton, G. (2000). Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins *40*, 502–511.

Drozdetskiy, A., Cole, C., Procter, J., and Barton, G. (2015). JPred4: a protein secondary structure prediction server. Nucleic Acids Res. *43*, W389-94.

Dunker, A., Brown, C., Lawson, J., Iakoucheva, L., and Obradovic, Z. (2002). Intrinsic disorder and protein function. Biochemistry *41*, 6573–6582.

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

Fuxreiter, M. (2011). Fuzziness: linking regulation to protein dynamics. Mol. Biosyst.

Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004). Preformed structural elements feature in partner recognition in intrinsically unstructured proteins. J. Mol. Biol. *338*, 1015–1026.

Louis-Jeune, C., Andrade-Navarro, M., and Perez-Iratxeta, C. (2012). Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. Proteins. Structure, Function, and Bioinformatics. Proteins *80*, 374–381.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., and Dunker, A.K. (2001). Sequence complexity of disordered protein. Proteins Struct. Funct. Genet. *42*, 38–48.

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

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

Vucetic, S., Brown, C., Dunker, A., and Obradovic, Z. (2003). Flavors of protein disorder. Proteins *52*, 573–584.

Chapter 5. Perspectives and future directions

We have shown that the structure of the $BCOR_{short}$-AF9 complex is highly similar to that of the other AF9 complexes we have solved. We have also demonstrated that the minimal AF9 binding motif in BCOR binds AF9 weakly, particularly when compared to the other binding partners. However, a longer BCOR peptide with residues unassignable by NMR binds much more tightly (Chapter 2). We have used this structural information to create a charge reversal point mutation on AF9 that disrupts binding of BCOR without affecting the other binding partners (Chapter 3). This mutant, AF9 E531R, does not affect colony formation in mouse bone marrow cells, but does have an effect on proliferation. This is unique among the AF9 binding partners, as selective perturbation of the binding of AF4, DOT1L, or CBX8 has a distinct effect on colony formation. Furthermore, MLL-AF9 E531R to cause leukemia *in vivo*. RNAseq of cells transformed with MLL-AF9 WT and E531R demonstrates that loss of BCOR-AF9 binding affects expression of a unique subset of genes. Their overlap with MLL-AF9 ChIP targets or genes with loss of H3K79me2/3 upon loss of DOT1L-AF9 binding is minimal, but there is more significant overlap with genes bound by PRC1.1 in MLL-AF9 cells. This may indicate that the effects of abrogating BCOR binding to AF9 may manifest themselves at points in the genome nearby in three-dimensional space. We have demonstrated that residues of $BCOR_{long}$ which are unassignable by NMR have helical characteristics in prediction based on primary sequence, prediction based on chemical shift on nearby assignable residues, and CD (Chapter 4). We have exploited the difference in chemical shift of certain residues between the $BCOR_{short}$- and $BCOR_{long}$-AF9 complexes as a tool for determining presence or absence of secondary structure at the $BCOR_{long}$ C-terminus. Mutation of residues to proline in the middle of the putative helix produces spectra reminiscent of the $BCOR_{short}$-AF9 helix, indicating disruption of secondary structure.

CSPs indicate that the point of contact for the putative helix may be along AF9 helix 2 near the beta hairpin that forms part of the BCOR binding interface. The C-terminal BCOR residues would form an attractive amphipathic helix, the hydrophobic residues of which could pack in a groove formed between AF9 helix 2 and the beta hairpin.

Our results are the first investigation into the role of direct BCOR recruitment by AF9. Though these results shed some light on the nature of that interaction and the biological effects of its disruption, they also raise new questions. Changes in AF9 binding affinity for AF4, DOT1L, and CBX8 result in significant changes in colony formation. Loss of binding of the transcriptional activators AF4 and DOT1L or an increase in CBX8 binding affinity all lead to reduced colony formation. Based on these results, we had expected loss of BCOR-AF9 binding to lead to a change in colony formation. Instead, the observed phenotype was an increase in proliferative ability. In addition, the genes affected by the AF9 E531R mutation have little overlap with known MLL-AF9 ChIP targets. This was highly unexpected, as it suggests the effects of BCOR binding to MLL-AF9 are not exerted at the point of recruitment. Though there is some overlap of E531R-affected genes with known PRC1.1 and PRC1/PRC1.1 targets in MLL-AF9 cells, we do not know if these are direct effects or downstream changes as a result of immediately genes affected.

