Structural and Biochemical Mechanisms of MLL1 Activation on Chromatin

by

Alex M. Ayoub

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in The University of Michigan
2022

Doctoral Committee:

Professor Yali Dou, Co-Chair
Professor Anna Mapp, Co-Chair
Professor Uhn-soo Cho
Professor Greg Dressler
Professor Min Su

Alex M. Ayoub

bouya@umich.edu

ORCID iD: 0000-0003-0341-8747

Dedication

This dissertation is dedicated to my mother, Maureen, who has worked tirelessly to support me throughout my life and academic career. She continuously sacrifices herself for the improvement of others in her life and I am no exception. I love you mom.

Acknowledgements

Who could have predicted that of all the roads I could have possibly traversed in my life, the one I chose would lead to a Ph.D.? Certainly not me. It has been a long and often challenging time while studying at the University of Michigan, and I have to show my appreciation for those here and elsewhere that have supported me through this journey. Despite how it may seem, without their support, this whole adventure would have ended abruptly years ago.

First, I naturally need to thank my advisor, Yali Dou. She is one of the most dynamic human beings I have ever met during my scientific journey. While she is exceptionally talented, brutally motivated, hyper attentive, passionate, and endlessly driven in her pursuits, she still manages to be surprisingly modest, caring, considerate, humorous, and eternally patient. I want to thank you for never giving up on me despite providing you ample opportunities to do so. I also need to thank you for never being satisfied with any result and constantly pushing ideas forward with me. Under you, I have become a much better independent scientific mind and am grateful for the latitude and support you always provided me.

I also need to thank my committee members: Anna, Uhn-soo, Greg, and Min for their constant guidance during my time here. You have constantly forced me to think more critically of my science, the motivations behind my work, and the questions being asked. Your willingness to have individual conversations have been instrumental in my continued pursuits. I am sincerely appreciative of my collaboration with Uhn-soo’s postdoc Sangho. Together, we worked hard and maintained effective communication leading to several manuscripts. Without the efforts of Uhn-soo and Sangho, the last few years would have been exceedingly difficult and much less prosperous, so I am forever in your debt. I need to also thank Anna for taking time of out her days on many occasions to have conversations (and a laugh or two) with me. This support was always welcome and appreciated.

I need to thank the Dou lab, too, for being a fun and thrilling lab to work in. We had our independent corner etched out in the empty hallways of Med Sci I and within that lab, we had an exceptional group of brilliant, caring, motivated scientists. I miss our days together in and out of lab and our conversations and interactions were a constant reminder why I joined Yali's group – thank you.

I want to specifically thank my previous postdoc mentor Young-tae Lee, who inspired me to continue my work in the field of structural biology. Young is an exceedingly talented, precise, organized, perceptive scientist who taught me the value in data interpretation and experimental setup. Working alongside Young taught be to be patient in my pursuits and I am forever thankful for his wisdom and guidance, which I will always keep in mind moving forward in my scientific career.

I want to thank the few people from Michigan whose friendship was essential to my sanity. Dr. Nick Borotto whose scientific chats, occasional beers, and late-night writing was always a welcome break from the strains of lab. Though he left Michigan for greener pastures, he still entertains my ideas and judges my scientific work. Thanks for being friend when it was needed and keeping me sane during my stay in Michigan.

Beyond and before Michigan, there were a multitude of people who helped me get here that deserve recognition. I have to thank my undergraduate advisor Will Pomerantz who despite his overwhelming talents, brilliance, and tireless efforts in the lab has to be one of the humblest people I have ever met. Without any good reason, Will took me on in his lab after taking his course at UMN. Will gave me unprecedented freedom in my projects and allowed me to tackle diverse, independent projects. Through his support, then and now, I pursued my doctoral degree and would never have likely done so without his words of support and motivation. To this day, Will continues to support me without question and I am forever grateful for him giving me a chance; he taught me the value of independent scientific investigation. He was the catalyst that got me where I am today.

I want to also thank my family. Since enrolling at the University of Minnesota, they have been patient with my lack of availability and time away from home. I want to thank my mother who is ceaselessly giving in all aspects of her life. Despite being a single working mother early in my life, she succeeded in raising three loving children and caring for her four grandchildren. She sacrifices everything for her family, and I have been the constant recipient of those sacrifices for my entire life. She inspires me to never forget about the people and things that matter and put family before everything else. I also need to thank my sisters, Rachel and Deena, for their unquestionable love and support despite my constant absence from home during the holidays and birthdays.

Table of Contents

CHAPTER 2. Cryo-EM Structure of the Human Mixed Lineage Leukemia-1 Complex

CHAPTER 3. Mechanism for DPY30 and Disordered Regions in ASH2L to Modulate the MLL/SET1 Activity on Chromatin

CHAPTER 4. Regulation of MLL1 Methyltransferase Activity in Two Distinct Nucleosome Binding Modes

CHAPTER 5. Conclusions and Future Directions

List of Figures

List of Tables

List of Abbreviations

53BP1	p53 binding protein
A	Alanine
A-loop	Anchoring loop
AF9	ALL-fused gene from chromosome 9
ALL	Acute lymphocytic leukemia
ARM	Arginine-rich motif
AS-ABM	Activation segment; ASH2L binding motif
ASH1L	Absent, small, or homeotic disc 1-like
ASH2L	Absent, small, or homeotic disc 2-like
ATP	Adenosine triphosphate
bp	Base-pair
BRD	Bromodomain
Bre2	Brefeldin-A sensitivity protein 2
BSA	Bovine serum albumin
C	Cysteine
CaCl2	Calcium chloride
cAMP	cyclic 3'5' adenosyl monophosphate
Cas9	CRISPR-associated protein 9
CBB	Coomassie Brilliant Blue
CBP	CREB-binding protein
CC	Correlation coefficient
CFP1	CxxC finger protein 1
CHD	Chromodomain-helicase DNA binding
ChIP	Chromatin immunoprecipitation
CpG	5'-C-phosphate-G-3'

Cps40	Complex protein associated with SET1 protein 40
Cps50	Complex protein associated with SET1 protein 50
Cps60	Complex protein associated with SET1 protein 60
CREB	cAMP-response element binding protein
CRISPR	Clustered regularly interspaced short palindromic repeats
Cryo-EM	Cryogenic-electron microscopy
Cryo-ET	Cryogenic-electron tomography
Cryo-FIB	Cryogenic-focused ion beam
CTF	Contrast transfer function
CUT&RUN	Cleavage under targets & release using nuclease
D	Aspartic acid
D2O	Deuterium oxide
DD	Dimerization domain
DMEM	Dulbecco's modified eagle medium
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DOC	Sodium deoxycholate
DOT1L	Disruptor of telomere silencing 1-like
DPY30	DumPY protein 30
DTT	Dithiothreitol
DUB	Deubiquitylase
E	Glutamic acid
E14	E14tg2a
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene-bis(oxyethylenenitrilo)tetraacetic acid
EMSA	Electrophoretic mobility shift assay
EOM	Ensemble-optimized method
ESCs	Embryonic stem cells
EZH2	Enhancer of zeste homolog 2
FG-MD	Fragment-guided molecular dynamics

FPLC	Fast protein liquid chromatography
FSC	Fourier shell correlation
G	Glycine
gRNA	Guide RNA
GST	Glutathione S-transferase
H	Histidine
H2A	Histone H2A
H2AK15ub	H2A K15 monoubiquitylation
H2B	Histone H2B
H2BK120ub	H2B K120 monoubiquitylation
H3	Histone H3
H3K27	Histone H3 lysine 27
H3K27me3	H3 K27 trimethylation
H3K36	Histone H3 lysine 36
H3K36me2	H3 K36 dimethylation
H3K4	Histone H3 lysine 4
H3K4me1	H3 K4 monomethylation
H3K4me2	H3 K4 dimethylation
H3K4me3	H3 K4 trimethylation
H3K79	Histone H3 lysine 79
H3K9	Histone H3 lysine 9
H4	Histone H4
H4K16	Histone H4 lysine 16
H4K20me2	H4 K20 dimethylation
HA	Hemagglutinin
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDM	Histone demethylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HKMT	Histone lysine methyltransferase
HMQC	Heteronuclear multiple quantum correlation

HMT	Histone methyltransferase
I	Isoleucine
I-loop	Inserting loop
I-TASSER	Iterative template-based fragment assembly refinement
IDR	Intrinsically-disordered region
IgG-HRP	Immunoglobulin G-conjugated horseradish peroxidase
ILV	Isoleucine-Leucine-Valine
INO80	Inositol-requiring mutant 80
IPTG	Isopropyl β-D-1thiogalactopyranoside
ISWI	Imitation Switch
K	Lysine
KH2PO4	Potassium phosphate monobasic
KIX	Kinase-inducible domain-interacting domain
L	Leucine
LANA	Latency-associated nuclear antigen
LEDGF	Lens epithelium-derived growth factor
LiCl	Lithium chloride
LIF	Leukemia inhibitory factor
lncRNA	Long non-coding RNA
M	Methionine
MC	Monte Carlo
MENIN	Multiple endocrine neoplasia 1
MgCl2	Magnesium chloride
MgSO4	Magnesium sulfate
MLL	Mixed lineage leukemia
MNase	Micrococcal nuclease
mRNA	Messenger RNA
N	Asparagine
Na2HPO4	Sodium phosphate dibasic
NaCl	Sodium chloride
NaHCO3	Sodium bicarbonate

NCP	Nucleosome core particles
NH4Cl	Ammonium chloride
Ni-NTA	Nickel nitrilotriacetic acid
NMR	Nuclear magnetic resonance
NP-40	Nonident P-40
OD	Optical density
P	Proline
PAF1C	Polymerase associating factor elongation complex
PAGE	Polyacrylamide gel electrophoresis
PBase	piggyBac transposase
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PHD	Plant homeodomain
PHD-WH	Plant homeodomain-winged helix
PMSF	Phenylmethylsulfonyl fluoride
PPI	Peptidyl prolyl isomerase
PPIs	Protein-protein interactions
PRC2	Polycomb repressive complex 2
PRD-BF1	Positive regulatory domain I-binding factor 1
PRDM	PRD-BF1 and RIZ domains
PTM	Post-translation modification
PVDF	Polyvinylidene difluoride
Q	Glutamine
qPCR	Quantitative polymerase chain reaction
R	Arginine
RbBP5	Retinoblastoma binding protein 5
RCC1	Regulator of chromatin condensation
REMC	Replica-exchange Monte Carlo
RIZ1	Retinoblastoma protein-interacting zinc finger gene 1
RNA	Ribonucleic acid
RNAPII	RNA polymerase II

RNase	Ribonuclease
RRM	RNA recognition motif
rTTA	Reverse tetracycline-controlled transactivator
S	Serine
SAGA	Spt-Ada-Gcn5 acetyltransferase
SAH	*S*-adenosyl-L-homocysteine
SAM	*S*-adenosyl-L-methionine
SANT	Sw3, Ada2, N-Cor, TFIIIB
SAXS	Small angle X-ray scattering
SBD	SANT1-like binding domain
Sdc1	Suppressor of CDC25 protein 1
SDI	Sdc1/DPY30 interaction
SDS	Sodium dodecyl sulfate
SEC-SAXS	Size-exclusion chromatography small angle x-ray scattering
SEM	Standard error of the mean
SET	Suppressor of variegation 3-9, enhancer of zeste, trithorax
SET-I	SET insertion
SHL	Superhelical location
siRNA	Small interfering RNA
SNF	Sucrose non-fermentable
SPP1	Suppressor of PRP protein 1
SPRY	SpIa and ryanodine receptor
SRM	Stimulation-response motif
SUMO	Small ubiquitin-related modifier
TAD	Trans-activation domain
TAE	Tris-acetate EDTA
TBE	Tris-borate EDTA
TBS	Tris-buffer saline
TCEP	Tris(2-carboxylethyl)phosphine
TE	Tris-EDTA
TEV	Tobacco etch virus

Tris-HCl	Tris(hydroxymethyl)aminomethane-hydrochloride
TROSY	Transverse relaxation-optimized spectroscopy
TRR	Trithorax-related
TRX	Trithorax
TSS	Transcriptional start site
ULP1	Ubiquitin-like-specific protease 1
UV	Ultraviolet
V	Valine
WD	Tryptophan-Aspartic acid
WDR5	WD repeat-containing protein 5
Zn	Zinc

List of Appendices

Abstract

Transcription is a critical process by which cells regulate temporal and spatial expression of genes. In eukaryotes, histone H3 lysine 4 methylation by the MLL/SET1 family histone methyltransferases is enriched at transcription regulatory elements including gene promoters and enhancers. The existence of six functionally distinct MLL/SET1 H3K4 methyltransferase family members further underscores the biochemical complexity of transcriptional regulation. The level of H3K4 methylation is highly correlated with transcription activation and is one of the most frequently used histone post-translational modifications to predict transcriptional outcome. Recently, it has been shown that rearrangement of the cellular landscape of H3K4 mono-methylation at distal enhancers precedes cell fate transition and is utilized for identification of novel regulatory elements for development and disease progression. Similarly, broad H3K4 tri-methylation regions have also been used to predict intrinsic tumor suppression properties of regulatory regions in a variety of cancer models.

Understanding the mechanisms of H3K4 methylation deposition and its regulation is of paramount importance as dysregulation of these enzymes almost universally results in disease establishment and/or progression of developmental disorders and malignant transformation in cancers. Therefore, determining how MLL/SET1 members engage and methylate chromatin, their native substrate, is central to deconstructing the mechanistic requirements for H3K4 methylation in cells. In this thesis, we will provide molecular insight

on how MLLs engage their substrate histone H3 lysine 4 (H3K4) on chromatin and how this interaction is modulated leading to functional impacts on transcriptional regulation.

First, we use structural and biochemical methods to investigate how MLL1 catalytic SET domain (MLL1^{SET}) binds to nucleosome core particles (NCPs), its native substrate. Using single particle cryo-EM, we show that MLL1 binds near the dyad axis through ASH2L and RbBP5 binding motifs, the majority of which interact with nucleosomal DNA. We show loss of these motifs attenuate MLL1^{SET} catalysis *in vitro* in an NCP-specific fashion underscoring their importance in MLL1^{SET} engaging chromatin.

Next, we use advanced NMR and cellular work to show that central to this MLL1^{SET}-NCP interaction is DPY30. We first show that MLL1^{SET} is capable of higher processivity on an NCP substrate compared with recombinant H3. We reveal that DPY30, which binds at the ASH2L C-terminus, is central to this effect, functions universally amongst MLL/SET family members and acts independently and cooperatively with the H2BK120ub activating mark. We find novel mechanisms regarding DPY30 in the MLL1 complex on the NCP. We show that DPY30 induces drastic changes in ASH2L intrinsically disordered regions (IDRs) resulting in newly resolved resonances. We find that loss of any ASH2L IDR attenuates DPY30-mediated stimulation and removal of DPY30 from the MLL1^{SET} complex induces immense rotational dynamics on the NCP interface with RbBP5 as the sole anchor and overall instability in the ASH2L arm. Lastly, we show that *de novo* H3K4me3 in cells depends strictly on DPY30.

Finally, we use cryo-EM to show that MLL1^{SET} complex engages with NCP in a dynamic interplay of two discrete interaction modes. Using a catalytically inert H3 K4-to-M NCP (NCP^{K4M}), we readily capture these distinct states. Specifically, we show biochemically that interaction motifs found in this alternative mode do not strongly affect MLL1^{SET} methylation or binding in an NCP-specific manner. Despite overall strong structural agreement with ySET1-NCP, our results suggest MLL1^{SET}-NCP regulation occurs divergently through unique rotational dynamics on the NCP interface with ASH2L acting as an anchor.

These findings provide significant structural and biochemical insights into several key aspects of MLL1^{SET} regulation on a native substrate, the nucleosome. Our structural findings provide distinct insight into how the MLL1^{SET} complex recognizes, binds, and is activated on the NCP. We also provide new evidence for divergent modes of regulation distinct from ancient yeast homolog, ySET, from which MLL1 derived namely through unique rotational dynamics. Additionally, unique from ySET1, we show critical functions for conserved subunit DPY30. We revealed novel roles in MLL1^{SET} activation on NCP through IDRs, maintenance of complex and NCP-bound stability, and acting independently of H2BK120ub. These findings provide critical structural and biochemical insights into catalytic regulation of MLL1^{SET} previously unknown. They provide mechanistic insights into a completely novel mechanism of methyltransferase regulation through intrinsically disordered regions (IDRs). Disordered regions exist abundantly in many epigenetic proteins and their complexes critically acting as interaction hubs in transduction pathways. As a result, these findings provide foundational evidence for novel catalytic regulatory roles for IDRs previously ignored.

CHAPTER 1.
Insights on the Regulation of the MLL1/SET1 Family Histone Methyltransferases

1.1 Introduction

Cells are complex, information-processing centers that handle an immense flow of signals often leading to fine tuning the expression of genes. DNA, and genes found therein, are wrapped around an octameric core of histone proteins, creating a singular functional unit, generally known as the nucleosome core particle (NCP) [7]. Sequentially linked NCPs, classically known as “beads on a string”, are further wrapped into 30 nm chromatin fibers as part of the chromatin compaction hierarchy [8, 9]. Finally, these are packed into the large structures known as chromosomes that store our genetic information [10]. Transcription is an essential molecular process by which genes and gene products are unwrapped from nucleosomes, becoming accessible and finally expressed, thereby fulfilling the evolving needs of the organism. In eukaryotes, this process involves the complex orchestration of proteins to access and read DNA sequences, transcribe a complementary strand of messenger RNA (mRNA) and ultimately translate into polypeptides by the ribosome. Given the many factors involved in

[1]The contents of this chapter were adapted from a published co-first authored manuscript: Sha, L.*, **Ayoub, A.*,** Cho, U.-S., Dou, Y. "Insights on the regulation of the MLL/SET1 family histone methyltransferases." *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* **1863**(*7*): 194561 (2020).

this process, precise regulation of each step is essential throughout the lifetime of the cell and organism as a whole. This regulation is one of the main facets of epigenetics and, despite the presence of nearly identical DNA in each cell, functions as the primary basis of unique cellular identity [11]. More specifically, the diverse pathways cells take in their lifetimes are determined by the generalized epigenetic landscape [12].

1.2 Cell-specific regulation by epigenetic processes

The exquisite epigenetic regulation is accomplished in part through chromatin post-translational modifications (PTMs) evolved to demarcate, among a mosaic of functions, actively transcribed genes from the inactive ones [11]. Through the combined functions of epigenetic writer, reader, and eraser proteins (Figure 1.1), the combination of these modifications directly impacts gene accessibility and expression [13]. While the field is

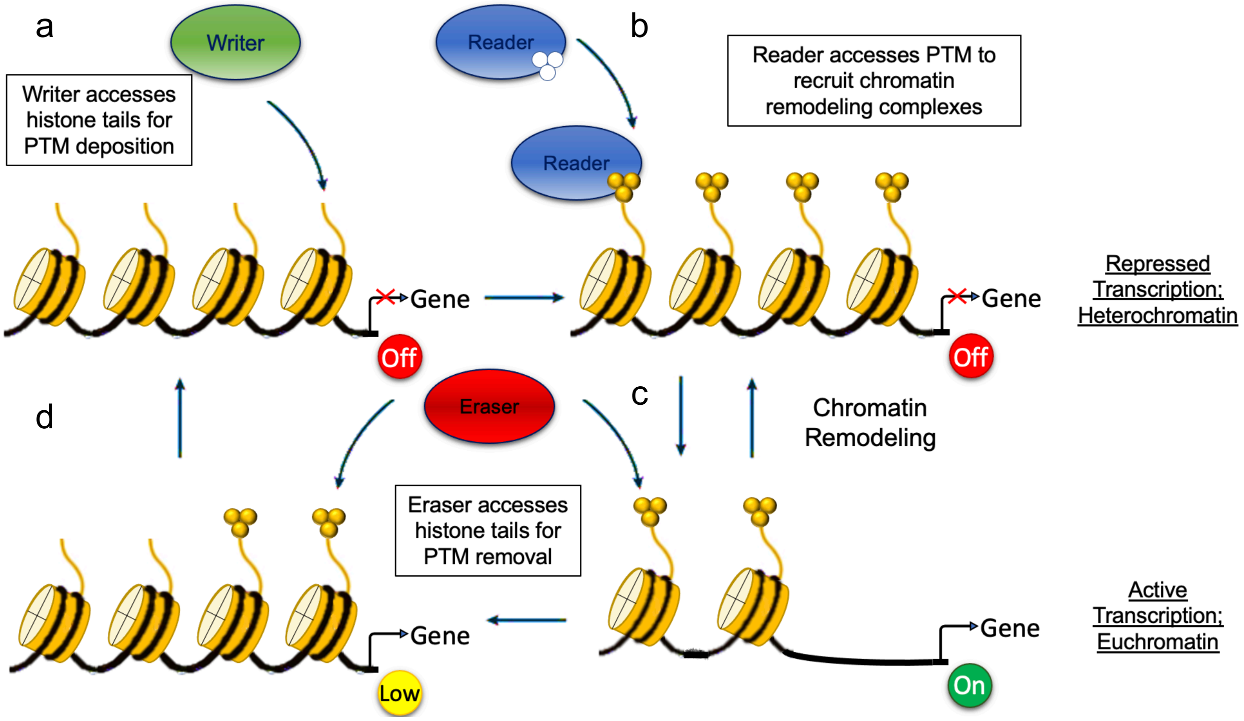

a
Writer
Writer accesses histone tails for PTM deposition
Reader
b
Reader accesses PTM to recruit chromatin remodeling complexes
Reader
Gene
Off
Gene
Off
Repressed Transcription; Heterochromatin
d
Eraser
c
Chromatin Remodeling
Eraser accesses histone tails for PTM removal
Gene
Low
Gene
On
Active Transcription; Euchromatin

Figure 1.1. Generalized epigenetic pathway for activating gene transcription. Clockwise from top left: **a**, Chromatin is constitutively repressed until a writer deposits a PTM; **b**, a reader accesses this PTM; **c**, recruiting chromatin remodeling complexes, allowing for gene transcription to take place. When the needs of the cell have been fulfilled, **d**, erasers remove the PTM and allow gene transcription to be tuned or turned off. Modified from [5]

constantly evolving and expanding through the discovery of new chromatin marks and the proteins responsible for the deposition, in the simplest terms this is achieved initially through the installation, or writing, of a modification on DNA or histone tails. “Writer” proteins, enzymes central to this process, include histone acetyltransferases (HATs), methyltransferases (HMTs), kinases, ubiquitin ligases, and DNA methyltransferase (DNMTs) [14-16], among others. The critical outcome of this modification depends on the recognition, or “reading”, of these marks by specific binding of domains including bromodomains (BRDs) [17] or chromodomains (chromodomain-helicase DNA binding, CHD) [18] and the chromatin remodeling complexes in which they reside [19]. Whether allowing for gene accessibility in euchromatin formation or gene silencing in repressive heterochromatin, the cycle completes with the removal, or “erasing”, of histone or DNA PTMs by deacetylases (HDACs) [20] or demethylases (HDMs) [21]. Given their role in transcriptional regulation, chromatin regulatory enzymes and proteins are actively investigated as therapeutic targets as their dysregulation often drives or sustains disease development. Here, we focus on a family of methyl-histone writer proteins shown to play a causal role in activating transcription.

Histone lysine methylation is a major PTM in eukaryotes. It occurs on the ε-amino group in three discrete states of mono-, di-, and tri-methylation through the methyl donation of *S*-adenosyl-L-methionine (SAM) [22-27]. Since the first histone lysine methyltransferase was discovered twenty years ago [28], over 60 histone methyltransferase enzymes have been putatively identified [29]. Histone methylation has been shown in both transcriptional activation and repression [30]. This duality exists as solvent-accessible histone tails contain a multitude of lysine residues capable of being methylated *in vivo* [30-32]. There are five highly studied, well-characterized lysine residues of histone H3 including H3K4, -9, -27, -36, and -79. They contribute to functions ranging from transcriptional regulation and chromatin dynamics to the DNA damage response [30, 33]. Unique amongst these is the H3K79-specific disruptor of telomere silencing 1-like (DOT1L) methyltransferase, which lacks a suppressor of variegation 3-9, enhancer of zeste, trithorax (SET) domain [34, 35]. Additionally, whereas DOT1L lacks a SET domain entirely, the positive regulatory domain I-binding factor 1 and retinoblastoma protein-interacting zinc finger gene 1 (PRD-BF1 and RIZ domains; PRDM) members contain an N-terminal PR domain that shares canonical SET domain methyltransferase fold with only around 20-30% SET sequence homology [36, 37].

Among well-characterized HKMTs, the highly conserved mixed lineage leukemia (MLL or KMT2) family of proteins is responsible for depositing the majority of histone 3 lysine 4 (H3K4) methylation in eukaryotes. Complexity of the H3K4 HMTs increases as eukaryotes evolved from single cell organisms to mammals, concomitant with increasing demands for spatial and temporal gene regulation. This is particularly evident where despite general conservation of the catalytic SET domain, each MLL/SET1 protein has non-redundant functions in development and is subject to distinct regulations [38]. Despite this distinct regulation, H3K4 methylation universally associated with transcriptional activation, being highly enriched at gene promoters and distal regulatory enhancers, and plays a pivotal role in the recruitment of basal transcription machinery [2, 39-41] and chromatin remodeling complexes [42-44] (Figure 1.2).

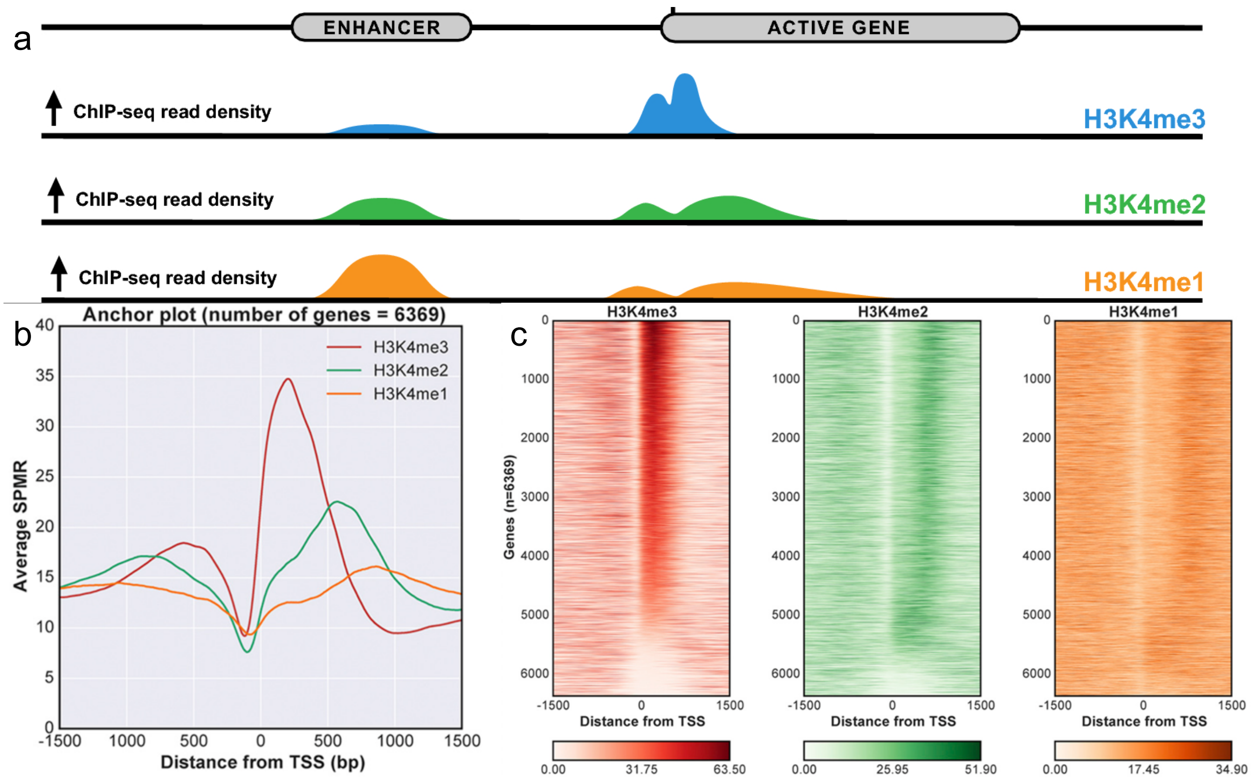


Figure 1.2. Distribution of H3K4 methylation in transcription. **a**, H3K4me3 exists primarily at the transcriptional start site (TSS) of actively transcribed genes whereas H3K4me1 is at upstream distal enhancer regions. **b**, overlaid anchor plot of distribution of H3K4me1/2/3 ± 1500 bp from TSS. **c**, heatmaps for detailed distribution amongst a wide array of genes showing a similar distribution as in **b**. Combined from [2, 3].

H3K4 methylation also promotes long-range chromatin interactions and higher order chromatin organization [45-47]. Dynamic interplay between H3K4me and co-transcriptional processes have also been reported [48, 49]. Human genetic studies have corroborated the functional importance of the MLL family enzymes: heterozygous mutations in MLLs are reported in congenital human Kabuki [50-55], Wiedemann-Steiner and Kleefstra spectrum syndromes [56-58]. Furthermore, MLL family proteins are among the most frequently mutated genes in human malignancies [38].

In budding yeast, *Saccharomyces cerevisiae*, ySET1 (Suppressor of Variegation 3-9, Enhancer of Zeste, Trithorax), an MLL homolog, is responsible for all H3K4 methylation [59-61]. Unlike their ancient homolog in *S. cerevisiae*, ySET1, the evolution of H3K4

HMTs in fruit flies to humans resulted in added complexity both in regulatory capacity and interactions. In *Drosophila melanogaster,* there are three MLL family enzymes: Trithorax (TRX), Trithorax-related (TRR), and drosophila SET1 (dSET1), that are responsible for global H3K4 methylation [38, 62]. Each of the three genes (i.e., *trx*, *trr* and *dset1*) are duplicated in mammals, giving rise to the six MLL/SET1 family members: MLL1 and MLL2 (KMT2A and 2B), MLL3 and MLL4 (KMT2C and 2D), and SET1A and SET1B (KMT2F and 2G), respectively.

1.3 MLLs make extensive domain-specific chromatin interactions

To fulfill the complex transcriptional regulatory demands of eukaryotes, the MLL/SET1 family enzymes contain multiple chromatin-interacting domains that are capable of recognizing specific patterns of histone and DNA modifications (Figure 1.3).

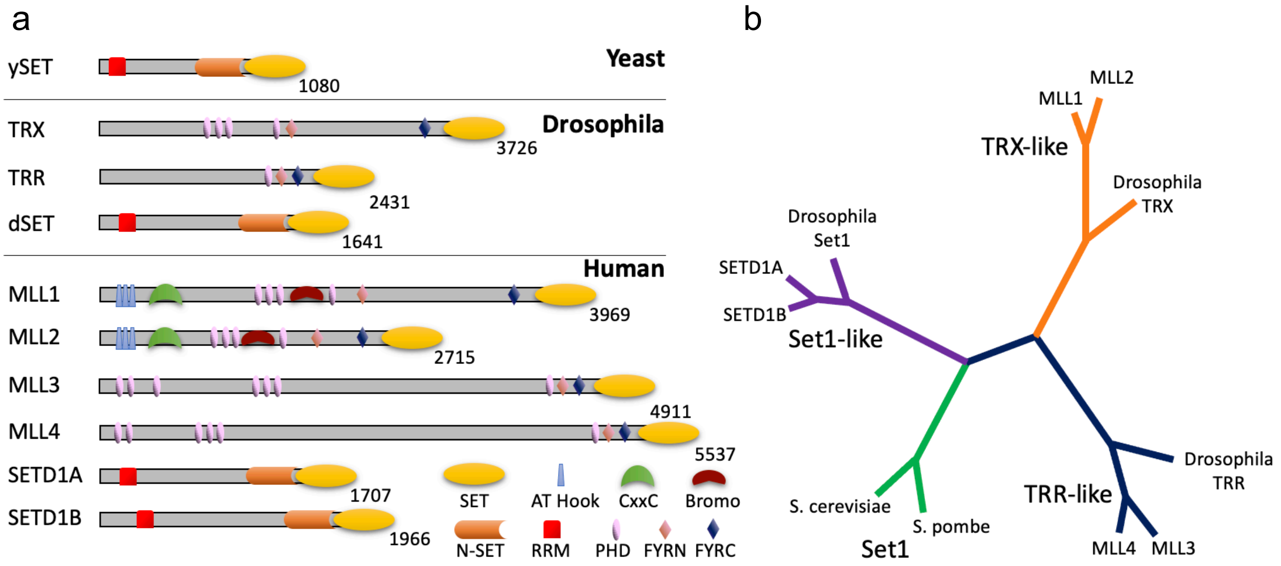


Figure 1.3. Yeast, Fly, and Human MLL family multi-domain distribution and phylogenetic tree. **a**, Evolutionarily derived domain complexity originating with ySET1 (top), eukaryotic transition and evolution to fly (middle), and duplication events resulting in human MLL family (bottom) **b**, Relatedness of yeast, fly, and human H3K4 HMTs. (Clockwise from left) human SETD1A/B from dSET1; human MLL1/2 from Drosophila TRX; MLL3/4 from Drosophila TRR; originating from progenitor SET1 of yeast. Inspired by [4]

These domain-specific interactions may contribute to loci-specific distribution of H3K4 methylation at transcriptionally-active gene promoters and distal regulatory enhancers [16] as well as to colocalization of H3K4 methylation with prominent co-transcriptional marks such as H3 acetylation, H3K79 and H3K36 methylation [16, 42, 63, 64]. Close correlation of H3K4 methylation has also been established with hypo-methylated DNA [65-68]. These interactions are central to our current understanding of MLL1-chromatin engagement and its functional interplay with transcriptional regulatory processes.

MLL1 CxxC domain engages unmethylated CpG islands

MLL1 contains a CxxC domain that is retained in the MLL1 fusion proteins after chromosome rearrangement. The MLL1 CxxC domain binds to unmethylated DNA [67, 69]. Structural studies show that the MLL1 CxxC domain makes rigid contacts with DNA nucleobases [70]. The methyl group on cytosine creates a steric clash in the CxxC binding pocket [70]. However, the CxxC mutant deficient in DNA binding does not affect MLL-AF9 (ALL1-fused gene from chromosome 9) binding to chromatin with high levels of DNA methylation, suggesting the CxxC domain is not a key contributor to overall binding affinity of MLL1 to chromatin. Instead, it acts to passively regulate DNA methylation by blocking access of DNA methyltransferases [68, 70]. In addition to interacting with DNA, the CxxC domain is also able to interact with the polymerase associating factor (PAF) elongation complex (PAF1C) in mammals [71-73]. This interaction involves a key arginine residue (R1153) of MLL1 that is not important for DNA binding [74]. The R1153 residue is not conserved in KMT2B (MLL2), which has alanine (A) in its place. When R1153 is mutated to A, as that of MLL2, it abolishes the interaction between MLL1 CxxC domain and PAF1C [71] and leads to reduced recruitment of MLL-AF9 to HOX targets and attenuation of the leukemic transformation [71-73]. These studies demonstrate that the CxxC domains in MLL1 and MLL2 have distinct functions, in part due to differential PAF1C interactions. The CxxC domain is not conserved in KMT2C, 2D, 2F and 2G. However, the function of CxxC domain is likely partially conserved through CxxC finger protein 1 (CFP1) [75, 76], a stable component of the SET1 complex. While CFP1 (and human SET1) do not interact with PAF1C [71], the CxxC domain of CFP1 is able to bind non-methylated CpG [77]. Distinct from that of MLL1, CFP1 is causally linked to *de novo* establishment of H3K4me3 at non-methylated CpG in mammalian cells [76, 78-80].

MLL1 PHDs are multifunctional epigenetic reader domains

MLL1 contains four plant homeodomains (PHD) and a bromodomain (BRD) immediately C-terminal to the CxxC domain. The PHD fingers are also present in other MLL family enzymes [38]. PHD fingers, together with the CxxC domain, are essential for recruitment of MLL1 to gene targets on chromatin [71, 73]. Specifically, PHD3 recognizes di- and tri-methylated H3K4 (H3K4me2/3) [71], contributing to spreading of H3K4 tri-methylation through the coupled 'writer-reader' regulation [16, 42]. Mutation of PHD3 attenuates chromatin recruitment of MLL1 and expression of MLL1 target genes [81]. Interestingly, PHD3 of MLL1 also interacts with Cyp33, which is required for histone deacetylase (HDAC)-dependent gene repression [82, 83]. Cyp33 contains a peptidyl prolyl isomerase (PPI) domain on the C-terminus [83, 84]. Cyp33 induces isomerization of the proline (P) 1629 in the MLL1 PHD3-bromodomain that allows PHD3 to directly interact with the RNA recognition motif (RRM) in Cyp33. Binding of PHD3 to H3K4me2/3 and Cyp33 RRM are mutually exclusive. Cyp33 overexpression dramatically decreases H3K4me3 at MLL1 target genes [84], enabling Cyp33 to act as a regulatory switch for gene regulation [84]. In MLL1-rearranged leukemia, PHD3 is not present in the fusion proteins. Loss of PHD3 and Cyp33-mediated repression potentially leads to a constitutively-activated leukemic program that leads to malignant transformation [85, 86]. Notably, Cyp33 overexpression also has inhibitory effects on leukemia without MLL1 rearrangement [85]. Cyp33 interacts specifically with the PHD3 domain of MLL1, but not that of MLL2, despite over 70% sequence homology between these two domains [83]. Functions of PHDs in other MLL family enzymes for recognition of H4K16 acetylation and protein degradation have also been reported [87, 88].

1.4 MLLs engage in complex protein-protein interaction networks

Beyond the domain-specific interactions that define the N-terminus of MLL1, there are also functional prerequisites required for MLL1 protein-protein interactions (PPIs) to occur. These extensive PPIs are necessary for MLL1 to exert its role in transcription [89, 90]. More rigidly, these complexes function in concert with an annotated core complex of proteins that directly interact with the MLL1 C-terminal catalytic domain [91-93].

MLL1 binds H3K36me2 via complexation with LEDGF and MENIN

MLL1 forms a tripartite complex with tumor suppressors MENIN (Multiple endocrine neoplasia 1) and LEDGF/p75 (Lens epithelium-derived growth factor) [94, 95]. Chromatin binding for each protein is mutually dependent. Deletion of MENIN or LEDGF/p75 significantly reduces MLL1 recruitment to the target genes (i.e., *HoxA9*, *p27*^{kip1} and *p18*^{ink4c}) [95-98]. Reciprocally, MLL1 also plays a critical role in supporting MENIN function *in vivo* [96]. Since LEDGF/p75 specifically binds H3K36me2 [99], it is able to recruit the MLL1 complex to genomic regions enriched for H3K36me2. In support of this, the histone H3K36 methyltransferase ASH1L co-localizes with MLL1 and LEDGF/p75 and is required for transcriptional activation [99, 100]. Similarly, H3K36 demethylase KDM2A promotes dissociation of MLL1 and LEDGF/p75 from chromatin and is functionally antagonistic to both MLL1 and ASH1L in leukemic transformation [99]. While MENIN is a stable component of the MLL2 complex [94, 101], it remains to be determined whether LEDGF/p75 is a *bona fide* component of the MLL2 complex. Furthermore, since MENIN and LEDGF/p75 interact with the oncogenic MLL1 fusion proteins [102], rationally targeting the MLL1-MENIN or LEDGF/p75-MLL1-MENIN interactions has shown great efficacy in blocking MLL1-rearranged leukemia [102-105].

MLL1 TAD allows for ternary complexation with c-Myb and CBP KIX

MLL1 has a conserved trans-activation domain (TAD) that interacts with CREB-binding protein (CBP) [106]. A solution structure of a ternary complex for the activation domain of transcription factor c-Myb, MLL1 TAD and CBP kinase-inducible domain-interacting domain (KIX) has been reported [107]. Binding of MLL1 TAD stabilizes the binary interaction between c-Myb and CBP through conformational changes in the disordered regions of the KIX domain [107]. MLL1 TAD binding also facilitates interactions between phosphorylated CREB and CBP [106]. The MLL1 TAD-mediated transactivation is largely suppressed by co-expression of adenovirus $E1A_{12S}$, a competitive inhibitor of CBP, or by MLL1 TAD mutants deficient in CBP-binding [106]. Interestingly, CBP seems to dictate MLL1 recruitment to either E2F1-mediated early-stage, pro-survival genes or late-stage, pro-apoptotic genes in a hepatocellular carcinoma mouse model [108]. The interaction between MLL1 and CBP is evolutionarily conserved. In *D. melanogaster*, TRX resides in a stable complex with dCBP, which cooperates with TRX in homeotic gene regulation [109]. Similarly, p300/CBP also interacts with the mammalian SET1 complex [40]. Tang and colleagues have elegantly demonstrated that both p300 and the SET1 complex are required for efficient p53-dependent transcription from a reconstituted chromatin template *in vitro* [40]. Although p300 is sufficient to initiate transcription on chromatin in a p53-dependent manner, recruitment of the SET1 complex by p300 enhances H3K4 methylation and further activates transcription in a p300 dose-dependent manner [40]. Knockdown of p300 by siRNA leads to global down regulation of H3K4me3 [40]. Beyond the intricate multivalent interactions through N-terminally situated domains that regulate methylation and downstream transcription, the highly conserved catalytic C-terminal SET domain depends on an entirely separate subset of complex interactions to regulate methylation, which is explored further within the contents of this document.

1.5 Early structural studies of MLL complex methyltransferase activity

MLL family enzymes reside in a conserved core complex

The MLL/SET1 family enzymes are large proteins with multiple functional domains as well as large stretches of disordered regions [110, 111]. While they share a highly conserved C-terminal SET domain that confers H3K4 methylation [112], they also have subclass specific domains such as the CxxC and bromodomain for KMT2A/2B, the PHD domains for KMT2A-D, and the RRM domain for KMT2F/G. Biochemical studies show that the catalytic SET domain of the MLL/SET1 enzymes has low intrinsic enzymatic activity [93, 112]. The SET domain activity can be drastically enhanced by interacting with a conserved core complex [93] (Figure 1.4a).

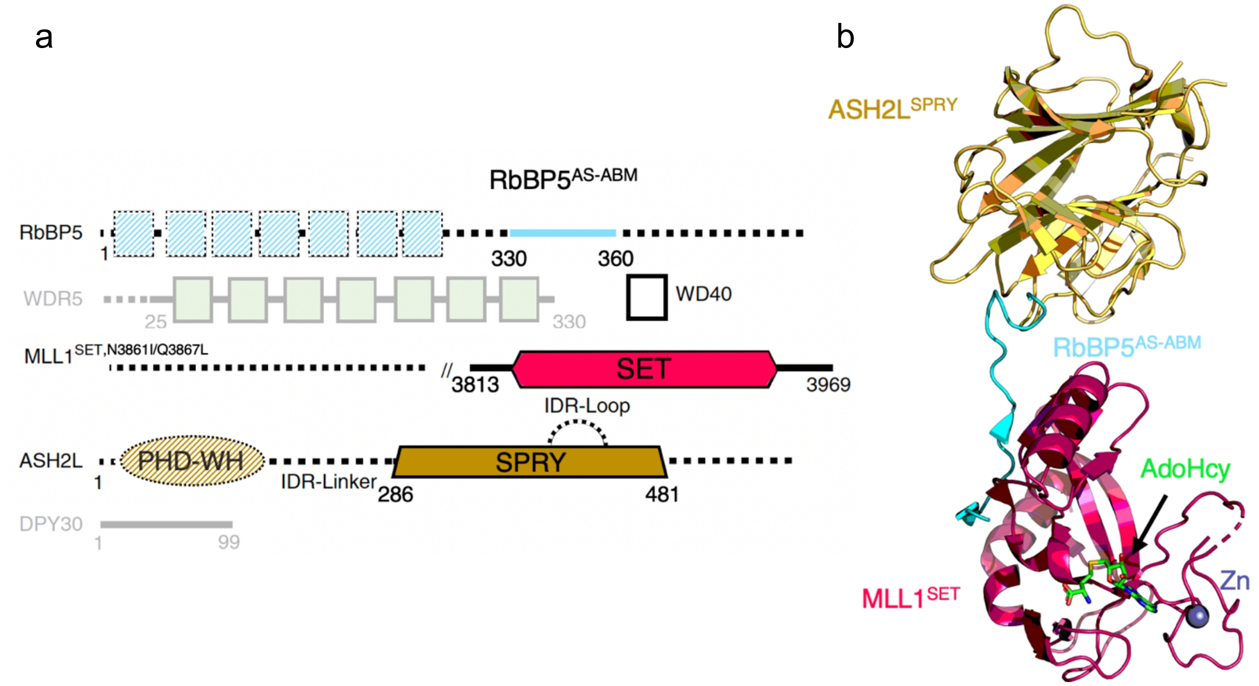


Figure 1.4. MLL^{SET} depends on WRAD proteins and their domains. **a**, RbBP5 (cyan), WDR5 (green), MLL1^{SET} (dark pink), ASH2L (yellow), DPY30 (black line). Faded or dashed regions are those omitted in the crystal structure. **b**, First crystal structure of MLL1 subcomplex MLL$^{SET,\ N3861I/Q3867L}$-ASH2L$^{SPRY,\Delta400\text{-}440}$RbBP5$^{AS\text{-}ABM}$ to highlight minimal structural domains from [1]. *S*-adenosyl-L-homocysteine (green) and Zinc (dark blue) also shown.

The core complex contains the highly conserved WDR5 (WD40 repeat-containing protein 5), RbBP5 (Retinoblastoma binding protein 5), ASH2L (Absent, small, homeotic disc 2-like), and DPY30 (DumPY protein 30) proteins (referred to as WARD) [38, 91, 93]. Among them, WDR5, RbBP5 and ASH2L, together with MLL1^{SET} are sufficient to reconstitute full activity of the MLL1 holo-complex on histone H3 [93, 113]. The core complex also acts as a platform for interacting with transcription factors, chromatin remodeling complexes and lncRNAs [38, 92, 114, 115], constituting a basic functional unit of the MLL/SET1 complexes that is essential in chromatin engagement.

Co-crystal structures of the MLL1/MLL3/ySET1 core complexes

Biochemical and structural studies have characterized inter-subunit interactions within the MLL/SET1 core complex [116-118]. Recent co-crystal structures of the MLL1, MLL3, and ySET1 complexes delineate detailed architectures of the core complex with or without substrates, i.e., *S*-adenosine-L-methionine (SAM) and histone H3 [1, 119]. Li and colleagues reported the first co-crystal structure of the MLL1 and 3 core complexes [1]. It shows that MLL3^{SET} makes extensive interactions with an acidic surface of the RbBP5-ASH2L heterodimer via a conserved SET-I arginine residue (Figure 1.4b) [1]. This interaction is stabilized by two hydrophobic residues in the SET-I, which are conserved in KMT2B-G. Interestingly, this interaction is not conserved for the MLL1^{SET} domain. Mutating MLL1 residues to MLL3-like sequences (N3861I/Q3867L, MLL1^{IL}) stabilizes RbBP5-ASH2L binding and circumvents WDR5 requirement for MLL1 activation [1]. Furthermore, RbBP5-ASH2L association with MLL3^{SET} and MLL1^{IL} reduces SET-I (SET-insertion between SET-C and SET-N) flexibility, allowing for stable substrate binding [1]. This study provides a structural basis for regulation of the MLL1/3 activities by core components [93, 113, 120] as well as the unique requirement of WDR5 in the MLL1 complex [121]. The SET-I is the least conserved region of MLL family SET regions and functions, alongside SET-C, through their respective paired acidic lobes to form a channel into which lysine 4 of the histone 3 peptide binds and orients it near the *S*-adenosyl-L-homocysteine binding site [92, 122]. Moreover, outside of the SET-I motif, divergent SET domain sequences confer distinct biochemical properties for the MLL/SET1 family HMTs, despite overall structural similarity [123-126]. The co-crystal structure of the ySET1 complex by Hsu and colleagues shows that ySET1 contains a unique glycine-centered motif (GI/NR)G(V/I/C/SS) that acts as a ‘hinge’ to control substrate access to the ySET1 catalytic site [119], rendering a naturally inactive state. Distinct primary sequences of the MLL3/4 SET domain also confer specific regulation of substrate state specificity [4, 119]. Targeting unique biochemical properties of individual MLL/SET1 complex has led to development of the MLL1-specific inhibitors that show good efficacy in cancer treatment as well as embryonic stem cell reprogramming [121, 127-129]. Such approaches can be envisioned to specifically target other MLL family enzymes as we learn more about their unique features in the future.

X-ray crystallography and histone-modifying enzyme complexes

As the gold standard in structural biology, X-ray crystallography is the tool of choice to resolving high resolution molecular details for apo- and small molecule-bound states of proteins [130]. It also remains the technique used by medicinal chemistry labs and the pharmaceutical industry due to the high efficiency crystallization screening process [131, 132]. Modern software advancements have allowed for more precise chemical docking, assisting in the process of drug development [133]. Further, the multitude of tools available for translating diffraction patterns to electron density maps and finally atomic models, has led to continuously expanding protein structure databases [134-137].

Indeed, the multitude of original structures of chromatin-modifying proteins were done using X-ray crystallography yielding new insights of recruitment to the nucleosome, interactions with histone tails and nucleosomal DNA interactions [138]. Early studies showed proteins like latency-associated nuclear antigen (LANA) exploited a now well-known motif of histone binding proteins known as an “arginine anchor” [139]. This arginine residue functions to “anchor” the histone core-interacting protein to the oppositely charged acidic patch of the nucleosome. Composed of residues in H2A (E56, E61, E64, D90, E91, and E92) and H2B (E105 and E113), the acidic patch has since been shown to readily bind several larger histone-binding proteins including the chromatin remodeler ISWI/SNF [140, 141]. While a novel early discovery, biochemical results already revealed a co-dependence for some complexes to recognize additional modifications on histone tails.