To deconvolute the RNAseq results and supply a mechanism by which BCOR affects gene expression in an AF9-dependent manner requires further experimentation which would constitute another project entirely. ChIPseq for BCOR in cells transformed with MLL-AF9 WT and E531R would identify the genomic locations which experience a decrease in levels of BCOR upon loss of AF9 binding. This would allow separation of direct BCOR targets from downstream effects in the RNAseq data. Given the evidence that BCOR's effects may be exerted distally relative to MLL-AF9, it may be more

informative to use chromosome conformation capture (3C) techniques, particularly Hi-C, to elucidate the three-dimensional component of the puzzle.

In addition to new questions, details of the BCOR-AF9 structure implicitly echo those we have previously raised. AF9 can bind four partners with a wide range of affinities (sub-nanomolar to likely low micromolar). The mechanism by which binding partners are exchanged remains unclear, though the ability to do so rapidly and in response to cellular conditions is a hallmark of IDPs. It may be mediated by PTMs, particularly phosphorylation near the binding site, which can have a dramatic effect on binding affinity. Additionally, locally high concentrations of a binding partner may allow it to outcompete others for AF9 occupancy. For instance, BCOR binding to AF9 may be mediated by BCOR recruitment to DNA by KDM2B as part of PRC1.1 or binding to BCL6. This would tether it near AF9, increasing the likelihood of displacing untethered AF4 or DOT1L, which bind more tightly. The modulation of BCOR-AF9 affinity by the putative BCOR C-terminal helix is also relevant to this discussion. The fuzziness of the complex means that the helix is not always in contact with AF9. Partial dissociation of BCOR from AF9 by virtue of this transient interaction may make BCOR more susceptible to displacement by AF4 or DOT1L. Alternatively, it could make BCOR more accessible for degradation or PTM. This remains highly speculative and requires more investigation to advance our understanding of the process(es) involved.

As mentioned in Chapter 1, AF9 shares significant sequence homology with ENL, another common MLL translocation partner. The AHD of AF9 is 80% identical to that of ENL, and this number is slightly higher when only structured residues are considered. Despite this, there are differences in MLL-AF9 and MLL-ENL leukemias. For instance, the former are most often myeloid in nature (AML), while the latter are generally lymphatic (ALL). Interestingly, the ENL AHD has different affinities for the

binding partners than AF9. In particular, CBX8 and BCOR bind much more tightly to ENL. These differences may explain some of the difference in MLL-AF9 and MLL-ENL leukemias. BCOR could play an intriguing role, via its corepression of BCL6 targets, as BCL6 is involved in lymphoid development and function. Of structured AF9 residues, only nine differ from their ENL counterparts. Four of these would be unlikely to affect binding based on distance from the binding site. There are two instances where the sequence HI in AF9 corresponds to NV in ENL; one of these occurs in the first AF9 beta strand. Though these residues do not make direct contacts at the binding interface, they see significant shift in ^{15}N-^{1}H HSQC spectra as a result of AF9 E531R, BCOR L1212P, and BCOR H1213P mutations. They could therefore be significant in making contacts with the BCOR C-terminal helix, but this is insufficient to explain the difference in affinity of AF9 and ENL for the $BCOR_{short}$ peptide. The most likely candidate for explaining the difference in affinity is AF9 C548, which corresponds to ENL F540. C548 is located just C-terminal to the second AF9 beta strand and has NOEs to residues in the BCOR peptide. It's seems unlikely that the large increase in side chain size from C to F at that location would have no effect on binding. Further study will be necessary to determine why the observed effect is much larger for the repressors than the activators. If this single amino acid difference is responsible, the changes in binding can be studied by two simple mutations, AF9 C548F and ENL F540C, but more effort will be required to investigate the structural basis for the differences.

Inhibition of DOT1L enzymatic activity has shown promise as treatment for MLL leukemias. Given the effects of abrogation of AF9 AHD binding to AF4, DOT1L, and BCOR, this domain provides another potential therapeutic target for treatment of MLL-AF9 leukemias. Because of the homology of AF9 to ENL, any treatments effective at inhibiting AF9 binding to its partners may have similar effects on ENL. Targeting the

unstructured AHD presents a unique challenge, however. Successful inhibition with small molecules generally relies on protein structure and topology, and the intrinsic disorder of this domain in the unbound state makes it more difficult to drug than the average transcription factor.