Recognition co-dependence is illustrated by Spt-Ada-Gcn5 acetyltransferase (SAGA) deubiquitylase (DUB) that binds and reads H2BK120ub whereas the catalytic domain contains a basic zinc finger module that interacts at the acidic patch interface [142]. Others, like those that function as ATP-dependent chromatin remodelers, including INO80 and CHD1, bind at the gyres of nucleosomal DNA [143-146]. However, complete structures resolved by cryo-EM provide a comprehensive picture of these complex nucleosomal interactions. From these early results, it is evident that new techniques are required for the modern study of histone-modifying enzyme complexes as the increase in macromolecular size and dynamics makes crystallization challenging [147]. More importantly, the intrinsically disordered regions (IDRs) in many of these proteins make for difficult X-ray crystallography targets. It is shown that 60% of lysine methyltransferases contain IDRs of 80 residues or more, with only 20% of other annotated proteins having IDRs of similar length [110]. Given the capacity for these flexible regions to partake in complex protein-protein interactions and facilitate phase transition and heterochromatin functions in cells, visualizing this complexity is essential in understanding cellular function [148-152]. Though still challenging, structural approaches like cryo-EM have recently gone through a major revolution allowing for the capture of large, dynamic macromolecular complexes. This reveals, for example, stabilization of stimulationresponsive motif (SRM) IDR of enhancer of zeste homolog 2 (EZH2) into a helical domain only in presence of an H3K27me3-substituted nucleosome as a requirement for polycomb repressive complex 2 (PRC2)-mediated H3K27me3 spreading [153]. Indeed, this transition becomes even more enticing when considering the advances in cryo-electron tomography (cryo-ET) for studying vitrified cells grown on grids using cryogenic-focused ion beam (cryo-FIB) milling [154] and for studying membrane proteins (Figure 1.5).

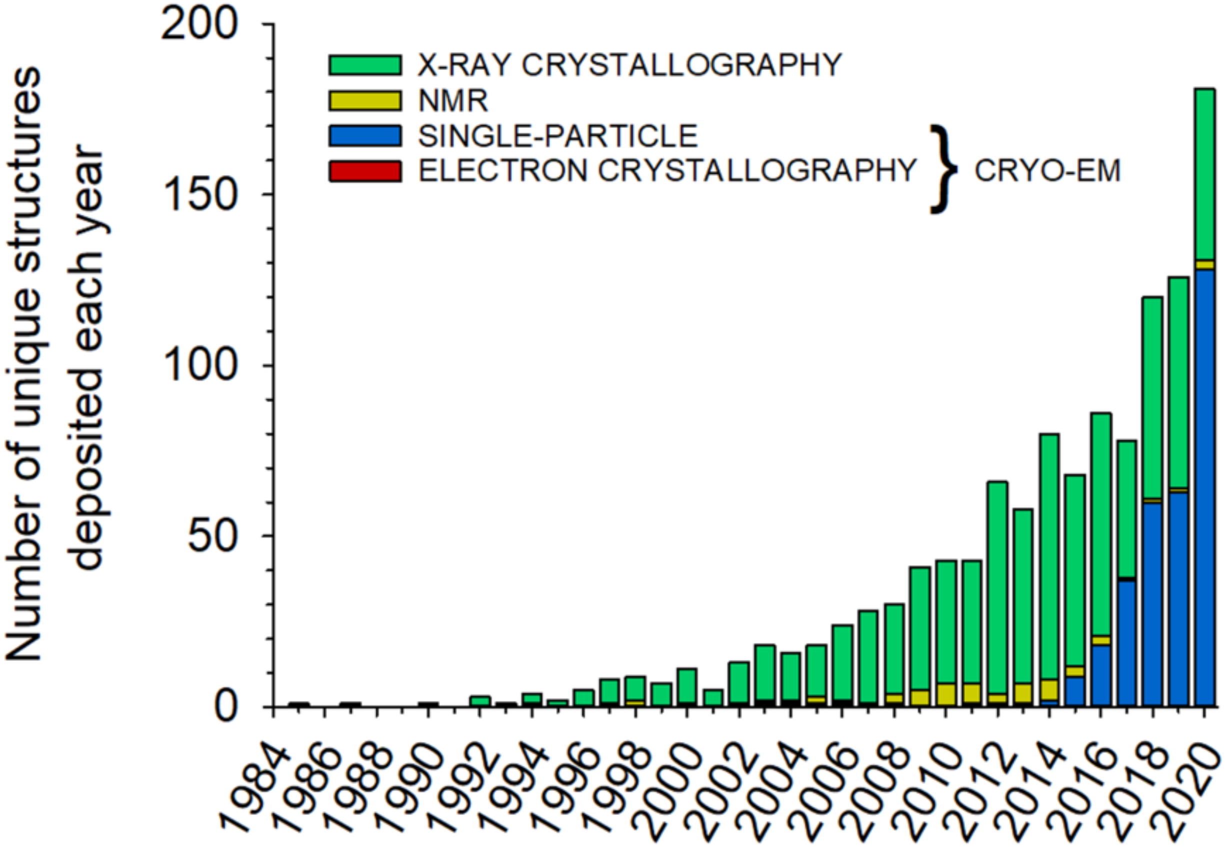


Figure 1.5. Single-particle cryo-EM and electron crystallography on the rise in the membrane protein field. Graph depicting the annual trend of unique membrane structures deposited in the PDB by traditional structural techniques. Graph reused with permission of the authors [6].

Cryo-EM studies in histone-modifying enzymes

Previous biochemical work has shown that the MLL1 core complex has much higher activity on a nucleosome core particle (NCP) substrate as compared to that on H3 peptide alone [155]. Further, histone-modifying enzymes proteins harbor extensive recognition domains and require complex interactions for functional regulation [156]. Therefore, transitioning to structural studies involving chromatin templates is a natural progression toward a better understanding of their regulation in cells. In 2016, a manuscript depicting p53 binding protein (53BP1) bound to an H4K20me2- and H2AK15ub-containing NCP was published [157]. Following this, a pair of cryo-EM manuscripts were published providing new molecular details for transcriptional process of RNA Polymerase II (RNAPII) paused at distinct histone-DNA landmarks and the intra-complex interactions stabilizing these pauses [158, 159]. These studies allow for the visualization of long elusive discrete steps of active transcription: Invasion of the nucleosome at SHL-6, peeling DNA at SHL-5, and pausing before the nucleosome dyad (SHL-1 and -2). After this, an H2A-H2B heterodimer loses DNA contact and is retained by the intramolecular interactions of the octameric core. This provides precise validation into prior biochemical studies. It also re-enforces consistent initial steps in nucleosome recognition other chromatin remodeling complexes exploit.

Relevant to this thesis is the study of H3K27 methyltransferase PRC2 simultaneously bound to a hetero-dinucleosome template [153], showing the intrinsic requirement for a singular pseudo-trimethylated H3K27 nucleosome attached via linker DNA to an unmodified one. The structure shows the requirement for an optimized linker both for consistent orientation of the complex and overall complex stability. For PRC2, Poepsel and colleagues show that a difference of even five basepairs (bp) in linker length (30 versus 35 versus 40) has dramatic implications. While 30 and 40 bp linkers allow for similar overall orientation, using 35 bp completely inverts the modified nucleosome, flipping it 180º. As a result, the PRC2 engagement points of 30 and 40 bp are similar aside from extended DNA exit points, yet due to the reorientation in the 35 bp linker construct, otherwise fuzzy densities in the 30 and 40 bp constructs are resolvable. Specifically, the SBD-SANT1 region of EZH2 is further stabilized through optimal contacts allowing for visualization [153].

These studies reveal why the cryo-EM field is ripe for answering questions involving epigenetic writer proteins and their complexes. Indeed, recent cryo-EM structures, including ours, reveal how the MLL1, MLL3 and ySET1 complexes bind to the nucleosome core particle (NCP) [123-126]. These studies shed light on distinct features of how these MLL complexes engage the H3 substrate in a more physiological context and, importantly, highlight divergent regulation of the MLL family enzymes on chromatin that may have implications for their respective regulation.

1.6 Dissertation Summary

The major goal of this dissertation is to dissect the mechanisms underlying DPY30-mediated activation of MLL1^{SET} on the nucleosome substrate. MLL1^{SET} functions through a conserved complex of protein partners to mono-, di- and tri-methylate histone H3 K4. Despite DPY30 having negligible effect on MLL1 activity on recombinant histone H3, somewhat paradoxically, loss of DPY30 has drastic effects on global H3K4me3 *in vivo*, in embryonic stem cells (ESCs) and hematopoietic stem cells [113, 160-164]. Further, no structural work for the MLL1 complex on the NCP had been reported, which suggests a gap in knowledge to be filled. Therefore, this system provides insight into both how MLL/SET family methyltransferases engage chromatin and how DPY30 regulates H3K4me3 on the nucleosome.

In chapter two, we use structural and biochemical techniques to show how MLL1 engages an unmodified nucleosome. Specifically, we show how MLL1 orients on the nucleosome through structures containing the DPY30 density. We reveal the dual anchoring motifs provided by ASH2L and RbBP5 that are required for efficient NCP catalysis. We finally determine if these MLL1-NCP interaction motifs affect MLL1 activation in an NCP-specific manner.

In chapter three, we extend our studies of DPY30 to reveal the unique mechanism behind DPY30-induced activation amongst MLL/SET family. We explore how the specific binding of DPY30 to ASH2L Sdc1/DPY30 interaction (SDI) motif affects ASH2L IDRs using directed NMR studies. Finally, we show the critical role of DPY30 in restricting the dynamics of the MLL1 complex on the NCP.

In chapter four, we use a catalytically inactive mutant containing H3 K-to-M at position four to capture discrete MLL1-NCP populations. We fully investigate the distinct binding modes motifs identified for the MLL1-NCP complex [124]. We specifically look at how rotational dynamics of the MLL1 complex regulates its activity on the NCP.

1.7 References

1. Li, Y., et al., *Structural basis for activity regulation of MLL family methyltransferases.* Nature, 2016. **530**(7591): p. 447-52.

2. Soares, L.M., et al., *Determinants of Histone H3K4 Methylation Patterns.* Molecular Cell, 2017. **68**(4): p. 773-785.e6.

3. Collins, B.E., et al., *Histone H3 lysine K4 methylation and its role in learning and memory.* Epigenetics & Chromatin, 2019. **12**(1): p. 7.

4. Zhang, Y., et al., *Evolving Catalytic Properties of the MLL Family SET Domain.* Structure, 2015. **23**(10): p. 1921-1933.

5. von Schaper, E., *Roche bets on bromodomains.* Nature Biotechnology, 2016. **34**(4): p. 361-362.

6. Le Bon, C., et al., *Amphipathic environments for determining the structure of membrane proteins by single-particle electron cryo-microscopy.* Quarterly Reviews of Biophysics, 2021. **54**: p. e6.

7. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.

8. Kornberg, R.D., *Chromatin Structure: A Repeating Unit of Histones and DNA.* Science, 1974. **184**(4139): p. 868-871.

9. Olins, A.L. and D.E. Olins, *Spheroid chromatin units (v bodies).* Science, 1974. **183**(4122): p. 330-332.

10. Kornberg, R.D. and Y. Lorch, *Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome.* Cell, 1999. **98**(3): p. 285-294.

11. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.

12. Goldberg, A.D., C.D. Allis, and E. Bernstein, *Epigenetics: a landscape takes shape.* Cell, 2007. **128**(4): p. 635-8.

13. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-45.

14. Kouzarides, T., *Chromatin Modifications and Their Function.* Cell, 2007. **128**(4): p. 693-705.

15. Allis, C.D. and T. Jenuwein, *The molecular hallmarks of epigenetic control.* Nat Rev Genet, 2016. **17**(8): p. 487-500.

16. Hyun, K., et al., *Writing, erasing and reading histone lysine methylations.* Exp Mol Med, 2017. **49**(4): p. e324.

17. Sanchez, R. and M.-M. Zhou, *The role of human bromodomains in chromatin biology and gene transcription.* Current opinion in drug discovery & development, 2009. **12**(5): p. 659-665.

18. Marfella, C.G.A. and A.N. Imbalzano, *The Chd family of chromatin remodelers.* Mutation research, 2007. **618**(1-2): p. 30-40.

19. Clapier, C.R. and B.R. Cairns, *The Biology of Chromatin Remodeling Complexes.* Annual Review of Biochemistry, 2009. **78**(1): p. 273-304.

20. Glozak, M.A. and E. Seto, *Histone deacetylases and cancer.* Oncogene, 2007. **26**(37): p. 5420-5432.

21. Kooistra, S.M. and K. Helin, *Molecular mechanisms and potential functions of histone demethylases.* Nature Reviews Molecular Cell Biology, 2012. **13**(5): p. 297-311.

22. Murray, K., *THE OCCURRENCE OF EPSILON-N-METHYL LYSINE IN HISTONES.* Biochemistry, 1964. **3**: p. 10-5.

23. Paik, W.K. and S. Kim, *Enzymatic methylation of protein fractions from calf thymus nuclei.* Biochemical and Biophysical Research Communications, 1967. **29**(1): p. 14-20.

24. Paik, W.K. and S. Kim, *Enzymatic methylation of histones.* Archives of Biochemistry and Biophysics, 1969. **134**(2): p. 632-637.

25. Kim, S. and W.K. Paik, *Studies on the Origin of ε-N-Methyl-l-lysine in Protein.* Journal of Biological Chemistry, 1965. **240**(12): p. 4629-4634.

26. Paik, W.K. and S. Kim, *Protein Methylation: Enzymatic methylation of proteins after translation may take part in control of biological activities of proteins.* Science, 1971. **174**(4005): p. 114-119.

27. Bannister, A.J. and T. Kouzarides, *Histone methylation: recognizing the methyl mark.* Methods in enzymology, 2003. **376**: p. 269-288.

28. Rea, S., et al., *Regulation of chromatin structure by site-specific histone H3 methyltransferases.* Nature, 2000. **406**(6796): p. 593-9.

29. Binda, O., *On your histone mark, SET, methylate!* Epigenetics, 2013. **8**(5): p. 457-463.

30. Margueron, R., P. Trojer, and D. Reinberg, *The key to development: interpreting the histone code?* Current Opinion in Genetics & Development, 2005. **15**(2): p. 163-176.

31. Zhang, L., et al., *Identification of novel histone post-translational modifications by peptide mass fingerprinting.* Chromosoma, 2003. **112**(2): p. 77-86.

32. Zhang, K., et al., *A mass spectrometric “Western blot” to evaluate the correlations between histone methylation and histone acetylation.* Proteomics, 2004. **4**(12): p. 3765-3775.

33. Martin, C. and Y. Zhang, *The diverse functions of histone lysine methylation.* Nature reviews Molecular cell biology, 2005. **6**(11): p. 838-849.

34. Jenuwein, T., *The epigenetic magic of histone lysine methylation: delivered on 6 July 2005 at the 30th FEBS Congress in Budapest, Hungary.* The FEBS journal, 2006. **273**(14): p. 3121-3135.

35. van Leeuwen, F., P.R. Gafken, and D.E. Gottschling, *Dot1p modulates silencing in yeast by methylation of the nucleosome core.* Cell, 2002. **109**(6): p. 745-56.

36. Hohenauer, T. and A.W. Moore, *The Prdm family: expanding roles in stem cells and development.* Development, 2012. **139**(13): p. 2267-82.

37. Huang, S., G. Shao, and L. Liu, *The PR domain of the Rb-binding zinc finger protein RIZ1 is a protein binding interface and is related to the SET domain functioning in chromatin-mediated gene expression.* Journal of Biological Chemistry, 1998. **273**(26): p. 15933-15939.

38. Rao, R.C. and Y. Dou, *Hijacked in cancer: the KMT2 (MLL) family of methyltransferases.* Nat Rev Cancer, 2015. **15**(6): p. 334-46.

39. Vermeulen, M., et al., *Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4.* Cell, 2007. **131**(1): p. 58-69.

40. Tang, Z., et al., *SET1 and p300 act synergistically, through coupled histone modifications, in transcriptional activation by p53.* Cell, 2013. **154**(2): p. 297-310.

41. Lauberth, S.M., et al., *H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation.* Cell, 2013. **152**(5): p. 1021-36.

42. Ruthenburg, A.J., et al., *Multivalent engagement of chromatin modifications by linked binding modules.* Nat Rev Mol Cell Biol, 2007. **8**(12): p. 983-94.

43. Wysocka, J., et al., *A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling.* Nature, 2006. **442**(7098): p. 86-90.

44. Taverna, S.D., et al., *How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers.* Nat Struct Mol Biol, 2007. **14**(11): p. 1025-1040.

45. Phillips, J.E. and V.G. Corces, *CTCF: master weaver of the genome.* Cell, 2009. **137**(7): p. 1194-211.

46. Tang, Z., et al., *CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription.* Cell, 2015. **163**(7): p. 1611-27.

47. Yan, J., et al., *Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers.* Cell Res, 2018. **28**(2): p. 204-220.

48. Sims, R.J., 3rd, et al., *Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing.* Molecular cell, 2007. **28**(4): p. 665-676.

49. Khan, D.H., et al., *Dynamic Histone Acetylation of H3K4me3 Nucleosome Regulates MCL1 Pre-mRNA Splicing.* Journal of Cellular Physiology, 2016. **231**(10): p. 2196-2204.

50. Ng, S.B., et al., *Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.* Nature genetics, 2010. **42**(9): p. 790-793.

51. Paulussen, A.D., et al., *MLL2 mutation spectrum in 45 patients with Kabuki syndrome.* Hum Mutat, 2011. **32**(2): p. E2018-25.

52. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature, 2011. **472**(7341): p. 120-4.

53. Micale, L., et al., *Mutation spectrum of MLL2 in a cohort of Kabuki syndrome patients.* Orphanet journal of rare diseases, 2011. **6**: p. 38-38.

54. Hannibal, M.C., et al., *Spectrum of MLL2 (ALR) mutations in 110 cases of Kabuki syndrome.* Am J Med Genet A, 2011. **155a**(7): p. 1511-6.

55. Kluijt, I., et al., *Kabuki syndrome–report of six cases and review of the literature with emphasis on ocular features.* Ophthalmic genetics, 2000. **21**(1): p. 51-61.

56. Jones, W.D., et al., *De novo mutations in MLL cause Wiedemann-Steiner syndrome.* The American Journal of Human Genetics, 2012. **91**(2): p. 358-364.

57. Mendelsohn, B.A., et al., *Advanced bone age in a girl with Wiedemann-Steiner syndrome and an exonic deletion in KMT2A (MLL).* Am J Med Genet A, 2014. **164a**(8): p. 2079-83.

58. Strom, S.P., et al., *De Novo variants in the KMT2A (MLL) gene causing atypical Wiedemann-Steiner syndrome in two unrelated individuals identified by clinical exome sequencing.* BMC medical genetics, 2014. **15**: p. 49-49.

59. Roguev, A., et al., *The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4.* The EMBO Journal, 2001. **20**(24): p. 7137-7148.

60. Briggs, S.D., et al., *Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae.* Genes Dev, 2001. **15**(24): p. 3286-95.

61. Nagy, P.L., et al., *A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3.* Proc Natl Acad Sci U S A, 2002. **99**(1): p. 90-4.

62. Morgan, M.A. and A. Shilatifard, *Drosophila sets its sights on cancer: Trr/MLL3/4 COMPASS-like complexes in development and disease.* Molecular and Cellular Biology, 2013. **33**(9): p. 1698-1701.

63. Ha, M., et al., *Coordinated histone modifications are associated with gene expression variation within and between species.* Genome Res, 2011. **21**(4): p. 590-8.

64. Barski, A., et al., *High-Resolution Profiling of Histone Methylations in the Human Genome.* Cell, 2007. **129**(4): p. 823-837.

65. Rose, N.R. and R.J. Klose, *Understanding the relationship between DNA methylation and histone lysine methylation.* Biochim Biophys Acta, 2014. **1839**(12): p. 1362-72.

66. Deaton, A.M. and A. Bird, *CpG islands and the regulation of transcription.* Genes Dev, 2011. **25**(10): p. 1010-22.

67. Birke, M., *The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation.* Nucleic Acids Research, 2002. **30**: p. 958-965.

68. Erfurth, F.E., et al., *MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression.* Proc Natl Acad Sci U S A, 2008. **105**(21): p. 7517-22.

69. Ayton, P.M., E.H. Chen, and M.L. Cleary, *Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein.* Mol Cell Biol, 2004. **24**(23): p. 10470-8.

70. Cierpicki, T., et al., *Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia.* Nat Struct Mol Biol, 2010. **17**(1): p. 62-8.

71. Milne, T.A., et al., *Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis.* Mol Cell, 2010. **38**(6): p. 853-63.

72. Speck, N.A. and C.R. Vakoc, *PAF is in the cabal of MLL1-interacting proteins that promote leukemia.* Cancer Cell, 2010. **17**(6): p. 531-2.

73. Muntean, A.G., et al., *The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis.* Cancer Cell, 2010. **17**(6): p. 609-21.

74. Bach, C., et al., *Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2.* Oncogene, 2009. **28**(6): p. 815-823.

75. Clouaire, T., et al., *Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells.* Genes & development, 2012. **26**: p. 1714-28.

76. Brown, D.A., et al., *The SET1 Complex Selects Actively Transcribed Target Genes via Multivalent Interaction with CpG Island Chromatin.* Cell Rep, 2017. **20**(10): p. 2313-2327.

77. Xu, C., et al., *The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain.* Nat Commun, 2011. **2**: p. 227.

78. Thomson, J.P., et al., *CpG islands influence chromatin structure via the CpG-binding protein Cfp1.* Nature, 2010. **464**(7291): p. 1082-6.

79. Clouaire, T., S. Webb, and A. Bird, *Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells.* Genome Biol, 2014. **15**(9): p. 451.

80. Yu, C., et al., *CFP1 Regulates Histone H3K4 Trimethylation and Developmental Potential in Mouse Oocytes.* Cell Rep, 2017. **20**(5): p. 1161-1172.

81. Chang, P.Y., et al., *Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL-dependent gene transcription.* J Mol Biol, 2010. **400**(2): p. 137-44.

82. Xia, Z.B., et al., *MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8342-7.

83. Wang, Z., et al., *Pro Isomerization in MLL1 PHD3-Bromo Cassette Connects H3K4me Readout to CyP33 and HDAC-Mediated Repression.* Cell, 2010. **141**(7): p. 1183-1194.

84. Park, S., et al., *The PHD3 domain of MLL acts as a CYP33-regulated switch between MLL-mediated activation and repression.* Biochemistry, 2010. **49**(31): p. 6576-86.

85. Fair, K., et al., *Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells.* Mol Cell Biol, 2001. **21**(10): p. 3589-97.

86. Muntean, A.G., et al., *The PHD fingers of MLL block MLL fusion protein-mediated transformation.* Blood, 2008. **112**(12): p. 4690-3.

87. Zhang, Y., et al., *Selective binding of the PHD6 finger of MLL4 to histone H4K16ac links MLL4 and MOF.* Nature Communications, 2019. **10**(1): p. 2314.

88. Wang, J., et al., *A subset of mixed lineage leukemia proteins has plant homeodomain (PHD)-mediated E3 ligase activity.* Journal of Biological Chemistry, 2012. **287**(52): p. 43410-43416.

89. Eidahl, J.O., et al., *Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes.* Nucleic Acids Res, 2013. **41**(6): p. 3924-36.

90. Slany, R.K., *The molecular mechanics of mixed lineage leukemia.* Oncogene, 2016. **35**(40): p. 5215-5223.

91. Cho, Y.W., et al., *PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex.* J Biol Chem, 2007. **282**(28): p. 20395-406.

92. Cosgrove, M.S. and A. Patel, *Mixed lineage leukemia: a structure-function perspective of the MLL1 protein.* FEBS J, 2010. **277**(8): p. 1832-42.

93. Dou, Y., et al., *Regulation of MLL1 H3K4 methyltransferase activity by its core components.* Nat Struct Mol Biol, 2006. **13**(8): p. 713-9.

94. Hughes, C.M., et al., *Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus.* Mol Cell, 2004. **13**(4): p. 587-97.

95. Yokoyama, A. and M.L. Cleary, *Menin critically links MLL proteins with LEDGF on cancer-associated target genes.* Cancer Cell, 2008. **14**(1): p. 36-46.

96. Milne, T.A., et al., *Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors.* Proc Natl Acad Sci U S A, 2005. **102**(3): p. 749-54.

97. Thiel, A.T., et al., *MLL-AF9-induced leukemogenesis requires coexpression of the wild-type Mll allele.* Cancer Cell, 2010. **17**(2): p. 148-59.

98. Chen, Y.X., et al., *The tumor suppressor menin regulates hematopoiesis and myeloid transformation by influencing Hox gene expression.* Proc Natl Acad Sci U S A, 2006. **103**(4): p. 1018-23.

99. Zhu, L., et al., *ASH1L Links Histone H3 Lysine 36 Dimethylation to MLL Leukemia.* Cancer Discov, 2016. **6**(7): p. 770-83.

100. Jones, M., et al., *Ash1l controls quiescence and self-renewal potential in hematopoietic stem cells.* J Clin Invest, 2015. **125**(5): p. 2007-20.

101. van Nuland, R., et al., *Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes.* Mol Cell Biol, 2013. **33**(10): p. 2067-77.

102. Yokoyama, A., et al., *The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis.* Cell, 2005. **123**(2): p. 207-18.

103. Caslini, C., et al., *Interaction of MLL amino terminal sequences with menin is required for transformation.* Cancer Res, 2007. **67**(15): p. 7275-83.

104. El Ashkar, S., et al., *LEDGF/p75 is dispensable for hematopoiesis but essential for MLL-rearranged leukemogenesis.* Blood, 2018. **131**(1): p. 95-107.

105. Borkin, D., et al., *Pharmacologic inhibition of the Menin-MLL interaction blocks progression of MLL leukemia in vivo.* Cancer Cell, 2015. **27**(4): p. 589-602.

106. Ernst, P., et al., *MLL and CREB Bind Cooperatively to the Nuclear Coactivator CREB-Binding Protein.* Molecular and Cellular Biology, 2001. **21**(7): p. 2249.

107. De Guzman, R.N., et al., *Structural Basis for Cooperative Transcription Factor Binding to the CBP Coactivator.* Journal of Molecular Biology, 2006. **355**(5): p. 1005-1013.

108. Swarnalatha, M., A.K. Singh, and V. Kumar, *Promoter occupancy of MLL1 histone methyltransferase seems to specify the proliferative and apoptotic functions of E2F1 in a tumour microenvironment.* J Cell Sci, 2013. **126**(Pt 20): p. 4636-46.

109. Petruk, S., et al., *Trithorax and dCBP acting in a complex to maintain expression of a homeotic gene.* Science, 2001. **294**(5545): p. 1331-4.

110. Lazar, T., et al., *Intrinsic protein disorder in histone lysine methylation.* Biology Direct, 2016. **11**(1): p. 30.

111. Szabó, B., et al., *Disordered Regions of Mixed Lineage Leukemia 4 (MLL4) Protein Are Capable of RNA Binding.* International Journal of Molecular Sciences, 2018. **19**(11): p. 3478.

112. Milne, T.A., et al., *MLL Targets SET Domain Methyltransferase Activity to Hox Gene Promoters.* Molecular Cell, 2002. **10**(5): p. 1107-1117.

113. Patel, A., et al., *On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex.* J Biol Chem, 2009. **284**(36): p. 24242-56.

114. Dou, Y., et al., *Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF.* Cell, 2005. **121**(6): p. 873-85.

115. Yang, Y.W., et al., *Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency.* Elife, 2014. **3**: p. e02046.

116. Avdic, V., et al., *Structural and biochemical insights into MLL1 core complex assembly.* Structure, 2011. **19**(1): p. 101-8.

117. Odho, Z., S.M. Southall, and J.R. Wilson, *Characterization of a Novel WDR5-binding Site That Recruits RbBP5 through a Conserved Motif to Enhance Methylation of Histone H3 Lysine 4 by Mixed Lineage Leukemia Protein-1.* Journal of Biological Chemistry, 2010. **285**(43): p. 32967-32976.

118. Patel, A., et al., *A Conserved Arginine-containing Motif Crucial for the Assembly and Enzymatic Activity of the Mixed Lineage Leukemia Protein-1 Core Complex.* Journal of Biological Chemistry, 2008. **283**(47): p. 32162-32175.

119. Hsu, P.L., et al., *Crystal Structure of the COMPASS H3K4 Methyltransferase Catalytic Module.* Cell, 2018. **174**(5): p. 1106-1116 e9.

120. Cao, F., et al., *An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain.* PLoS One, 2010. **5**(11): p. e14102.

121. Cao, F., et al., *Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia.* Mol Cell, 2014. **53**(2): p. 247-61.

122. Southall, S.M., et al., *Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks.* Mol Cell, 2009. **33**(2): p. 181-91.

123. Hsu, P.L., et al., *Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS.* Molecular Cell, 2019. **76**(5): p. 712-723.e4.

124. Park, S.H., et al., *Cryo-EM structure of the human MLL1 core complex bound to the nucleosome.* Nature Communications, 2019. **10**(1): p. 5540.

125. Worden, E.J., X. Zhang, and C. Wolberger, *Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome.* eLife, 2020. **9**: p. e53199.

126. Xue, H., et al., *Structural basis of nucleosome recognition and modification by MLL methyltransferases.* Nature, 2019. **573**(7774): p. 445-449.

127. Grebien, F., et al., *Pharmacological targeting of the Wdr5-MLL interaction in C/EBPα N-terminal leukemia.* Nature chemical biology, 2015. **11**(8): p. 571-578.

128. Zhang, H., et al., *MLL1 Inhibition Reprograms Epiblast Stem Cells to Naive Pluripotency.* Cell Stem Cell, 2016. **18**(4): p. 481-94.

129. Aho, E.R., et al., *Displacement of WDR5 from chromatin by a WIN site inhibitor with picomolar affinity.* Cell reports, 2019. **26**(11): p. 2916-2928. e13.

130. van Montfort, R.L.M. and P. Workman, *Structure-based drug design: aiming for a perfect fit.* Essays in biochemistry, 2017. **61**(5): p. 431-437.

131. Stevens, R.C., *High-throughput protein crystallization.* Curr Opin Struct Biol, 2000. **10**(5): p. 558-63.

132. McPherson, A. and J.A. Gavira, *Introduction to protein crystallization.* Acta crystallographica. Section F, Structural biology communications, 2014. **70**(Pt 1): p. 2-20.

133. Halgren, T.A., et al., *Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening.* J Med Chem, 2004. **47**(7): p. 1750-9.

134. Skubák, P. and N.S. Pannu, *Automatic protein structure solution from weak X-ray data.* Nature communications, 2013. **4**: p. 2777-2777.

135. Afonine, P.V., et al., *Real-space refinement in PHENIX for cryo-EM and crystallography.* Acta Crystallogr D Struct Biol, 2018. **74**(Pt 6): p. 531-544.

136. Liebschner, D., et al., *Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix.* Acta Crystallogr D Struct Biol, 2019. **75**(Pt 10): p. 861-877.

137. UniProt, C., *UniProt: a worldwide hub of protein knowledge.* Nucleic acids research, 2019. **47**(D1): p. D506-D515.

138. Zhou, K., G. Gaullier, and K. Luger, *Nucleosome structure and dynamics are coming of age.* Nature Structural & Molecular Biology, 2019. **26**(1): p. 3-13.

139. Barbera, A.J., et al., *The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA.* Science, 2006. **311**(5762): p. 856-61.

140. Dann, G.P., et al., *ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference.* Nature, 2017. **548**(7669): p. 607-611.

141. Dao, H.T., et al., *A basic motif anchoring ISWI to nucleosome acidic patch regulates nucleosome spacing.* Nature chemical biology, 2020. **16**(2): p. 134-142.

142. Morgan, M.T., et al., *Structural basis for histone H2B deubiquitination by the SAGA DUB module.* Science, 2016. **351**(6274): p. 725-728.

143. Eustermann, S., et al., *Structural basis for ATP-dependent chromatin remodelling by the INO80 complex.* Nature, 2018. **556**(7701): p. 386-390.

144. Ayala, R., et al., *Structure and regulation of the human INO80–nucleosome complex.* Nature, 2018. **556**(7701): p. 391-395.

145. Farnung, L., et al., *Nucleosome-Chd1 structure and implications for chromatin remodelling.* Nature, 2017. **550**(7677): p. 539-542.

146. Sundaramoorthy, R., et al., *Structural reorganization of the chromatin remodeling enzyme Chd1 upon engagement with nucleosomes.* Elife, 2017. **6**: p. e22510.

147. Callaway, E., *Revolutionary cryo-EM is taking over structural biology.* Nature, 2020. **578**(7794): p. 201.

148. Oldfield, C.J. and A.K. Dunker, *Intrinsically disordered proteins and intrinsically disordered protein regions.* Annu Rev Biochem, 2014. **83**: p. 553-84.

149. Wright, P.E. and H.J. Dyson, *Intrinsically disordered proteins in cellular signalling and regulation.* Nat Rev Mol Cell Biol, 2015. **16**(1): p. 18-29.

150. van der Lee, R., et al., *Classification of intrinsically disordered regions and proteins.* Chemical reviews, 2014. **114**(13): p. 6589-6631.

151. Gibson, B.A., et al., *Organization of Chromatin by Intrinsic and Regulated Phase Separation.* Cell, 2019. **179**(2): p. 470-484 e21.

152. Musselman, C.A. and T.G. Kutateladze, *Characterization of functional disordered regions within chromatin-associated proteins.* iScience, 2021. **24**(2): p. 102070.

153. Poepsel, S., V. Kasinath, and E. Nogales, *Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes.* Nat Struct Mol Biol, 2018. **25**(2): p. 154-162.

154. Wagner, F.R., et al., *Preparing samples from whole cells using focused-ion-beam milling for cryo-electron tomography.* Nat Protoc, 2020. **15**(6): p. 2041-2070.

155. Patel, A., et al., *A novel non-set domain multi-subunit methyltransferase required for sequential nucleosomal histone H3 methylation by the mixed lineage leukemia protein-1 (MLL1) core complex.* Journal of Biological Chemistry, 2010.

156. Smith, E. and A. Shilatifard, *The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes.* Molecular cell, 2010. **40**(5): p. 689-701.

157. Wilson, M.D., et al., *The structural basis of modified nucleosome recognition by 53BP1.* Nature, 2016. **536**(7614): p. 100-103.

158. Kujirai, T., et al., *Structural basis of the nucleosome transition during RNA polymerase II passage.* Science, 2018. **362**(6414): p. 595-598.

159. Farnung, L., S.M. Vos, and P. Cramer, *Structure of transcribing RNA polymerase II-nucleosome complex.* Nature Communications, 2018. **9**(1): p. 5432.

160. Haddad, J.F., et al., *Structural Analysis of the Ash2L/Dpy-30 Complex Reveals a Heterogeneity in H3K4 Methylation.* Structure, 2018.

161. Shinsky, S.A. and M.S. Cosgrove, *Unique Role of the WD-40 Repeat Protein 5 (WDR5) Subunit within the Mixed Lineage Leukemia 3 (MLL3) Histone Methyltransferase Complex.* J Biol Chem, 2015. **290**(43): p. 25819-33.

162. Jiang, H., et al., *Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains.* Cell, 2011. **144**(4): p. 513-25.

163. Yang, Z., et al., *The DPY30 subunit in SET1/MLL complexes regulates the proliferation and differentiation of hematopoietic progenitor cells.* Blood, 2014. **124**(13): p. 2025-33.

164. Yang, Z., et al., *Dpy30 is critical for maintaining the identity and function of adult hematopoietic stem cells.* The Journal of experimental medicine, 2016. **213**(11): p. 2349-2364.

CHAPTER 2.

Cryo-EM Structure of the Human Mixed Lineage Leukemia-1 Complex[2]

2.1 Abstract

Mixed Lineage Leukemia (MLL) family histone methyltransferases are the key enzymes that deposit histone H3 lysine 4 (K4) mono-, di-, and tri-methylation and regulate gene expression in mammals. Despite extensive structural and biochemical studies, the molecular mechanism by which the MLL complexes recognize histone H3K4 within the nucleosome core particle (NCP) remains unclear. Here, we report the single-particle cryo-electron microscopy (cryo-EM) structure of the human MLL1 core complex bound to the NCP. The MLL1 core complex anchors on the NCP through RbBP5 and ASH2L, which interacts extensively with nucleosomal DNA as well as the surface close to histone H4 N-terminal tail. Concurrent interactions of RbBP5 and ASH2L with the NCP uniquely align the catalytic MLL1^{SET} domain at the nucleosome dyad, allowing symmetrical access to both H3K4 substrates within the NCP. Our study sheds light on how the MLL1 complex engages chromatin and how chromatin binding promotes MLL1 tri-methylation activity.

[2] The contents of this chapter were adapted from a published second authored manuscript: Park, S. H., **Ayoub, A.**, Lee, Y. T., Xu, J., Kim, H., Zheng, W., Zhang, B., Sha, L., An, S., Zhang, Y., Cianfrocco, M. A., Su, M., Dou, Y., and Cho, U. S. “Cryo-EM structure of the human MLL1 core complex bound to the nucleosome,” *Nature Communications,* **10**, 5540 (2019)

2.2 Introduction

The nucleosome core particle (NCP), consisting of an octameric core of histone proteins (two of each H2A, H2B, H3, and H4) and 146 basepairs of genomic DNA, represents the first level of eukaryotic DNA packaging [4]. It is further organized into higher order chromatin structures. Cell specific transcriptional programming, in large part, is governed by chromatin accessibility, which is actively regulated by histone modifying enzymes and ATP-dependent chromatin remodeling complexes. In recent years, X-ray crystallography and single-particle cryo-EM studies have shed light on how these chromatin-associating complexes interact with the NCP for respective physiological functions. Most, if not all, chromatin complexes engage the 'acidic-patch' region of the NCP through variations of an 'arginine anchor' motif [8-11], highlighting common features among chromatin interacting protein complexes. It remains unclear whether the recognition mode of the NCP is universal for chromatin interacting complexes.

Among histone post-translational modifications, the states of histone H3 lysine 4 methylation (H3K4me) (i.e., mono-, di-, tri-methylation) are exquisitely modulated at important DNA regulatory regions including active gene promoters, gene bodies and distal regulatory enhancers [14]. In particular, H3K4me3 is highly correlated to transcriptionally active- and open-chromatin regions [15, 16] and is shown to actively recruit the basic transcription machinery, ATP-dependent chromatin remodeling complexes, and histone acetyltransferases [15, 17-20]. In contrast, H3K4me1 is a prevalent mark often found at poised or active distal enhancers [22]. Specific regulation of the H3K4me states may play a critical role in important physiological processes in cells.

The mixed lineage leukemia (MLL) family enzymes, including MLL1-4/KMT2A-2D, SET1A/KMT2F and SET1B/KMT2G), are the major histone lysine 4 (K4) methyltransferases in mammals. They contain an evolutionarily conserved catalytic suppressor of variegation 3-9, enhancer of zeste, trithorax (SET) domain [23]. Biochemical studies, by us and others, have shown that the SET domain stably interacts with four highly conserved proteins, i.e. RbBP5 (retinoblastoma-binding protein 5), ASH2L (Absent, small, homeotic disks-2-like), WDR5 (WD40 repeat-containing protein 5) and DPY30 (DumPY protein 30) [24, 25]. The MLL1 core components are able to increases the MLL1^{SET} activity on mono- and di-methylation of histone H3K4 by ~600 fold [26]. The molecular mechanism by which MLL1 core components stimulate MLL1^{SET} activity has been elegantly demonstrated by a series of structural studies including the human MLL1/3^{SET}-ASH2L$^{SPRY,\Delta 400\text{-}440}$-RbBP5$^{330\text{-}375}$ subcomplex [1], the homologous yeast SET1 complexes [7, 21] as well as individual mammalian core components [14, 27, 28]. However, it is still unclear how tri-methylation activity of MLL1, which has been widely reported *in vivo* [29-32], is regulated. To date, the structures of the MLL family enzymes are determined with either no H3 or H3 peptide as the substrate. It remains unclear how the MLL1 complex binds and catalyzes H3K4 methylation on the NCP and more importantly, how MLL1 activity, especially tri-methylation activity, is regulated on chromatin.

Mutations of MLL proteins have been widely reported in a variety of congenital human syndromes including Kabuki [33-38], Wiedemann-Steiner [39-41], and Kleefstra spectrum syndromes [41] as well as a wide spectrum of human malignancies [14]. Similarly, aberrant expression and recurrent mis-sense mutations of ASH2L and RbBP5 have also been identified in human cancers of different origins [42-45]. It is important to understand how MLL activity on chromatin is modulated during disease development.

Here, we report the single-particle cryo-EM structure of the human MLL1 core complex bound to the NCP. It not only reveals the overall architecture of the human MLL1 complex with full-length core components, but also illustrates how the MLL1 core complex engages chromatin. Importantly, we show that the MLL1 core complex docks on the NCP through concurrent interactions of ASH2L/RbBP5 with nucleosomal DNA and histone H4. This unique configuration aligns the catalytic MLL1^{SET} domain at the nucleosome dyad, which allows the symmetric access to both H3K4 substrates. Our structure sheds new light on how the MLL1 complex binds to chromatin and how its activity for H3K4me3 is regulated.

2.3 Results

Architecture of the human MLL1 core complex bound to the NCP

Recombinant human MLL1 core complex (MLL1^{RWSAD}) containing RbBP5 (residues 1–538), WDR5 (residues 25–330), MLL1^{SET} (residues 3762–3969), ASH2L (residues 1–534) and DPY30 (residues 1–99) was reconstituted *in vitro* (Figures 2.1a and Figures S2.1a and S2.1b in Appendix A). It had drastically enhanced activity for H3K4me2 and H3K4me3 when the NCP was used as the substrate (Figure 2.1b).

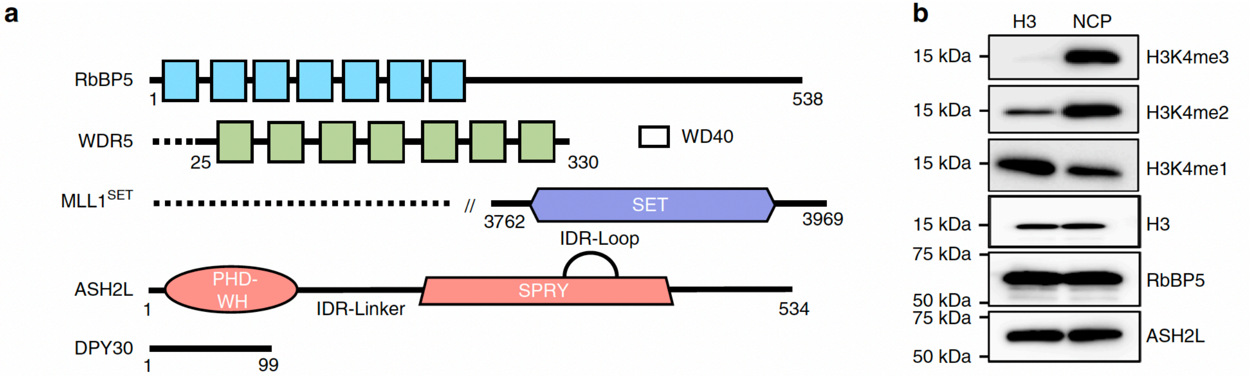


Figure 2.1. MLL1^{RWSAD} complex domains and its activation on H3 versus NCP. **a**, Schematic domain architectures for the core components of the human MLL1 complex used in this cryo-EM study. Dashed lines are truncations made. **b**, Immunoblot to detect H3K4 methylation in the *in vitro* histone methyltransferase assay. Antibodies used are indicated on the right. Substrates used were recombinant histone H3 (lane 1) and NCP (lane 2). Immunoblots for unmodified H3, RbBP5 and ASH2L included as controls.

To understand the underlying mechanism, we determined the single-particle cryo-EM structure of reconstituted MLL1^{RWSAD} bound to the recombinant NCP. The cryo-EM structure of MLL1^{RWSAD}-NCP was determined at a 6.2 Å resolution (Figures 2.2a and Figures S2.2, S2.3a and Table 2.1 in Appendix A). The composite map of MLL1^{RWSAD}-NCP was generated after local filtering to the estimated resolution to avoid over-interpretation (Figures 2.2a and S2.2 in Appendix A). In parallel, the cryo-EM maps of RbBP5-NCP and RbBP5-WDR5-MLL1^{SET} (MLL1^{RWS})-NCP subcomplexes were derived from the MLL1^{RWSAD}-NCP dataset, and reconstructed at 4.2 Å and 4.5 Å resolution, respectively (Figures S2.2 and S2.3b-d in Appendix A). The model structure of the MLL1^{RWSAD}-NCP complex (Figure 2.2b) was built by rigid-body fitting and real space refinement using previously published crystal structures of mouse RbBP5 (PDB ID: 5OV3) [12], human WDR5 (PDB ID: 2H14) [3], human MLL1^{SET}-ASH2L^{SPRY}-RbBP5$^{330\text{-}375}$ (PDB ID: 5F6L) [1], a DPY30 dimer (PDB ID: 6E2H) [46], and the 601-NCP (PDB ID: 3MVD) [9].

The overall architecture of the MLL1^{RWSAD}-NCP complex showed that MLL1^{RWSAD} anchors at the edge of the NCP through two core components, RbBP5 and ASH2L, simultaneously (Figure 2.2b). In the NCP, DNA superhelical location 7 (SHL7) and SHL1.5, together with H4 N-terminal tail, were involved in the interaction with MLL1^{RWSAD} (Figure 2.2b). Notably, domains in the MLL1^{RWSAD}-NCP complex were dynamically associated with each other and showed multiple conformations (Figure S2.3e in Appendix A). However, the overall architecture was conserved with respect to all sub-classes of the MLL1^{RWSAD}-NCP structures (Figure S2.3e in Appendix A). Distinct from many of previously reported NCP-recognizing protein or protein complexes [47], the MLL1 core complex did not interact with the acidic patch region of the NCP.

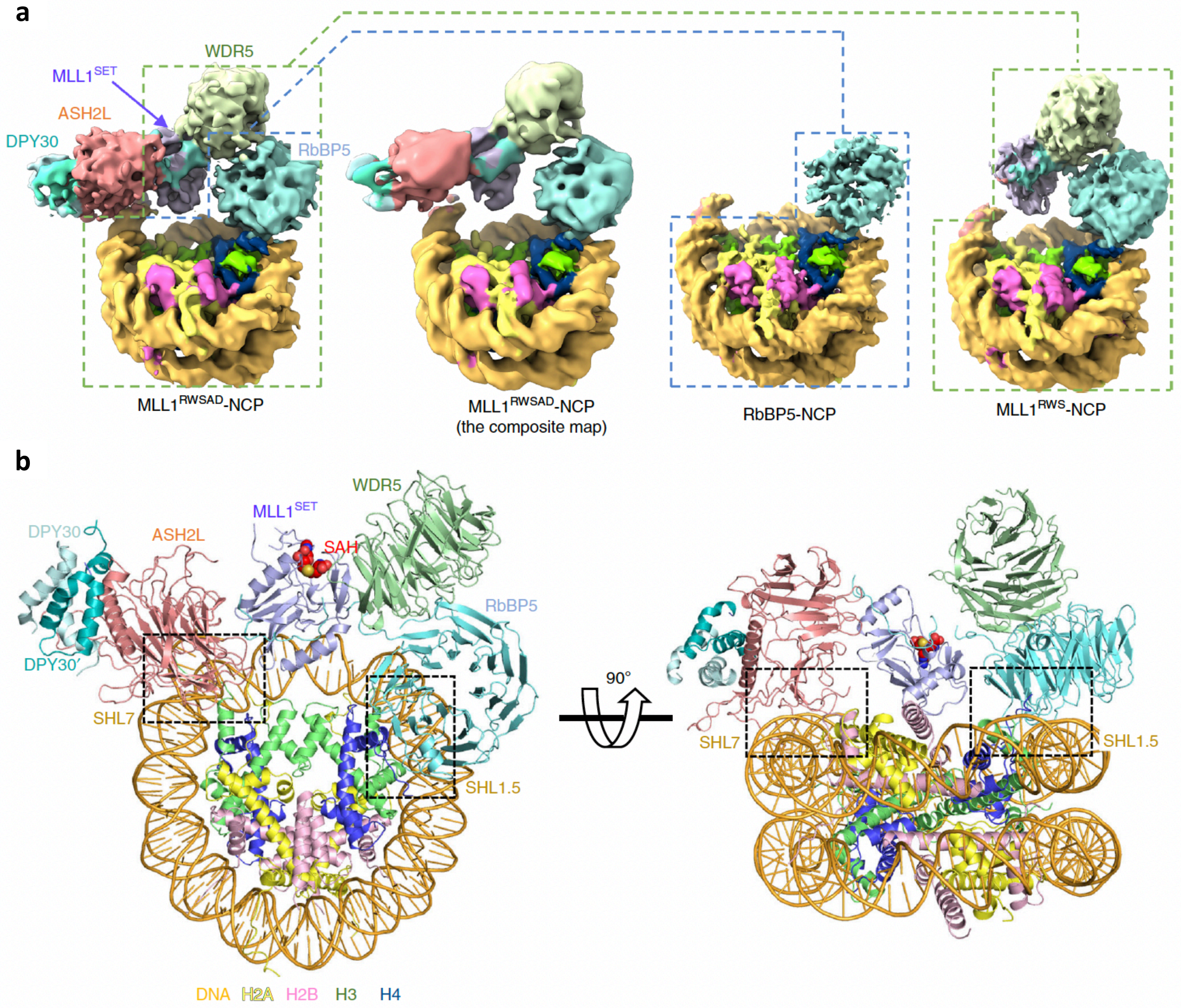


Figure 2.2. Cryo-EM 3D reconstruction of the MLL1^{RWSAD}-NCP complex. **a**, The composite map of MLL1^{RWSAD}-NCP was locally filtered to the estimated resolution. The subcomplexes, i.e., RbBP5-NCP and MLL1^{RWS}-NCP, are shown in dashed boxes. **b**, Top (left) and front (right) views of the MLL1^{RWSAD}-NCP structure. The *S*-adenosyl-L-homocysteine (SAH) was represented as a sphere (red) and the MLL1 core components shown in cartoon representation (RbBP5: cyan, WDR5: green, MLL1^{SET}: slate, ASH2L: pink, and DPY30 dimer: cerulean and teal). Widom 601 DNA and four histones were colored as indicated on the bottom. Two blacked, dashed squares highlighted the nucleosome contact points near SHL1.5 and SHL7 by MLL1^{RWSAD}. Illustrations of the protein structure and cryo-EM maps used in all figures were generated with PyMOL (Delano Scientific, LLC) and Chimera [6]/ChimeraX [13] by USC and SHP.

RbBP5 binds the NCP through both DNA and histone H4 tail recognition

In the MLL1^{RWSAD}-NCP complex, the RbBP5-NCP interfaces were less dynamic. The sub-population particles of RbBP5-NCP from the MLL1^{RWSAD}-NCP dataset were resolved at 4.2 Å resolution (Figures 2.3a, S2.2, and S2.3b in Appendix A). The regions of mouse RbBP5 in model fitting shared 100% sequence identity with human RbBP5 (Figure S2.4a in Appendix A). The structure showed that RbBP5 bound to the NCP by simultaneously engaging DNA (SHL1.5) and histone H4 N-terminal tail. The interactions involved six consecutive loops emanating from the WD40 repeats of RbBP5 (Figure 2.3a). Characteristic features of RbBP5 (e.g., unique helix, anchoring loop, and insertion loop) were well matched into the cryo-EM map of RbBP5-NCP subcomplex (Figure S2.4b in Appendix A). Notably, RbBP5 interacted with DNA SHL1.5 through four positively-charged arginine residues (Quad-R) located at β18–β19 (R220), β20–β21 (R251), β22–β23 (R272), and β24–β25 (R294) loops, respectively (Figure 2.3b). The Quad-R participated in electrostatic interactions with the DNA phosphate backbone (Figure 2.3b).

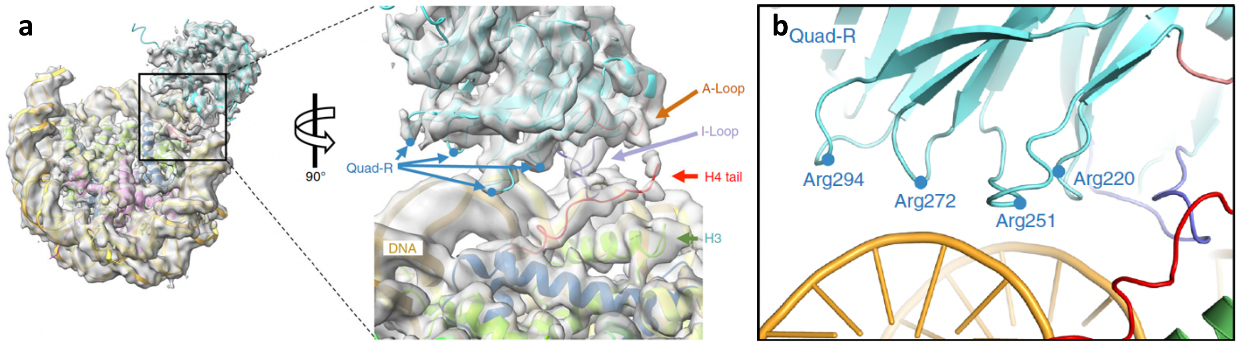


Figure 2.3. RbBP5 interaction with the NCP. **a**, The cryo-EM structure of the RbBP5-NCP subcomplex (4.2 Å). The interaction interface was enlarged and shown on the right. Insertion (I)-loop, Anchoring (A)-loop, and Quad-R of RbBP5, as well as the H4 tail highlighted in purple, orange, blue, and red, respectively. Histone H3 shown in green. **b**, Interaction of Quad-R, as indicated, with DNA backbone. Red line represented H4 tail. Figure prepared by USC and SHP

Disruption of RbBP5-NCP interaction significantly reduced the activity of the MLL1 core complex. Mutations of the Quad-R residues to alanine (A) led to reduction of H3K4me3 and to a lesser degree, H3K4me2 (Figure 2.4a). The effect was more drastic when Quad-R residues were mutated to glutamic acid (E) (Figure 2.4b). Systematic alteration of three, two or one arginine residue(s) in Quad-R showed that at least two arginine residues were required for optimal H3K4me3 activity (Figure 2.4a).

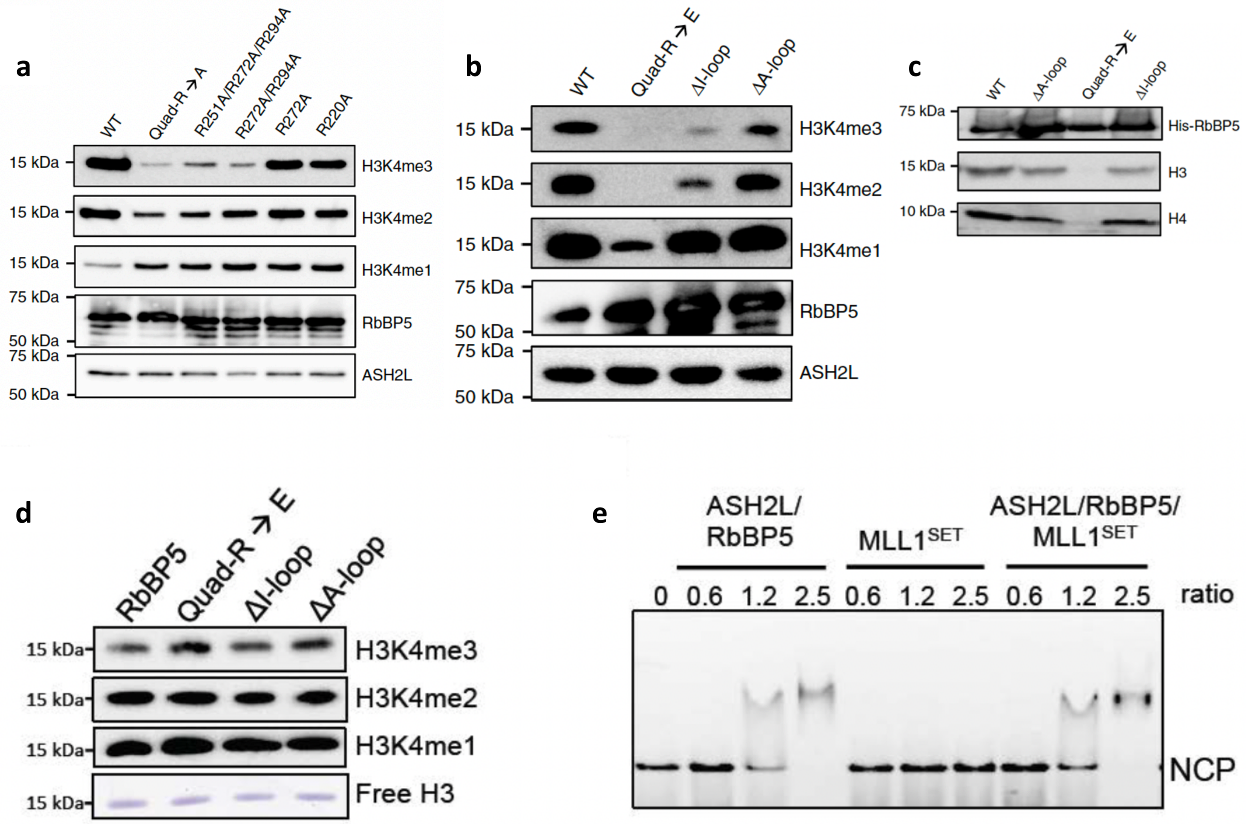


Figure 2.4. RbBP5 requirement for binding to and methylation of NCP. **a**, Immunoblot to detect *in vitro* histone methyltransferase activity with NCP as the substrate. The MLL1^{RWSAD} complex was reconstituted with wild-type and Quad-R mutated RbBP5 indicated on top; **b**, The MLL1^{RWSAD} complex was reconstituted with wild-type and deletion mutant proteins as shown on top. **c**, *In vitro* pulldown assay for RbBP5 and the NCP. Ni-NTA-bound fractions were shown and His-tagged wild-type or mutant RbBP5 proteins shown on top. Immunoblot for H3 used to detect the NCP in the bound fractions and for H4 as a control. **d**, Immunoblot to detect *in vitro* histone methyltransferase activity with H3 as the substrate. Reconstituted MLL1^{RWSAD} complexes containing wild-type and mutant RbBP5 were used, as indicated on top. Coomassie stain of H3 was included as a control. **e**, Electrophoretic mobility shift assay for the NCP with or without binding of protein or protein complexes as indicated on top. Molar ratio of protein(s) to the NCP was indicated on top. Westerns in **a** and **c** completed by JX; EMSA in **e** completed by YTL

The second RbBP5-NCP interface includes two loops, an insertion loop (β16–β17 loop, referred to herein as I-loop) and an anchoring loop (β19–β20 loop, referred to herein as A-loop), of RbBP5 (Figures 2.3a, 2.5a, 2.5c and 2.5d). Both I- and A-loops are evolutionarily conserved in higher eukaryotes (Figure S2.4b in Appendix A). The I-loop was positioned between the N-terminal tail of histone H4 and nucleosomal DNA (Figure 2.5a). The A-loop ran parallel to H4 tail, which was positioned between the I-/A-loops of RbBP5 and the helix α1 (Leu65–Asp77) of histone H3 (Figure 2.5a). This H4 tail-mediated nucleosome recognition of RbBP5 resembles that of the active-state DOT1L [48, 49] (Figure 2.5b). Similar to Quad-R, deletion of I-loop and to a lesser degree A-loop, reduced the activity of the MLL1 core complex for H3K4me3 and H3K4me2 (Figure 2.4b). Importantly, RbBP5-NCP interaction is specifically required for MLL1 activity on the NCP

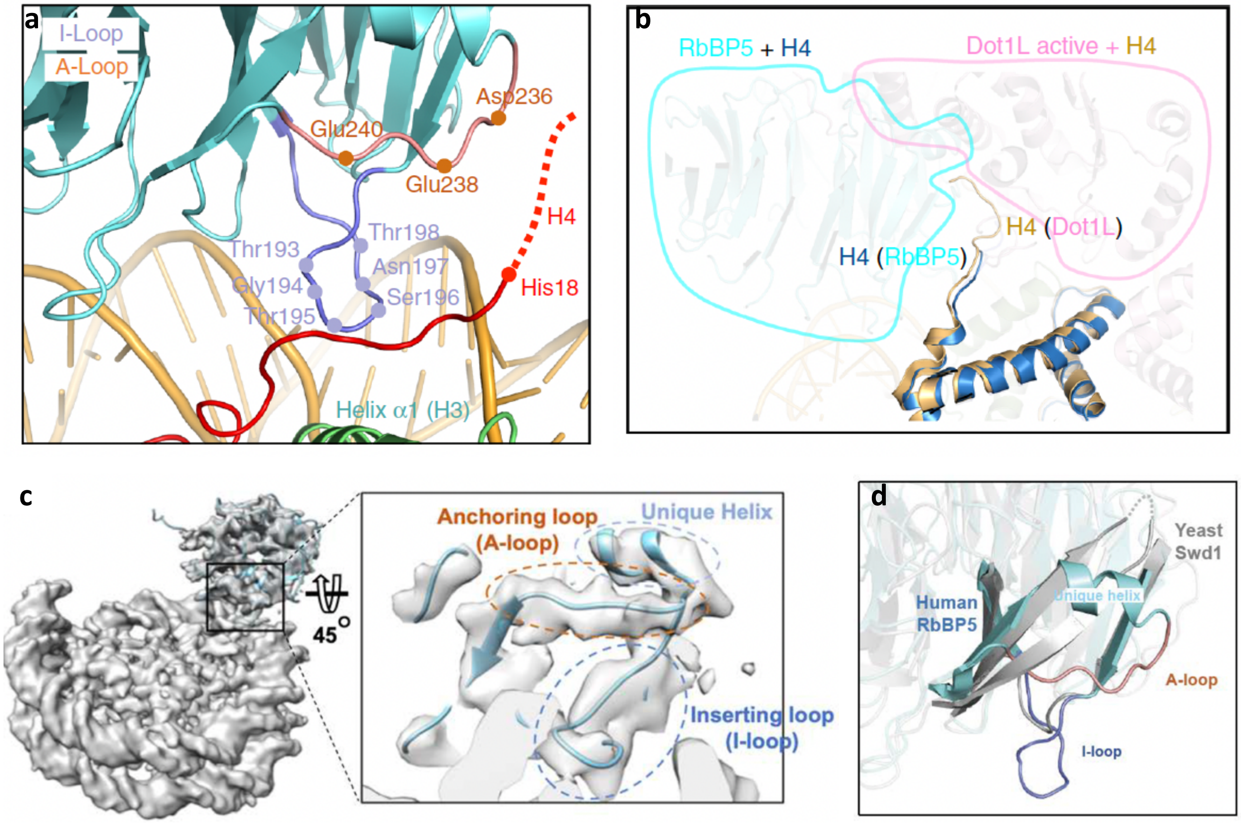
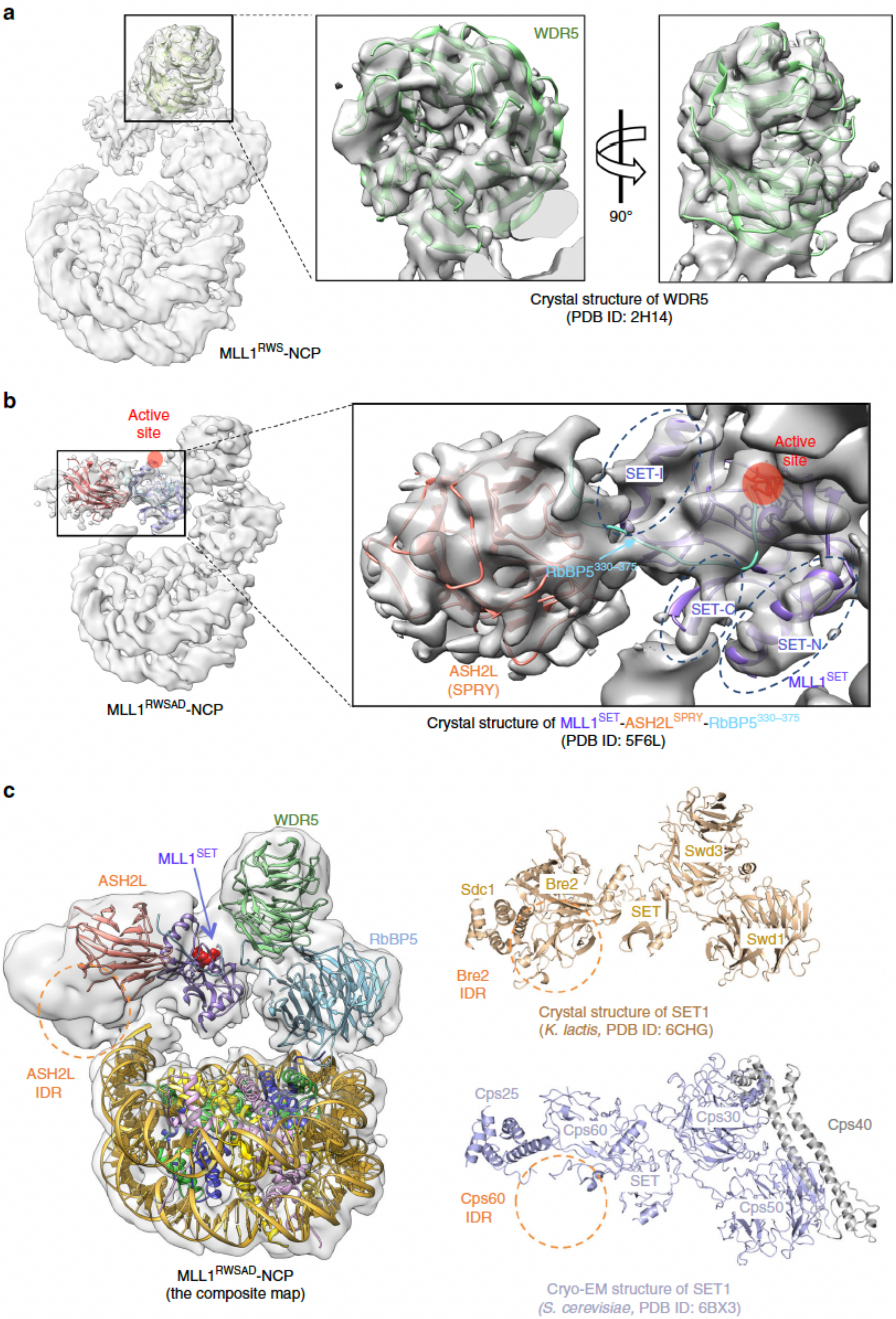


Figure 2.5. RbBP5 interactions with core histones. **a**, The interface between RbBP5 and the H4 tail. Key residues on RbBP5 I-/A-loops indicated. The H4 tail (His18 to core) represented by a red line and the extended tail beyond His18 represented by a dashed line. **b**, Structural superposition of the H4 tails upon RbBP5 (cyan) and DOT1L (PDB ID: 6NJ9) [2] binding. The RbBP5 and DOT1L at the interface enclosed by the blue and pink outlines, respectively. **c**, Rigid-body fitting of RbBP5 (PDB ID: 5OV3) [12] in the RbBP5-NCP cryo-EM map. Characteristic I- and A-loops as well as the unique helix (zoomed in, top) of RbBP5 fit nicely into the cryo-EM map. **d**, Structural overlay of human RbBP5 and yeast Swd1 (PDB ID: 6CHG) [21]. RbBP5 (cyan) and Swd1 (gray) had distinct features for I- and A-loops. Figure prepared by USC and SHP

(Figure 2.4b). Among RbBP5-NCP interactions, Quad-R was the main contributor to NCP binding. The mutation of Quad-R significantly reduced RbBP5 binding to the NCP, while the deletion of I- and A-loops only modestly affected NCP binding (Figure 2.4c). Mutations in Quad-R, I-loop, and A-loop had no effects on mono-, di-, tri-methylation of free H3 (Figure 2.4d).

Structure of a WDR5, MLL1^{SET}, and ASH2L^{SPRY} subcomplex

To resolve the structural organization of WDR5, MLL1^{SET}, and ASH2L^{SPRY} sub-complex, we reconstructed the MLL1^{RWS}-NCP subcomplex (Figure 2.2a and Figure S2.2 in Appendix A) and successfully docked the crystal structures of human WDR5 (PDB ID: 2H14) [3] and MLL1^{SET}-ASH2L^{SPRY}-RbBP5$^{330-375}$ (PDB ID: 5F6L) [1] into the cryo-EM maps (Figure 2.6a and 2.6b). The secondary structural components of MLL1^{SET} [50], including α-helices and β-hairpin of the SET-I, SET-N, and SET-C domains (dotted circles), fit well into the cryo-EM map (Figure 2.6b). Similar to MLL1^{SET}, distinctive features of WDR5 and ASH2L^{SPRY} were also well-defined in the cryo-EM structure (Figure 2.6a and 2.6b). Importantly, our structure indicated that the WDR5-MLL1^{SET}-ASH2L^{SPRY} sub-complex did not make direct contacts with nucleosomal DNA, which was experimentally confirmed by the gel mobility assays (Figure 2.4e and data not shown). The catalytic site of the MLL1^{SET} domain was pointing outward, which might confer distance restraint on substrate accessibility (Figure 2.6b, semi-transparent red circle). Furthermore, the overall domain architecture of the MLL1 core complex within MLL1^{RWSAD}-NCP was largely conserved in the NCP-free yeast SET1 complexes [7, 21], suggesting that NCP binding may not require or induce major conformational changes in the MLL1 core complex (Figure 2.6c).



Crystal structure of WDR5 (PDB ID: 2H14)

Crystal structure of MLL1^{SET}-ASH2L^{SPRY}-RbBP5$^{330–375}$ (PDB ID: 5F6L)

MLL1^{RWSAD}-NCP (the composite map)

Crystal structure of SET1 (*K. lactis*, PDB ID: 6CHG)

Cryo-EM structure of SET1 (*S. cerevisiae*, PDB ID: 6BX3)

Figure 2.6. The WDR5-MLL1$^{\mathrm{SET}}$-ASH2L$^{\mathrm{SPRY}}$-NCP subcomplex. **a**, Rigid-body fitting of human WDR5 crystal structure (PDB ID: 2H14) [3] into the cryo-EM map of MLL1$^{\mathrm{RWS}}$-NCP. Secondary structures of WDR5 were shown in green. **b**, Rigid-body fitting of MLL1$^{\mathrm{SET}}$ and ASH2L$^{\mathrm{SPRY}}$ into the cryo-EM map of MLL1$^{\mathrm{RWSAD}}$-NCP. The MLL1$^{\mathrm{SET}}$-RbBP5$^{330\text{-}375}$-ASH2L$^{\mathrm{SPRY}}$ crystal structure (PDB ID: 5F6L) [1] was used. Characteristic secondary structure of MLL1$^{\mathrm{SET}}$ (SET-I, SET-C and SET-N) were shown with black dashed circles. The catalytic active site was represented by a semi-transparent red circle. **c**, Comparison of human MLL1-NCP structure with ySET1 crystal structure (*K. lactis*, dark goldenrod, PDB ID: 6CHG, right top) [5] and cryo-EM structure (*S. cerevisiae*, light slate blue, PDB ID: 6BX3, right bottom) [7]. Orange circles indicated IDR regions of ASH2L (*Homo sapiens*), Bre2 (*K. lactis*), and Cps60 (*S. cerevisiae*). An extra domain in *S. cerevisiae* SET1 complex, Cps40, or in *K. lactis*, Spp1, colored gray. Figure prepared by USC and SHP

Dynamic ASH2L-NCP interaction is critical for H3K4me3 activity

The second docking point of the MLL1 core complex to the NCP was provided by the intrinsically disordered regions (IDRs) of ASH2L (Figures 2.2b and 2.6c). The ASH2L-NCP interface was highly dynamic in solution (Figure S2.3e in Appendix A), making it challenging to visualize the molecular details. Similar dynamic behaviour was observed for IDR of the yeast homologue Cps60 (Figure 2.6c), which was not resolved in the cryo-EM structure of the yeast SET1 complex [7]. Since the crystal structure of full-length human ASH2L has not been reported, we employed the protein structure prediction approach using the iterative template-based fragment assembly refinement (I-TASSER) method [51, 52]. The crystal structure of yeast Bre2 was used as template (PDB ID: 6CHG) [21] to build the ASH2L plant homeodomain-wing helix (PHD-WH)/IDRs model (Figures 2.7a and S2.5a in Appendix A). After resolving minor clashes, we were able to reliably dock ASH2L IDRs into the cryo-EM map of MLL1$^{\mathrm{RWSAD}}$-NCP (Figures 2.7a, S2.5a, and S2.6b in Appendix A). The model of MLL1$^{\mathrm{RWSAD}}$-NCP showed that ASH2L IDRs interacted with SHL7 of nucleosomal DNA (Figures 2.2b and 2.7b). Surprisingly, the PHD-WH domain of ASH2L was located outside of the cryo-EM map (Figure S2.6a in Appendix A) despite the reported function in DNA binding [53, 54].

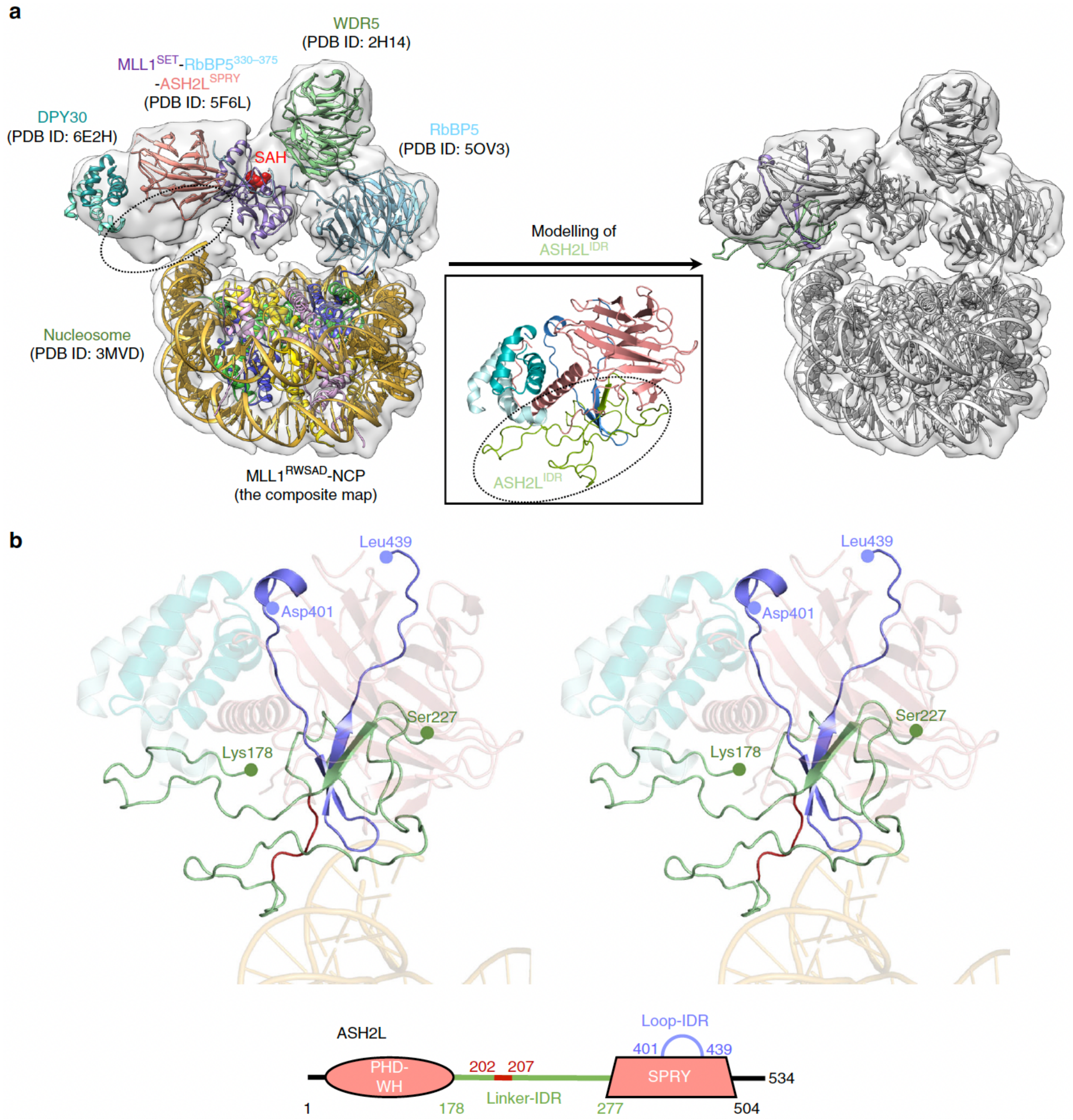


Figure 2.7. ASH2L interacts with nucleosomal DNA through IDRs. **a**, Structure prediction of ASH2L^{IDR}. The ASH2L IDR regions was not available and thus not assigned in the corresponding cryo-EM map (dashed circle). The structure prediction approach was employed to model ASH2L IDR regions as described in the methods. Linker-IDR and Loop IDRs were colored green and blue, respectively in the ASH2L^{IDR} model structure. **b**, Stereo-view of the ASH2L-DPY30 model structure and its contacts with DNA. The structure of ASH2L is a composite from the crystal structure of ASH2L^{SPRY} (PDB ID: 5F6L) [1] and the modeled ASH2L^{IDR}. The schematics of ASH2L were shown at the bottom and key residues 202-207 in ASH2L^{IDR} were highlighted in red. Figure prepared by USC and SHP; modeling in **a** and **b** done by WZ

Our MLL1^{RWSAD}-NCP model pinpointed a short stretch of positively charged residues (i.e., K205/R206/K207, $_{205}$KRK$_{207}$) in the ASH2L Linker-IDR to make contacts with nucleosomal DNA (Figure 2.7b). These positively charged residues were highly conserved in ASH2L homologs in higher eukaryotes (Figure 2.8a). To biochemically validate the model structure of ASH2L IDRs, we first examined the importance of key

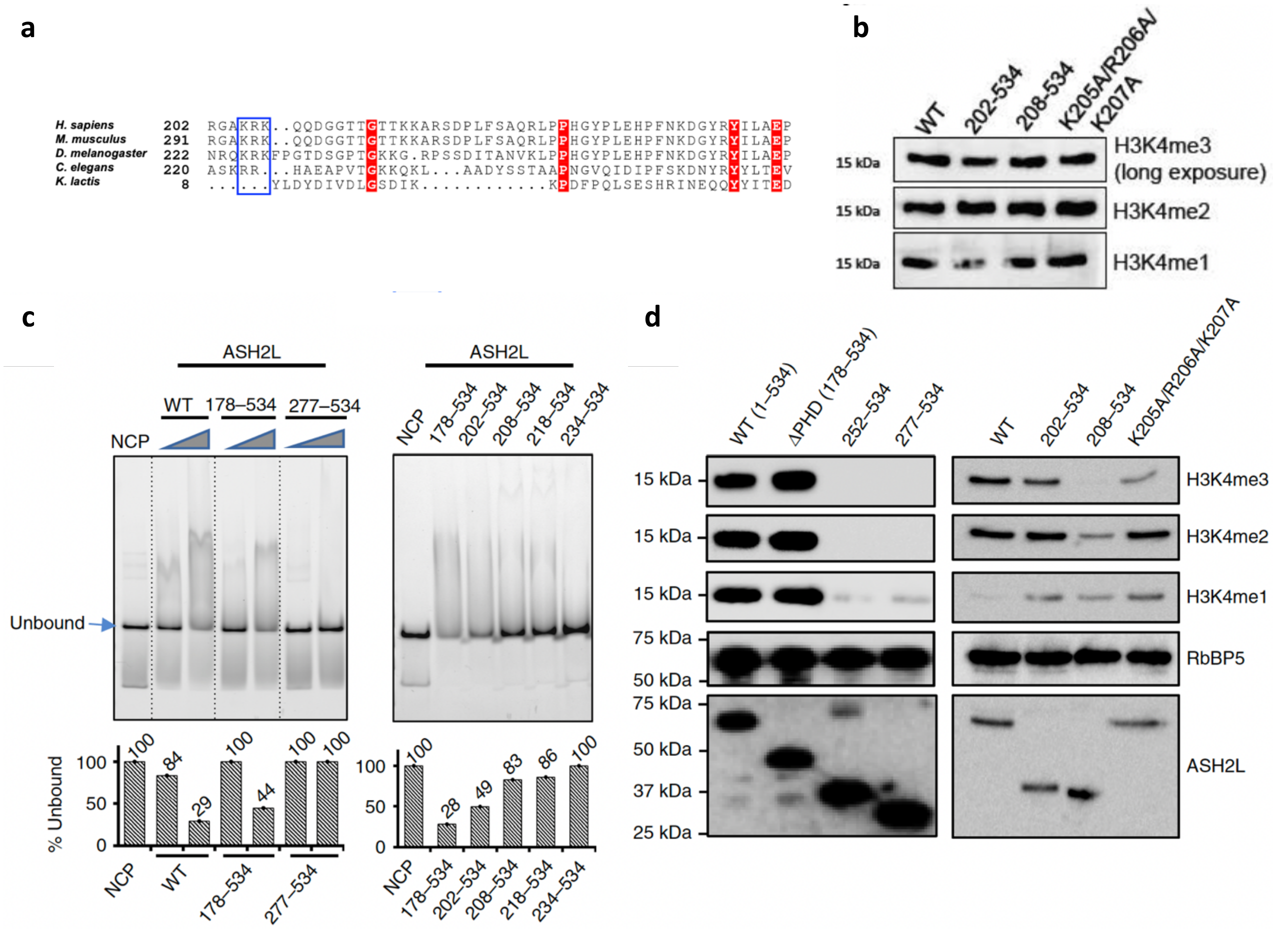


Figure 2.8. ASH2L Linker IDR is important for NCP binding and methyltransferase activity. **a**, Multiple sequence alignment of ASH2L Linker-IDR region (residues 202-254). The blue box indicated $_{205}$KRK$_{207}$, key residues for NCP recognition. **b**, Immunoblot to detect *in vitro* histone methyltransferase activity with the free H3 as the substrate. Reconstituted MLL1^{RWSAD} complexes containing wild-type and mutant ASH2L used as indicated on top. Immunoblots of RbBP5 and ASH2L included as controls. **c**, Top, electrophoretic mobility shift assay of ASH2L and ASH2L mutants as indicated on top. Bottom, the unbound NCP in the gel image was quantified by ImageJ and presented after normalization against the NCP alone signal, which was arbitrarily set as 1 (100%). This experiment was repeated separately to confirm the main conclusions. **d**, Immunoblot to detect *in vitro* histone methyltransferase activity with the NCP as the substrate. Reconstituted MLL1^{RWSAD} complexes containing wild-type and mutant ASH2L used as indicated on top. Immunoblots of RbBP5 and ASH2L included as controls. Western in **b** done by JX; EMSA in **c** done by YTL.

residues involved in the NCP interaction. As shown in Figure 2.8c, ASH2L directly interacted with the NCP, resulting in a

mobility shift in the native gel. However, deletion of both PHD-WH (residues 1–178) and Linker-IDR (residues 178–277), but not PHD-WH alone, abolished ASH2L interaction with the NCP (Figure 2.8c). Further truncation of ASH2L Linker-IDR identified that residues 202–207 were important for NCP interaction, consistent with our ASH2L model (Figure 2.7b). Binding of ASH2L to the NCP was critical for MLL1 methyltransferase activity. Deletion of ASH2L Linker-IDR completely abolished the MLL1 activity on the NCP (Figure 2.8d, left). Similarly, deletion of ASH2L residues 202–207 or mutation of $_{205}$KRK$_{207}$ to alanine significantly reduced MLL1 activity on the NCP for H3K4me3 (Figure 2.8d, right), but not on free H3 (Figure 2.8b). This result, together with that of RbBP5, suggests that MLL1-NCP interactions specifically promote tri-methylation of H3K4. Notably, deletion of ASH2L Linker-IDR led to more drastic reduction of overall H3K4me, suggesting additional mechanisms by which Linker-IDRs may contribute to MLL1 regulation (see Discussion).

Alignment of MLL1^{SET} at the nucleosome dyad

Given the binding of RbBP5 and ASH2L at the edge of the NCP (SHL1.5 and SHL7), the catalytic MLL1$^{\mathrm{SET}}$ domain was positioned at the nucleosome dyad (Figure 2.9a). In this arrangement, the active site of the MLL1$^{\mathrm{SET}}$ domain pointed outward (Figure 2.9a). The NCP structure was well-resolved in the cryo-EM map of MLL1$^{\mathrm{RWSAD}}$-NCP. Both histone H3 tails emanated from between two gyres of nucleosomal DNA, with Lys37 as the first observable residue on histone H3 N-terminal tails (Figure 2.9a). The distance between Lys37 on each histone H3 tail and the active site of MLL1$^{\mathrm{SET}}$ was ~ 60 Å. Thus, K4 residues on both H3 tails were almost equidistant to the MLL1^{SET} active site (Figure 2.9a). The distance constraint restricted access of the MLL1^{SET} domain to only K4 and K9 on H3 N-terminus (Figure 2.9a). More importantly, since MLL1^{SET} is a non-processive enzyme [26], close proximity of the MLL1^{SET} domain to K4 on both H3 tails likely played a significant role in promoting its activity on higher H3K4me states on the NCP (Figure 2.9b, see discussion).

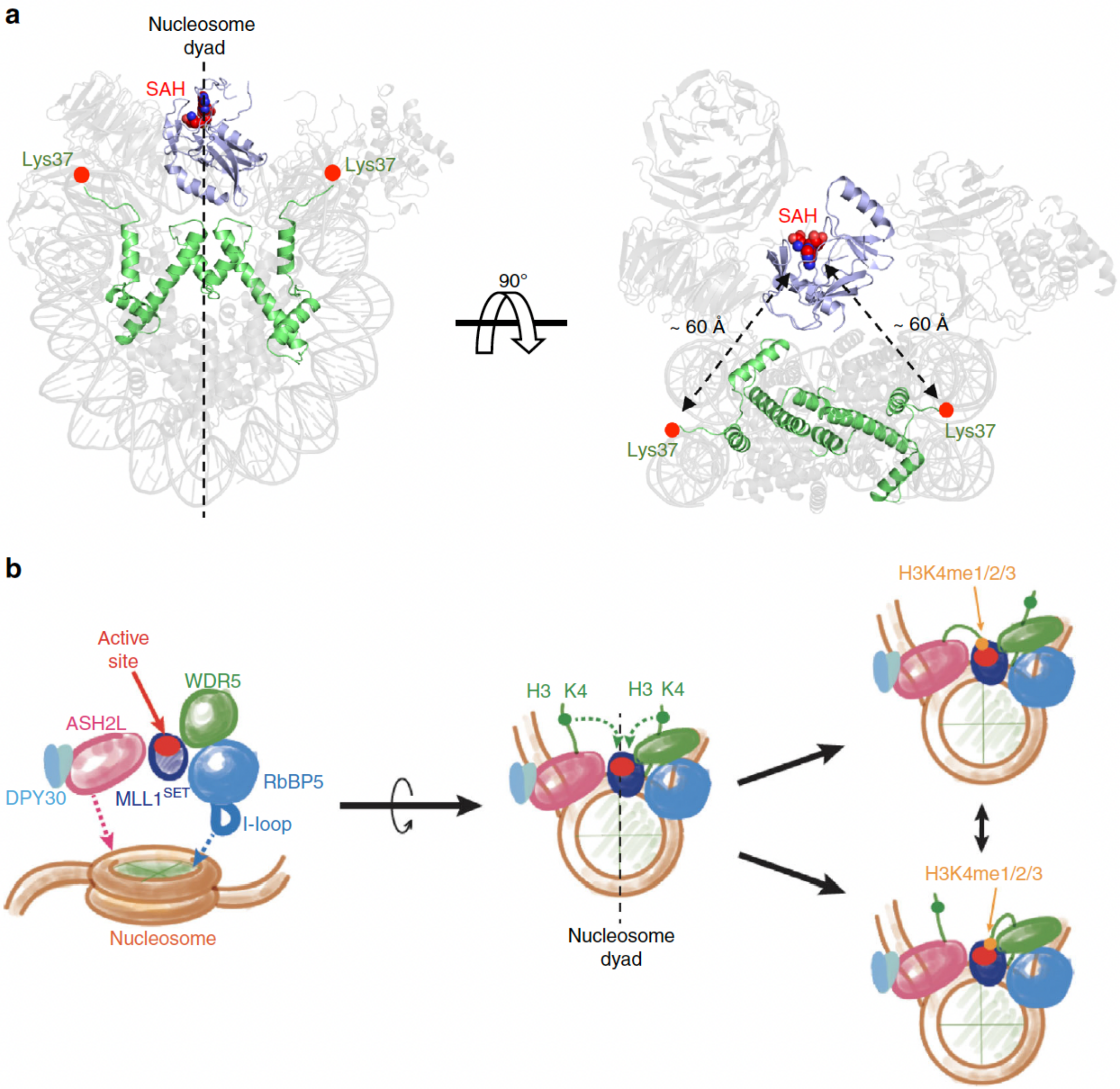


Figure 2.9. The MLL1^{SET} domain aligns at the nucleosome dyad. **a**, MLL1^{SET} (slate) and two copies of histone H3 (green) were highlighted against the faded MLL1^{RWSAD}-NCP structure. The. SAD molecule represented as a sphere and marked the catalytic active site of MLL1^{SET} **b**, Schematic model of NCP recognition mediated by the MLL1 core complex. The I-loop of RbBP5 and the active site of MLL1^{SET} were colored blue and red, respectively. Arrangement and mechanism prepared by USC

2.4 Discussion

Here we report the single-particle cryo-EM structure of the human MLL1 core complex bound to the NCP. It shows that the MLL1 core complex anchors on the NCP through motifs emanating from between WD40 domains of RbBP5 and Linker-IDR of ASH2L. This dual interaction positions the catalytic MLL1^{SET} domain near the nucleosome dyad, allowing symmetric access to both H3K4 substrates within the NCP. Disruption of MLL1-NCP interaction specifically reduces MLL1 activity for nucleosomal H3K4me3. Our study sheds light on how the MLL1 core complex engages chromatin and how chromatin binding promotes MLL1 tri-methylation activity.

Unique MLL1-NCP recognition among chromatin recognition complexes

One of the well-known features of the NCP is the acidic patch, which is a negatively charged and solvent exposed surface. It is organized by a series of negatively-charged residues in histones H2A and H2B [47]. The acidic patch interacts with the basic patch on histone H4 of adjacent nucleosomes, which underlies inter-nucleosome interactions in higher order chromatin structures [55]. Structures of NCP-protein complexes demonstrate that the acidic patch is recognized by NCP-interacting proteins in many cases through diverse arginine finger motifs (e.g., LANA [8], RCC1 [9], 53BP1 [56]). Our study demonstrates that the MLL1 core complex binds to a unique surface of the NCP that does not involve the acidic patch. The main contributors to the NCP interaction are the electrostatic interactions between positively charged residues in RbBP5 and ASH2L and the DNA backbone in the NCP. Extensive DNA-interactions were also observed for the histone H3K27 methyltransferase PRC2 in complex with dinucleosomes [57]. However, unlike PRC2, all MLL1-NCP interactions occur within a single nucleosome. It is possible that other domains of MLL1 (e.g., MLL1 PHD-BRD) that are not included in our study are important for engaging adjacent nucleosomes and spreading the H3K4me marks.

In our structure, the I-loop of RbBP5, inserting between the H4 tail and nucleosomal DNA (SHL1.5), provides the specific docking point of the MLL1 core complex to the NCP. Once RbBP5 is docked, the distance between RbBP5 and ASH2L (~ 70 Å) limits ASH2L binding to SHL7 of the NCP (Figure S2.6c in Appendix A). Interestingly, despite importance of RbBP5 I-loop in specifying the orientation of MLL1 on the NCP, it did not contribute significantly to binding of RbBP5 to the NCP (Figure 2.4c). Nonetheless, dual recognition through both specific and non-specific interactions of RbBP5 and ASH2L likely enables the MLL1 core complex to bind the NCP in a unique configuration for optimal access to both H3K4 substrates (Figure 2.9b). Notably, H4 interactions are also used by other chromatin interacting proteins e.g., DOT1L, SNF, ISWI [48, 49, 58-61], raising the possibility of another NCP docking site in addition to the acidic patch.

Structural conservation and divergence of the human MLL1 core complex

The composition of catalytically-active human MLL/SET and yeast SET1 complexes is largely conserved [14]. Our study shows that human MLL1 core complex and the yeast SET1 complexes [7, 21] have the same overall architecture, with the catalytic MLL1^{SET} domain sandwiched by RbBP5-WDR5 (Swd1-Swd3 in *Kluveromyces lactis*/Cps50-Cps30 in *Saccharomyces cerevisiae*) and ASH2L-DPY30 (Bre2-Sdc1/Cps60-Cps25 in yeast) on each side (Figure 2.6c). Furthermore, the crystal structure of the yeast SET1 complex

[21] overlays well with the MLL1 core complex in our structure. This may suggest that the yeast SET1 complex adopts a similar configuration on the NCP.

Given that RbBP5 and ASH2L are shared among all MLL family protein complexes in mammals, it is likely that our study reveals a general mechanism for how the mammalian MLL complexes engage chromatin and gain access to the H3K4 substrates. Sequence alignments show significant conservation of RbBP5 I-/A-loops as well as the basic residues in ASH2L in higher eukaryotes. It supports the functional importance of these regions in chromatin binding and H3K4me regulation. It also suggests that the mechanism by which MLL family enzymes engage chromatin in higher eukaryotes is likely to be conserved. Importantly, recent genome sequencing studies have identified mutations in these conserved regions in human malignancies [62, 63], which warrants future studies. We would like to point out that the interface of RbBP5 and ASH2L with the NCP are not well conserved in homologous yeast Swd1/Cps50 and Bre2/Cps60 protein (Figures 2.8a and S2.4b in Appendix A). The I-loop is much shorter, and the A-loop is missing in the yeast Swd1/Cps50 protein (Figure 2.5d), suggesting potential divergence of detailed yeast SET1-NCP interactions at the molecular level, although the overall NCP recognition pattern might be similar.

Contribution of ASH2L IDRs in NCP recognition and H3K4me regulation

Previous studies have shown that several components of the MLL1 core complex are capable of interacting with DNA or RNA, including WDR5, RbBP5 and ASH2L [12, 53, 54, 64]. This raises the question of how these interactions contribute to recruitment of the MLL1 core complex to chromatin. Our structure indicates that WDR5, MLL1^{SET} as well as

PHD-WH domain of ASH2L do not directly interact with the NCP, suggesting that these proteins probably interact with nucleosome-free DNAs and indirectly contribute to the stability of the MLL1 core complex on chromatin. Our study reveals a previously uncharacterized function of ASH2L Linker-IDRs in chromatin function. This region was not studied in the previous MLL1^{SET}-ASH2L$^{SPRY,\Delta400-440}$-RbBP5$^{330-375}$ structure [1]. The ASH2L IDRs contain evolutionarily conserved sequences (Figure 2.8a and S2.4b in Appendix A) and exhibit dynamic properties on the NCP. Notably, we take advantage of protein structure prediction approach to identify the essential interface between ASH2L IDRs and the NCP. This approach subsequently allowed us to uncover a basic patch region in ASH2L ($_{205}$KRK$_{207}$) that significantly contributes to MLL1 binding to the NCP and MLL1 tri-methylation activity. However, it is likely that other regions of ASH2L also contribute to NCP binding since deletion of the ASH2L residues 202–254 leads to more prominent reduction of H3K4me (Figure 2.8d). Molecular dissection of ASH2L IDRs awaits future studies.

Regulation of MLL1 activity for H3K4me3 on the NCP

Single turnover kinetic experiments revealed that the MLL1 core complex uses a non-processive mechanism for catalysis [26], requiring capture and release H3K4 after each round of the methylation reaction. Thus, it is not kinetically favorable to achieve tri-methylation state when enzyme and substrate have stochastic encounters in solution. Our study here demonstrates that the MLL1 core complex stably associates with NCP via RbBP5 and ASH2L, which uniquely positions the MLL1^{SET} domain at the nucleosome dyad with near symmetric access to both H3K4 substrates (Figure 2.9a). Stable settlement on the NCP allows close physical proximity and optimal orientation of the MLL1^{SET} catalytic site to both H3K4 substrates on the NCP, which significantly favor the kinetics of successive methylation reactions. In support, disruption of the MLL1-NCP interactions significantly reduces H3K4me3 activity on the NCP without affecting MLL1 activity on free H3. Given enhanced MLL1 activity on the NCP, we envision that stabilizing the MLL1 complex on chromatin by transcription factors and cofactors will further enhances overall H3K4me3, which in turn recruits additional transcription cofactors [15, 17-20]. This leads to a feedback loop for optimal gene expression and spreading of H3K4me3 at the actively transcribed genes in cells. Position of the MLL1^{SET} domain at the nucleosome dyad also raises the question of potential interplay with linker histones, which bind near the nucleosome dyad in the heterochromatic regions in eukaryotes [65-67]. It would be interesting to test whether linker histone inhibits MLL1 activity and thus promotes closed chromatin conformation in future.

2.5 Conclusion

Historic biochemical studies of the MLL1 complex have focused on simple peptidic or recombinant histone H3 and prior structural work used a minimal MLL1 subcomplex that eliminated much of the disordered, flexible regions of MLL1 and its binding partners RbBP5 and ASH2L [1, 26, 68]. Here we find that MLL1 engages a nucleosome using RbBP5 and ASH2L as anchors at SHL1.5 and 7, respectively. Motifs essential to this interaction from RbBP5 include a quadruple arginine motif, an inserting loop (I-loop) and anchoring loop (A-loop) that mainly interact with DNA phosphate backbone at SHL1.5. On the opposing end of the complex, at SHL7, ASH2L provides anchoring points through newly formed β-sheet between the flexible linker ($_{205}$KRK$_{207}$) and loop IDRs ($_{419}$KFK$_{421}$) unique to mammalian ASH2L.

Given our initial biochemical findings revolving around DPY30-mediated activation on the nucleosome alone, this model of active binding is attractive due to the near symmetrical access of each H3 tail to the MLL1^{SET} active site. Despite this, we noted another population in which the MLL1 complex sits across the nucleosome disc, however, the density associated with this subset lacked DPY30. Whether these two states represent discrete orientations or simply are endpoints of a continuum remains unclear given MLL1 has previously been shown to be non-processive [26]. However, these findings clearly suggested MLL1 engages in a dynamic exchange on the nucleosome interface and future work would resolve the significance of each orientation.

2.6 Materials and Methods

Protein expression and purification

The core subunits of the MLL1^{RWSAD} complex (MLL1$^{SET,3762-3969}$, ASH2L$^{1-534}$ and mutants, RbBP5$^{1-538}$ and mutants, WDR5$^{23-334}$, and DPY30$^{1-99}$) were expressed and purified using the pET28a His$_6$-small ubiquitin-related modifier (SUMO) vector as previously described [1, 69]. Mutations of RbBP5 and ASH2L were generated by overlap PCR-based site-directed mutagenesis. Individual components of the MLL1^{RWSAD} complex were purified on Ni-NTA column and equimolar quantities were mixed and purified by gel filtration chromatography as previously described [1, 69]. Full-length *X. laevis* histones H2A, H2B, H3, and H4 were expressed and purified using the one-pot purification method [70]. Assembly of the nucleosome core particle using 147 base-pair Widom 601 DNA [71] was done by salt dialysis as previously described [72, 73].

In vitro histone methyltransferase assay

The *in vitro* histone methyltransferase assay was carried out by incubating the MLL1^{RWSAD} complex (0.3 µM) with either nucleosome (0.965 µM) or free recombinant histone H3 (0.098 µM) for 1 hour at room temperature. The reaction buffer contained 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM DTT, 5 mM $MgCl_2$ and 10% v/v glycerol in a total volume of 20 µL. Reactions were quenched with 20 µL of 2X Laemmli Sample Buffer (Bio-Rad cat. #161-0737). H3K4 methylation was detected by western blot using antibodies for $H3K4me_1$ (1:20000, Abcam cat. No. ab8895), $H3K4me_2$ (1:40000, EMD-Millipore cat. #07-030), or $H3K4me_3$ (1:10000, EMD-Millipore cat. #07-473) for either 1 h at room temperature or overnight at 4 ºC. The blot was then incubated with IgG-HRP (Santa Cruz Biotechnology cat. #sc-486) for 1 h at room temperature. The membrane was developed using ECL (Pierce cat. #32106) and visualized by chemiluminescence (Bio-Rad ChemiDoc Imaging System).

Electrophoretic Mobility Shift Assay (EMSA)

EMSA assay was carried out using 0.1 µM nucleosomes and increasing concentration of MLL1 subunits. The protein mixture was run on the 6% 0.2X TBE gel that was pre-run for 1.5 hours, 150 V at 4 ºC. The gel was visualized by incubating in 100 mL of TBE with 1:20000 diluted ethidium bromide for 10 minutes at room temperature. Gels were then incubated in distilled water for 10 minutes and visualized by UV transillumination (Bio-Rad ChemiDoc Imaging System). The results were quantified by ImageJ software [74].

His_6 Pull-down Assay

His_6-fusion proteins were incubated with the NCP in BC150 (20 mM Tris-HCl, pH 7.5, 350 mM NaCl, 20 mM imidazole, 0.05% v/v NP-40, 10 mM DTT, 1 mg/ml BSA, PMSF and inhibitor cocktail) for 2 hours at 4 ºC. After several washes with BC150, the beads were boiled in SDS loading buffer and analyzed by Western blot.

Cryo-EM sample preparation

The cryo-EM sample was prepared by the GraFix method [75]. Specifically, the reconstituted $MLL1^{RWSAD}$ complex (30 μM) was incubated with nucleosomes (10 μM) in GraFix buffer (50 mM HEPES, pH7.5, 50 mM NaCl, 1 mM $MgCl_2$, 1 mM TCEP) with added 0.5 mM SAH for 30 min at 4 °C. The sample was applied onto the top of the gradient solution (0-60% glycerol gradient with 0-0.2% glutaraldehyde, in GraFix buffer) and was centrifuged at 48,000 rpm at 4 °C for 3 hours. After ultracentrifugation, 20 μl fractions were manually collected from the top of the gradient. The crosslinking reaction was terminated by adding 2 μl of 1 M Tris-HCl, pH7.5 into each fraction. Glycerol was removed by dialyzing the sample in GraFix buffer using centrifugal concentrator (Sartorius Vivaspin 500) before making cryo-EM grids.

Cryo-EM data collection and processing

A protein sample at 1 mg/ml concentration was plunge-frozen on 200 mesh Quantifoil R1.2/1.3 grids (Electron Microscopy Sciences) using a Mark IV Vitrobot (Thermo Fisher Scientific) with settings as 4 °C, 100% humidity and 4 sec blotting time. Cryo-EM grids were imaged on a FEI Titan Krios operating at 300 keV at liquid nitrogen temperature. The Gatan K2 Summit direct electron detector was used at a nominal magnification of 29,000X in a counting mode with a pixel size of 1.01 Å/pixel. A dose rate of 8 electrons/Å^2/s and defocus values ranging from -1.5 to -3.5 μm were used. Total exposure of 8 sec per image was dose-fractionated into 40 movie frames, resulting in an accumulated dose of 64 electrons per Å^2. A total of 4717 movies were collected for the MLL1^{RWSAD}-NCP dataset.

Micrograph movie stacks were first subjected to MotionCor2 for whole-frame and local drift correction [76]. For each micrograph, CTFFIND4.1 was used to fit the contrast transfer function [77]. The estimated resolution of micrographs lower than 5 Å were excluded from further processing, which resulted in 3896 micrographs. Particle picking was performed using the Warp [78], which picked total 712,198 particles. Using particle coordinates obtained from the Warp, the particles were extracted with the box size of 350 Å using RELION 3 program package [79]. Extracted particles were then imported into cryoSPARC [80] for 2D classification in 200 classes. After removal of bad classes, the total of 694,180 particles were subjected to *ab initio* 3D classification (Figure S2.2 in Appendix A). The major class (323,408 particles) contained the MLL1 core complex and the NCP, which was then subjected for the heterogeneous refinement. This led to the identification of ten subclasses. One subclass showed the partial cryo-EM density for the MLL1 core complex, thus excluded for the further processing. The remaining nine subclasses (252,109 particles) maintained intact MLL1^{RWSAD}-NCP complex. These nine subclasses also used for the rigid-body fitting of individual component of the MLL1^{RWSAD}-NCP complex to visualize the dynamics of each component against the NCP (Figure S2.3E in Appendix A). 252,109 particles were imported in RELION and performed the 3D classification without alignment (10 classes, 35 cycles, T=40). One out of 10 classes (8,433 particles) exhibited the well-defined map of MLL1^{RWSAD}-NCP. These particles were used for 3D refinement in RELION and post-processed to a resolution of 6.2 Å and a B factor of -189 Å^2. This cryo-EM map was local filtered using RELION to the local resolution to avoid over-interpretation.

To obtain a cryo-EM map for RbBP5-NCP and MLL1^{RWS}-NCP subcomplexes, we utilized RELION's multi-body refinement procedure with 252,109 particles (Figure S2.2 in Appendix A). RbBP5-NCP (32,563 particles), MLL1^{WSAD}, MLL1^{RWS}-NCP (21,114 particles), and MLL1^{AD} were separately masked during the multibody refinement [81]. The partial signal subtraction was performed to generate the particle set for RbBP5-NCP and MLL1^{RWS}-NCP. Further 3D classifications without alignment (5 classes, 35 cycles, T=40 for RbBP5-NCP and 10 classes, 35 cycles, T=40 for MLL1^{RWS}-NCP) were performed and the best maps based on the resolution and occupancy of RbBP5 and MLL1^{RWS} densities were selected for further refinement and post-processing (Figure S2.2 in Appendix A). The reported final resolution of each cryo-EM structure was estimated by RELION with Fourier shell correlation (FSC) at criteria of 0.143 (Figure S2.3a-c in Appendix A).

Modeling, rigid body fitting, and model refinement

We built a 3D atomic model of the human ASH2L protein by I-TASSER [51, 52, 82] assisted by deep-learning based contact-map prediction [83]. The fragment-guided molecular dynamics refinement software, FG-MD [84], was utilized to remove the steric clash between ASH2L model and other molecules and further refine the local structures (Figure S2.5b in Appendix A). Finally, our in-house EM-fitting software, EM-Ref (Zhang et al, in preparation), was used to fit the ASH2L model and other parts of human MLL1 core complex to the density maps to get final atomic models.

I-TASSER utilized LOMETS, which consisted of 16 individual threading programs [85], to generate templates as the initial conformation. Human ASH2L protein consisted of three domains, while the 2nd, 3rd domains (Linker-IDR and ASH2L^{SPRY}) and C-terminal SDI motif can be covered by templates (PDB ID: 6E2H and 6CHG, B chain, crystal structure of the yeast SET1 H3K4 methyltransferase catalytic module [21]) in most of the top threading alignments. The 1st domain (PHD-WH domain) was covered by another template (PDB ID: 3S32, A chain, the crystal structure of ASH2L N-terminal domain) [54]. Therefore, these three proteins were used as the main templates for building the full-length ASH2L model, where structural assembly simulation was guided by the contact-maps from the deep-learning program, ResPRE [83]. Finally, the first model of I-TASSER was selected as the potential ASH2L model, where the estimated TM-score [86] for the C-terminal domain is 0.71 ± 0.12, suggesting that the confidence of the I-TASSER model is high. Superposing ASH2L model (Linker-IDR and ASH2L^{SPRY}) with the experimental structure (ASH2L^{SPRY}) is shown in Figure 2.7a.

Monte Carlo (MC) simulation was employed to fit and refine the complex model structures based on the experimental density map. During the MC simulations, individual domain structures were kept as the rigid-body, where global translation and rotation of the domains were performed, which would be accepted or rejected based on Metropolis algorithm [87]. The total number of translation and rotation was 50,000 in the MC simulation. The MC energy function (Equation 2.1) used in the simulation was a linear combination of correlation coefficient (CC) between structural models and the density map data and the steric clashes between the atomic structures, i.e.,

$$E_{main} = w_1(1.0 - \frac{\sum_{y\in \mathrm{DM}}(\rho_o(y)-\bar{\rho}_o)(\rho_c(y)-\bar{\rho}_c)}{\sqrt{\sum_{y\in \mathrm{DM}}(\rho_o(y)-\bar{\rho}_o)^2}\sqrt{\sum_{y\in \mathrm{DM}}(\rho_c(y)-\bar{\rho}_c)^2}}) + w_2\sum_{i\in L}\sum_{j\in L,i\neq j}\varepsilon_{ij}[(\frac{r_{ij}}{d_{ij}})^{12} - 2(\frac{r_{ij}}{d_{ij}})^{6}].$$

Equation 2.1. MC Energy Function

Where $\rho_c(y)$ was the calculated density map on grid [88]. $\rho_o(y)$ was obtained from the experimental density map. $\bar{\rho}_c$ and $\bar{\rho}_o$ were the average of calculated density map and experimental density map, respectively. DM and L represented tbe density map and the length of protein, respectively. d_{ij} was the distance between the two atoms i and j. r_{ij} was the sum of their van der Waals atomic radii and ε_{ij} was the combined well-depth parameter for atoms i and j, which were all taken from the CHARMM force field [89]. $w_1 = 100$ and $w_2 = 1$ were the weights for correlation coefficient item and clash item, respectively.

For the nucleosome model, the crystal structure of nucleosome (PDB ID:3MVD) [9] was used for rigid-body fitting. In the cryo-EM structure of RbBP5-NCP, the histone H4 tail region was manually rebuilt where the density allowed using the program COOT [90]. Three model structures of MLL1$^{\mathrm{RWSAD}}$-NCP, RbBP5-NCP, and MLL1$^{\mathrm{RWS}}$-NCP were subjected to the real-space refinement using PHENIX after rigid-body fitting. Validations of three model structures were performed by MolProbity [91]. The final structures were further validated by calculating map-model FSC curves using phenix.mtriage in the PHENIX program package (Figure S2.3d in Appendix A) [92]. The computed FSC between the model and map agreed reasonably well as shown in Figure S2.3d in Appendix A. Statistics for data collection, refinement, and validation summarized in Table 2.1 in Appendix A.

Model quality estimation of the ASH2L IDR region

The estimated TM-score of the entire model is 0.67 ± 0.13 and the CC between the predicted model and cryo-EM density map was 0.696. These data showed that the predicted model was a confident model and there was a good fitting between the predicted model and the density map. To further check the local model quality, especially for the IDR region, we gave the residue-level B-factor predicted by ResQ [93] and the CC score between the predicted model and the cryo-EM density map in Figure S2.7 in Appendix A.

B-factor was estimated by ResQ, which uses support vector regression that makes use of the local structural information between the model and (1) threading templates, (2) structure alignment templates, (3) reference decoys, and (4) sequence-based secondary structure and solvent accessibility predictions [93]. Since IDR region of the ASH2L model are mainly loops, this region was more flexible and had relatively high B-factors (Figure S2.7 in Appendix A). However, it is difficult to simply say that IDR region has a good model quality. Therefore, after fitting the model to the density map, we gave the residue-level CC score between each residue of the ASH2L IDR model and the corresponding residue of the density map (Figure S2.7 in Appendix A) to further test the quality of the IDR region. The residue-level CC score can be calculated using Equation 2.2. The masking distance in Equation 2.2 is 5 Å if every atom is used to compute ρ_c, where ρ_c is the density map calculated from the fitted model. ρ_o is experimental density map. y is the grid point where its distance to atoms of residue i is < 5 Å [88]. Positive value of CC score indicates good fitting quality of the model and the density map. In the IDR region, especially for the DNA binding interface residues (residues 205–207), most of all residues had positive CC scores, indicating that by combining the information from the predicted model and the density map, our model for the IDR region is trustable.

$$CC(R_i) = \frac{\sum_{y \in R_i} (\rho_o(y) - \bar{\rho}_o)(\rho_c(y) - \bar{\rho}_c)}{\sqrt{\sum_{y \in R_i} (\rho_o(y) - \bar{\rho}_o)^2} \sqrt{\sum_{y \in R_i} (\rho_c(y) - \bar{\rho}_c)^2}}$$

Equation 2.2. Residue-level CC score calculation

2.7 References

1. Li, Y., et al., *Structural basis for activity regulation of MLL family methyltransferases.* Nature, 2016. **530**(7591): p. 447-52.

2. Worden, E.J., X. Zhang, and C. Wolberger, *Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome.* eLife, 2020. **9**: p. e53199.

3. Couture, J.F., E. Collazo, and R.C. Trievel, *Molecular recognition of histone H3 by the WD40 protein WDR5.* Nat Struct Mol Biol, 2006. **13**(8): p. 698-703.

4. Kornberg, R.D., *Chromatin Structure: A Repeating Unit of Histones and DNA.* Science, 1974. **184**(4139): p. 868-871.

5. Hsu, P.L., et al., *Crystal Structure of the COMPASS H3K4 Methyltransferase Catalytic Module.* Cell, 2018. **174**: p. 1106-1116.e9.

6. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

7. Qu, Q., et al., *Structure and Conformational Dynamics of a COMPASS Histone H3K4 Methyltransferase Complex.* Cell, 2018. **174**(5): p. 1117-1126 e12.

8. Barbera, A.J., et al., *The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA.* Science, 2006. **311**(5762): p. 856-61.

9. Makde, R.D., et al., *Structure of RCC1 chromatin factor bound to the nucleosome core particle.* Nature, 2010. **467**(7315): p. 562-566.

10. Armache, K.-J., et al., *Structural Basis of Silencing: Sir3 BAH Domain in Complex with a Nucleosome at 3.0 Å Resolution.* Science, 2011. **334**(6058): p. 977.

11. Kato, H., et al., *A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C.* Science, 2013. **340**(6136): p. 1110-3.

12. Mittal, A., et al., *The structure of the RbBP5 beta-propeller domain reveals a surface with potential nucleic acid binding sites.* Nucleic Acids Res, 2018. **46**(7): p. 3802-3812.

13. Goddard, T.D., et al., *UCSF ChimeraX: Meeting modern challenges in visualization and analysis.* Protein Sci, 2018. **27**(1): p. 14-25.

14. Rao, R.C. and Y. Dou, *Hijacked in cancer: the KMT2 (MLL) family of methyltransferases.* Nat Rev Cancer, 2015. **15**(6): p. 334-46.

15. Ruthenburg, A.J., et al., *Multivalent engagement of chromatin modifications by linked binding modules.* Nat Rev Mol Cell Biol, 2007. **8**(12): p. 983-94.

16. Chen, K., et al., *Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes.* Nat Genet, 2015. **47**(10): p. 1149-57.

17. Wysocka, J., et al., *A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling.* Nature, 2006. **442**(7098): p. 86-90.

18. Taverna, S.D., et al., *How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers.* Nat Struct Mol Biol, 2007. **14**(11): p. 1025-1040.

19. Lauberth, S.M., et al., *H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation.* Cell, 2013. **152**(5): p. 1021-36.

20. Vermeulen, M., et al., *Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4.* Cell, 2007. **131**(1): p. 58-69.

21. Hsu, P.L., et al., *Crystal Structure of the COMPASS H3K4 Methyltransferase Catalytic Module.* Cell, 2018. **174**(5): p. 1106-1116 e9.

22. Herz, H.-M., et al., *Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4.* Genes & development, 2012. **26**(23): p. 2604-2620.

23. Rea, S., et al., *Regulation of chromatin structure by site-specific histone H3 methyltransferases.* Nature, 2000. **406**(6796): p. 593-9.

24. Dou, Y., et al., *Regulation of MLL1 H3K4 methyltransferase activity by its core components.* Nat Struct Mol Biol, 2006. **13**(8): p. 713-9.

25. van Nuland, R., et al., *Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes.* Mol Cell Biol, 2013. **33**(10): p. 2067-77.

26. Patel, A., et al., *On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex.* J Biol Chem, 2009. **284**(36): p. 24242-56.

27. Avdic, V., et al., *Structural and biochemical insights into MLL1 core complex assembly.* Structure, 2011. **19**(1): p. 101-8.

28. Cosgrove, M.S. and A. Patel, *Mixed lineage leukemia: a structure-function perspective of the MLL1 protein.* FEBS J, 2010. **277**(8): p. 1832-42.

29. Katada, S. and P. Sassone-Corsi, *The histone methyltransferase MLL1 permits the oscillation of circadian gene expression.* Nat Struct Mol Biol, 2010. **17**(12): p. 1414-21.

30. Dou, Y., et al., *Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF.* Cell, 2005. **121**(6): p. 873-85.

31. Wysocka, J., et al., *WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development.* Cell, 2005. **121**(6): p. 859-72.

32. Guenther, M.G., et al., *Global and Hox-specific roles for the MLL1 methyltransferase.* Proc Natl Acad Sci U S A, 2005. **102**(24): p. 8603-8.

33. Ng, S.B., et al., *Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.* Nature genetics, 2010. **42**(9): p. 790-793.

34. Paulussen, A.D., et al., *MLL2 mutation spectrum in 45 patients with Kabuki syndrome.* Hum Mutat, 2011. **32**(2): p. E2018-25.

35. Li, Y., et al., *A mutation screen in patients with Kabuki syndrome.* Hum Genet, 2011. **130**(6): p. 715-24.

36. Micale, L., et al., *Mutation spectrum of MLL2 in a cohort of Kabuki syndrome patients.* Orphanet journal of rare diseases, 2011. **6**: p. 38-38.

37. Hannibal, M.C., et al., *Spectrum of MLL2 (ALR) mutations in 110 cases of Kabuki syndrome.* Am J Med Genet A, 2011. **155a**(7): p. 1511-6.

38. Kluijt, I., et al., *Kabuki syndrome–report of six cases and review of the literature with emphasis on ocular features.* Ophthalmic genetics, 2000. **21**(1): p. 51-61.

39. Jones, W.D., et al., *De novo mutations in MLL cause Wiedemann-Steiner syndrome.* The American Journal of Human Genetics, 2012. **91**(2): p. 358-364.

40. Mendelsohn, B.A., et al., *Advanced bone age in a girl with Wiedemann-Steiner syndrome and an exonic deletion in KMT2A (MLL).* Am J Med Genet A, 2014. **164a**(8): p. 2079-83.

41. Strom, S.P., et al., *De Novo variants in the KMT2A (MLL) gene causing atypical Wiedemann-Steiner syndrome in two unrelated individuals identified by clinical exome sequencing.* BMC medical genetics, 2014. **15**: p. 49-49.

42. Bochynska, A., J. Luscher-Firzlaff, and B. Luscher, *Modes of Interaction of KMT2 Histone H3 Lysine 4 Methyltransferase/COMPASS Complexes with Chromatin.* Cells, 2018. **7**(3).

43. Ge, Z., et al., *WDR5 high expression and its effect on tumorigenesis in leukemia.* Oncotarget, 2016. **7**(25): p. 37740-37754.

44. Butler, J.S., et al., *Low expression of ASH2L protein correlates with a favorable outcome in acute myeloid leukemia.* Leuk Lymphoma, 2017. **58**(5): p. 1207-1218.

45. Magerl, C., et al., *H3K4 dimethylation in hepatocellular carcinoma is rare compared with other hepatobiliary and gastrointestinal carcinomas and correlates with expression of the methylase Ash2 and the demethylase LSD1.* Hum Pathol, 2010. **41**(2): p. 181-9.

46. Haddad, J.F., et al., *Structural Analysis of the Ash2L/Dpy-30 Complex Reveals a Heterogeneity in H3K4 Methylation.* Structure, 2018.

47. Zhou, B.-R., et al., *Atomic resolution cryo-EM structure of a native-like CENP-A nucleosome aided by an antibody fragment.* Nature Communications, 2019. **10**(1): p. 2301.

48. Worden, E.J., et al., *Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L.* Cell, 2019. **176**(6): p. 1490-1501.e12.

49. Anderson, C.J., et al., *Structural Basis for Recognition of Ubiquitylated Nucleosome by Dot1L Methyltransferase.* Cell Rep, 2019. **26**(7): p. 1681-1690 e5.

50. Southall, S.M., et al., *Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks.* Mol Cell, 2009. **33**(2): p. 181-91.

51. Zhang, Y., *I-TASSER server for protein 3D structure prediction.* BMC Bioinformatics, 2008. **9**: p. 40.

52. Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: a unified platform for automated protein structure and function prediction.* Nat Protoc, 2010. **5**(4): p. 725-38.

53. Chen, Y., et al., *Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding.* EMBO Rep, 2011. **12**(8): p. 797-803.

54. Sarvan, S., et al., *Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain.* Nat Struct Mol Biol, 2011. **18**(7): p. 857-9.

55. Kalashnikova, A.A., et al., *The role of the nucleosome acidic patch in modulating higher order chromatin structure.* J R Soc Interface, 2013. **10**(82): p. 20121022.

56. Wilson, M.D., et al., *The structural basis of modified nucleosome recognition by 53BP1.* Nature, 2016. **536**(7614): p. 100-103.

57. Poepsel, S., V. Kasinath, and E. Nogales, *Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes.* Nat Struct Mol Biol, 2018. **25**(2): p. 154-162.

58. Yan, L., et al., *Structures of the ISWI–nucleosome complex reveal a conserved mechanism of chromatin remodeling.* Nature Structural & Molecular Biology, 2019. **26**(4): p. 258-266.

59. Jang, S., et al., *Structural basis of recognition and destabilization of the histone H2B ubiquitinated nucleosome by the DOT1L histone H3 Lys79 methyltransferase.* Genes & development, 2019. **33**(11-12): p. 620-625.

60. Valencia-Sánchez, M.I., et al., *Structural Basis of Dot1L Stimulation by Histone H2B Lysine 120 Ubiquitination.* Mol Cell, 2019. **74**(5): p. 1010-1019.e6.

61. Yao, T., et al., *Structural basis of the crosstalk between histone H2B monoubiquitination and H3 lysine 79 methylation on nucleosome.* Cell Research, 2019. **29**(4): p. 330-333.

62. Biankin, A.V., et al., *Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes.* Nature, 2012. **491**: p. 399-405.

63. Kim, T.-M., et al., *The mutational burdens and evolutionary ages of early gastric cancers are comparable to those of advanced gastric cancers.* The Journal of Pathology, 2014. **234**: p. 365-374.

64. Yang, Y.W., et al., *Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency.* Elife, 2014. **3**: p. e02046.

65. Hayes, J.J., D. Pruss, and A.P. Wolffe, *Contacts of the globular domain of histone H5 and core histones with DNA in a "chromatosome".* Proc Natl Acad Sci U S A, 1994. **91**(16): p. 7817-21.

66. Lu, X., et al., *Drosophila H1 regulates the genetic activity of heterochromatin by recruitment of Su(var)3-9.* Science, 2013. **340**(6128): p. 78-81.

67. Zhou, B.R., et al., *Structural Mechanisms of Nucleosome Recognition by Linker Histones.* Mol Cell, 2015. **59**(4): p. 628-38.

68. Shinsky, S.A. and M.S. Cosgrove, *Unique Role of the WD-40 Repeat Protein 5 (WDR5) Subunit within the Mixed Lineage Leukemia 3 (MLL3) Histone Methyltransferase Complex.* J Biol Chem, 2015. **290**(43): p. 25819-33.

69. Cao, F., et al., *An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain.* PLoS One, 2010. **5**(11): p. e14102.

70. Lee, Y.T., et al., *One-pot refolding of core histones from bacterial inclusion bodies allows rapid reconstitution of histone octamer.* Protein Expr Purif, 2015. **110**: p. 89-94.

71. Lowary, P.T. and J. Widom, *New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning.* J Mol Biol, 1998. **276**(1): p. 19-42.

72. Luger, K., T.J. Rechsteiner, and T.J. Richmond, *Preparation of nucleosome core particle from recombinant histones.* Methods Enzymol, 1999. **304**: p. 3-19.

73. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.

74. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-5.

75. Kastner, B., et al., *GraFix: sample preparation for single-particle electron cryomicroscopy.* Nat Methods, 2008. **5**(1): p. 53-5.

76. Zheng, S.Q., et al., *MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy.* Nat Methods, 2017. **14**(4): p. 331-332.

77. Rohou, A. and N. Grigorieff, *CTFFIND4: Fast and accurate defocus estimation from electron micrographs.* J Struct Biol, 2015. **192**(2): p. 216-21.

78. Tegunov, D. and P. Cramer, *Real-time cryo-EM data pre-processing with Warp.* bioRxiv, 2018.

79. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3.* Elife, 2018. **7**.

80. Punjani, A., et al., *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination.* Nat Methods, 2017. **14**(3): p. 290-296.

81. Nakane, T., et al., *Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION.* Elife, 2018. **7**.

82. Zheng, W., et al., *I-TASSER gateway: A protein structure and function prediction server powered by XSEDE.* Future Gener Comput Syst, 2019. **99**: p. 73-85.

83. Li, Y., et al., *ResPRE: high-accuracy protein contact prediction by coupling precision matrix with deep residual neural networks.* Bioinformatics, 2019. **35**(22): p. 4647-4655.

84. Zhang, J., Y. Liang, and Y. Zhang, *Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling.* Structure, 2011. **19**(12): p. 1784-95.

85. Zheng, W., et al., *LOMETS2: improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins.* Nucleic Acids Res, 2019. **47**(W1): p. W429-W436.

86. Xu, J. and Y. Zhang, *How significant is a protein structure similarity with TM-score = 0.5?* Bioinformatics, 2010. **26**(7): p. 889-95.

87. Binder, K., et al., *Monte Carlo Simulation in Statistical Physics.* Computers in Physics, 1993. **7**(2): p. 156-157.

88. DiMaio, F., et al., *Refinement of protein structures into low-resolution density maps using rosetta.* J Mol Biol, 2009. **392**(1): p. 181-90.

89. MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* J Phys Chem B, 1998. **102**(18): p. 3586-616.

90. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

91. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.

92. Afonine, P.V., et al., *New tools for the analysis and validation of cryo-EM maps and atomic models.* Acta Crystallogr D Struct Biol, 2018. **74**(Pt 9): p. 814-840.

93. Yang, Z., et al., *Dpy30 is critical for maintaining the identity and function of adult hematopoietic stem cells.* The Journal of experimental medicine, 2016. **213**(11): p. 2349-2364.

CHAPTER 3.

Mechanism for DPY30 and Disordered Regions in ASH2L to Modulate the MLL/SET1 Activity on Chromatin[3]

3.1 Abstract

Recent cryo-EM structures show the surprisingly dynamic nature of the MLL1-NCP interaction. Functional implication and regulation of such dynamics remain unclear. Here we show that DPY30 and the intrinsically disordered regions (IDRs) of ASH2L play important roles in restricting the rotational dynamics of the MLL1 complex on the NCP. We show that DPY30 binding to ASH2L leads to drastic changes in ASH2L IDRs and stabilize multiple new contacts at the MLL1-NCP interface. Both ASH2L IDRs and DPY30 are required for the dramatic increase of processivity and activity of the MLL1 complex. This regulation is NCP-specific and applies to all members of the MLL/SET1 family of enzymes. We further show that DPY30 is causal for *de novo* establishment of H3K4me3 in ESCs. Our study provides a new paradigm of how H3K4me3 is regulated on chromatin and how H3K4me3 heterogeneity can be modulated by ASH2L IDR interacting proteins.

[3] The contents of this chapter were adapted from a published second authored manuscript: Lee, Y. T., **Ayoub, A.**, Park, S. H., Sha, L., Xu, J., Mao, F. B. A., Zheng, W., Zhang, Y., Cho, U. S., and Dou, Y., "Mechanism for DPY30 and ASH2L intrinsically disordered regions to modulate the MLL/SET1 activity on chromatin," *Nature Communications,* **12**, 2953 (2021)

3.2 Introduction

Cells are complex, information-processing centers that handle an immense flow of signals often leading to fine tuning the expression of genes. To achieve exquisite regulation, chromatin post-translational modifications (PTMs) have evolved to demarcate, among a mosaic of functions, actively transcribed genes from the inactive ones [4]. The mixed lineage leukemia (MLL) family of histone methyltransferases (HMTs) catalyzes the deposition of histone H3 lysine 4 methylation (H3K4me) associated with active transcription [5, 6]. H3K4 methylation is highly enriched at gene promoters and distal regulatory enhancers, and plays a pivotal role in the recruitment of basal transcription machinery [8-10] and chromatin remodeling complexes [11-13]. It also promotes long-range chromatin interactions and higher order chromatin organization [14-16]. Dynamic interplay between H3K4me and co-transcriptional processes have also been reported [17, 18]. Human genetic studies have corroborated the functional importance of the MLL family enzymes: heterozygous mutations in MLLs are reported in congenital human Kabuki [19-24], Wiedemann-Steiner and Kleefstra spectrum syndromes [25-27]. Furthermore, MLL family proteins are among the most frequently mutated genes in human malignancies[29].

The MLL/SET1 family enzymes interact with several evolutionarily conserved proteins, WDR5, ASH2L, RbBP5 and DPY30, through the C-terminal catalytic SET domain [30-32]. We and others have previously shown that these core components are essential for MLL1 catalytic activity on histone H3 [32-34]. In particular, WDR5 is required to stabilize the trimeric RbBP5-ASH2L-MLL1 complex [35, 36], a role exploited for the development of MLL1-specific inhibitors [37, 38]. In-depth biochemical studies also show that these core components have multiple relatively weak interactions amongst themselves [39-41]. Recently, a co-crystal structure of the minimal trimeric complex (ASH2L$^{SPRY,\Delta 400\text{-}440}$-RbBP5$^{330\text{-}375}$-MLL1/3^{SET}) [35] and cryo-EM structures of the MLL1-NCP (nucleosome core particle) complex [2, 7] have revealed the overall architecture of the MLL1 core complex as well as its engagement with a physiological substrate (i.e. NCP). These studies, together with solution structures of MLL1 [40], show a surprisingly dynamic nature of the MLL1 core complex, especially the MLL1^{SET} domain and the RbBP5-NCP interface. Despite these studies, regulation of structural dynamics of the MLL1 complex on the NCP and its functional implications remain largely unknown.

Compared to the well-studied WDR5, RbBP5 and ASH2L^{SPRY} proteins, the function of DPY30 and the extended intrinsically disordered regions (IDRs) of ASH2L in the MLL1 complex remains a mystery. The biochemically defined minimal core complex showed negligible DPY30 contribution to the activity of the MLL/SET1 family enzymes using free histone H3 or peptidic H3 as substrates [41-43]. On the other hand, DPY30 is capable of regulating global H3K4 methylation *in vivo* [44] and DPY30 knockdown or knockout leads to global reduction of H3K4me3 in embryonic stem cells (ESCs) and hematopoietic stem cells [45, 46]. It is proposed as a potential therapeutic target for MLL1-rearranged leukemia [47]. The conflicting reports of the minimal *in vitro* DPY30 activity versus its importance in H3K4me3 *in vivo* remain unresolved.

Here we show that DPY30 greatly stimulates MLL1 activity on the NCP. By combined NMR, SAXS, cryo-EM and biochemical approaches, we find that DPY30 functions through the extended IDRs of ASH2L to restrict the rotational dynamics of the MLL1 complex on the NCP and thereby promoting H3K4 methylation, especially at higher states. The NCP-specific regulation by DPY30 and ASH2L IDRs generally applies to all MLL/SET1 family enzymes. Cellular studies further confirm importance of DPY30 in *de novo* establishment of H3K4me3 on chromatin. Taken together, we have established a new paradigm of how the disordered regions in the chromatin modifying complexes may exert loci-specific histone methylation and confer heterogeneity in the cellular epigenetic landscape.

3.3 Results

Activity of the MLL family enzymes on the NCP requires DPY30

To examine regulation of the MLL1 methyltransferase activity on the NCP *in vitro*, we performed the HMT assays using either free recombinant histone H3 or NCP as substrates. The overall activity of the MLL1 core complex was much higher on the NCP as compared to the free histone H3 (Figure 3.1a and Figure S3.1a in Appendix B). DPY30 was essential for the drastic increase of H3K4 methylation on the NCP (Figure 3.1a), especially for higher H3K4 methylation states (i.e., H3K4me2 and H3K4me3) at the expense of H3K4me1 (Figure 3.1a). As the control, DPY30 had no effect on MLL1 activity or processivity when recombinant H3 was used as the substrate (Figure 3.1a and Figure 3.1d), consistent with previous studies [41-43]. To test whether DPY30-dependent regulation on the NCP is a general mechanism for all MLL/SET1 family enzymes, we examined H3K4 methylation by MLL2-4 and SET1A/1B in the presence or absence of DPY30. As shown in Figure 3.1b and Figure 3.1d, DPY30 was able to significantly enhance methylation activity of all MLL/SET1 complexes in an NCP-specific manner. Domain mapping confirmed that the dimerization domain (DD, 45-99) of DPY30, which forms a hydrophobic groove that directly interacts with the ASH2L Sdc-DPY30-Interacting domain (SDI, 504-525) [48], was sufficient to stimulate MLL1 activity on the NCP (Figure 3.1c).

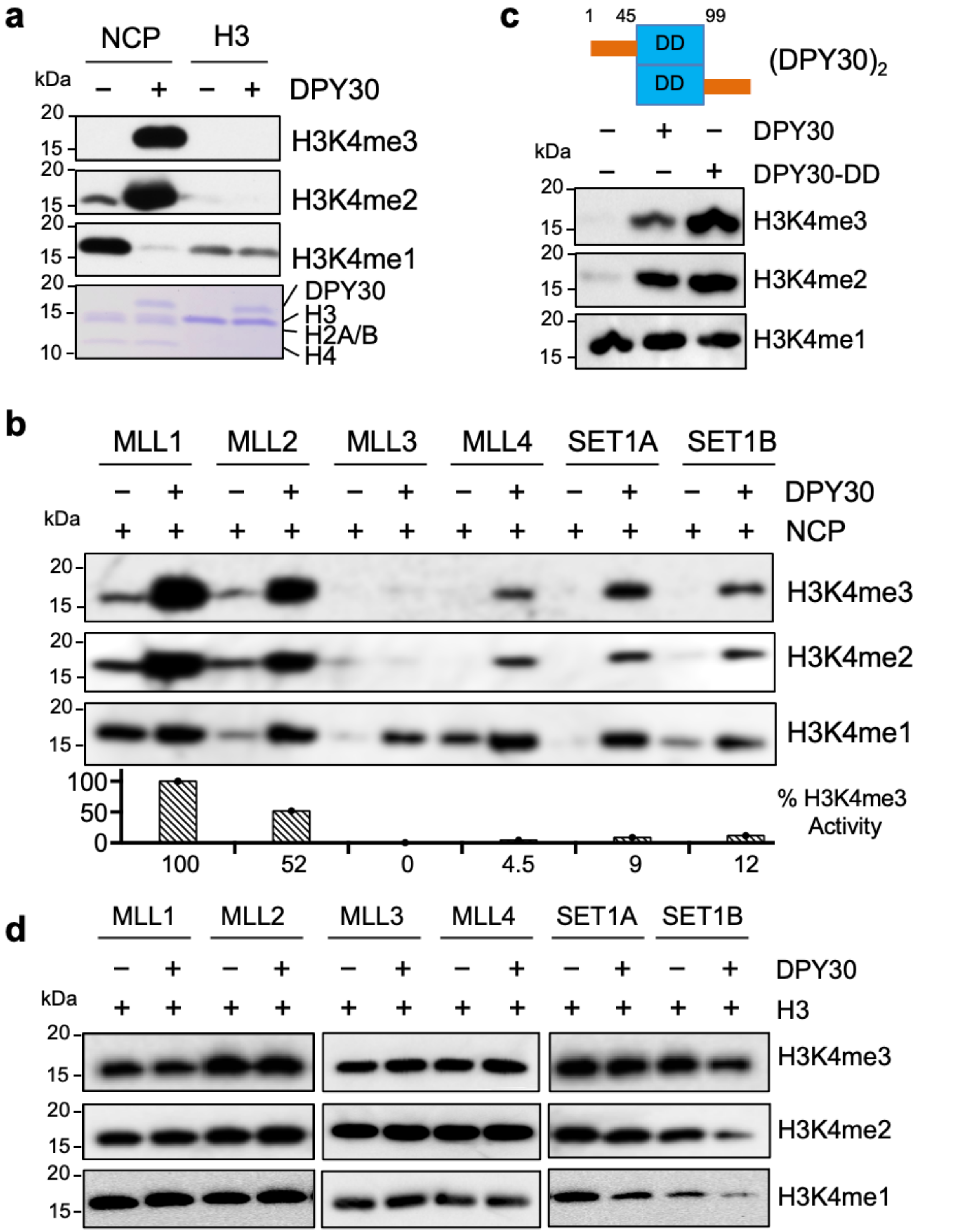


***Figure 3.1*.** DPY30 specifically stimulates MLL1 activity on the NCP. **a**, *In vitro* HMT assay for the MLL1 core complex using either the NCP (nucleosome core particle) or recombinant histone H3 as substrates, which were indicated on top. The MLL1 core complex (i.e., MLL1^{SET}, WDR5, RbBP5 and ASH2L) was added with or without DPY30. Histones were run on 15% SDS-PAGE and blotted with anti-H3K4me1, H3K4me2 and H3K4me3 antibodies as indicated at right. The Coomassie gel was included as the loading control at bottom. **b**, *In vitro* HMT assay for the core complexes of the MLL/SET1 family methyltransferases using the NCP as the substrate. The MLL/SET1 core complexes were added with or without DPY30 as indicated on top. **c**, Top, Domain structure for the DPY30 dimers. DD, dimerization domain (blue). Bottom, *in vitro* HMT assay for the MLL1 core complex with no, full length DPY30, or dimerization domain only The NCP was used as the substrate in all reactions. Quantification completed using ImageJ [3] with %activity calculated relative to wild-type ASH2L-containing complex. **d**, *In vitro* HMT assay for the MLL/SET1 family histone methyltransferases using recombinant H3 as the substrate. The reactions were carried out for longer time than those on the NCP and the immunoblots were subject to a longer exposure due to weaker enzymatic activity of these enzymes on recombinant histone H3. Antibodies used for immunoblots were indicated on right. Western blot in **a** done by YTL.

DPY30-dependent stimulation requires intrinsically disordered regions (IDR) in ASH2L

The recent cryo-EM studies of the MLL1/3-NCP complexes show that DPY30 does not make direct contacts with the NCP [2, 7]. Consistently, when we tested the binding of the MLL1 complex to the NCP with or without DPY30, DPY30 did not alter MLL1-NCP interaction in a gel mobility shift assay (Figure 3.2a). We next tested whether DPY30-mediated stimulation is redundant with that of H2BK120 ubiquitylation (H2BK120ub), which enhances H3K4 methylation without altering binding affinity of the ySET1 complex to the NCP [28]. As shown in Figure 3.2b, DPY30 was able to further enhance activities of SET1A and MLL1 on the H2BK120ub-containing NCP, suggesting that it probably functions through a mechanism distinct from that of H2BK120ub.

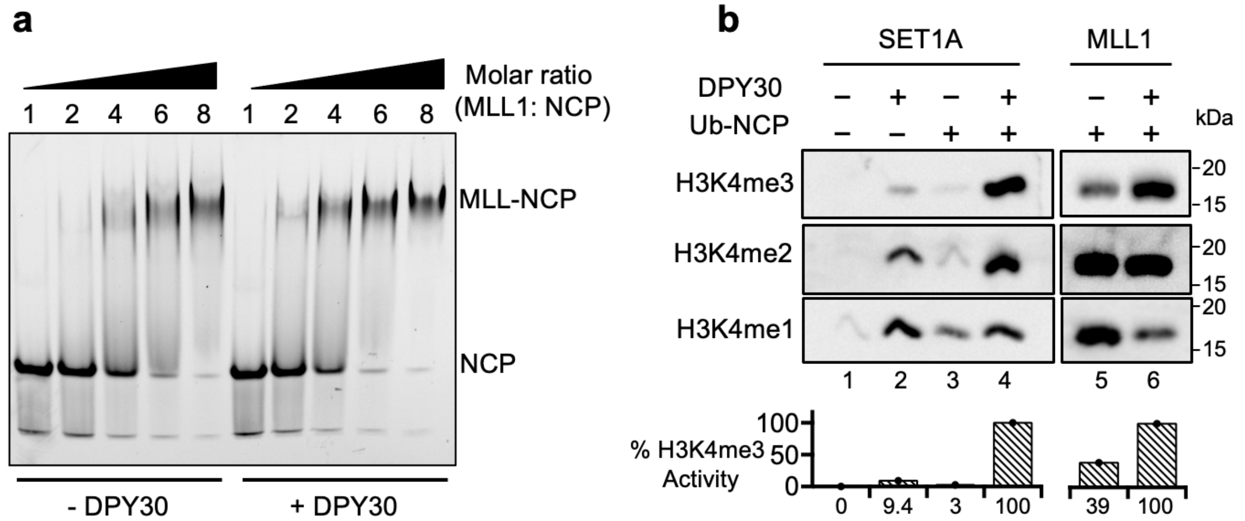


Figure 3.2. DPY30 can stimulate MLL1 activity independent of, and cooperatively with, K120ub. **a** DPY30 does not affect binding of the MLL1 complex to the NCP. Electrophoretic mobility shift assay of the MLL1 core complex binding to nucleosome in the presence or absence of DPY30. Molar ratio of the MLL1 complex to NCP was indicated on top. NCP concentration was 0.4 μM. **b**, DPY30 and H2BK120ub stimulate SET1 and MLL1 activity through distinct mechanisms. *In vitro* HMT assay for the SET1A and the MLL1 core complexes on unmodified NCP (denoted as '- ') or H2BK120ub-NCP ('+') substrates as indicated on top. Quantification of H3K4me3 was done using ImageJ [3] and presented as relative %activity to that of lane 4 and lane 6, respectively. EMSA in **a** done by YTL

As ASH2L is only the direct binding partner of DPY30 in the MLL1 core complex, we examined the role of ASH2L in DPY30-dependent regulation. ASH2L contains the structurally defined N-terminal PHD/WH domain (aa 1-178) [49, 50] and C-terminal split SPRY domain [51] as well as three intrinsically disordered regions (IDRs) (Schematic in Figure 3.3a), including Linker (aa 178-285), Loop (aa 400-440), and SDI (aa 504-525). The SDI of ASH2L directly interacts with DPY30 [48, 52], while both Linker and Loop are IDRs have not been previously characterized. In fact, Loop IDRs were removed in the previous structural studies without disrupting ASH2L SPRY structural integrity [35, 53]. We made selective serial deletions for each of these domains or regions in ASH2L (Schematic in Figure 3.3a) to test their respective contribution to DPY30-dependent stimulation in the *in vitro* HMT assays. As shown in Figure 3.3b, while SDI deletion increased activity of the MLL1 complex, likely by preventing SDI dimerization in solution

[53], it completely eliminated DPY30-dependent stimulation on the NCP (Figure 3.3b, lane 2 versus lane 4). Deletion of PHD-WH-Linker or Loop, but not PHD-WH alone, also abolished DPY30-dependent regulation (Figure 3.3c and 3.3d).

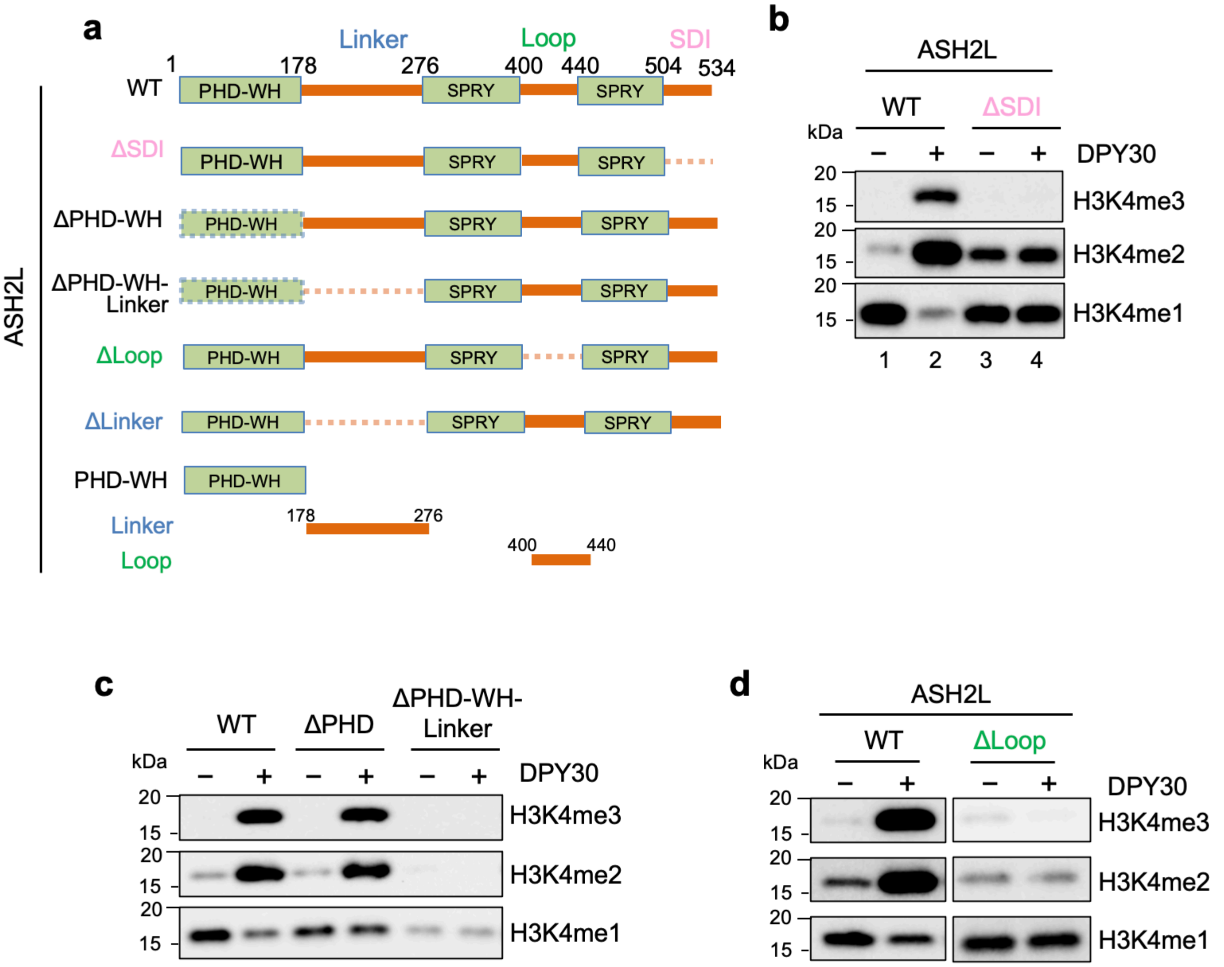


Figure 3.3. DPY30 requires ASH2L IDRs to stimulate MLL1 activity on chromatin. **a**, Human ASH2L truncation and deletion mutants used in the *in vitro* HMT experiments. ASH2L contains two structural domains, the PHD-WH (plant homeodomain-winged helix) domain on the N-terminus and a split-SPRY domain on the C-terminus. It also contains three IDRs, Linker (179-275), Loop (400-440) and SDI (504-525). **b**, *In vitro* HMT assay for the MLL1 core complex with either wild type or ΔSDI ASH2L. **c**, *In vitro* HMT assay for the MLL1 core complex with wild type or ASH2L mutants as indicated on top. **d**, *In vitro* HMT assay for the MLL1 core complex with wild type or ΔLoop ASH2L. For **b-d**, *in vitro* HMT assays were performed with or without DPY30. Equal amount of the NCP was used as substrate in each reaction and histone methylation was detected by immunoblot using antibodies as indicated on right. WBs by YTL

Interestingly, both PHD-WH-Linker and Linker fragments were able to stimulate MLL1 activity in a DPY30-dependent manner *in trans*, albeit at a lower level compared to *cis*-regulation (Figure 3.4a and 3.4b). This property was not shared by Loop IDR in the HMT assay (Figure 3.4b), which cannot function *in trans*.

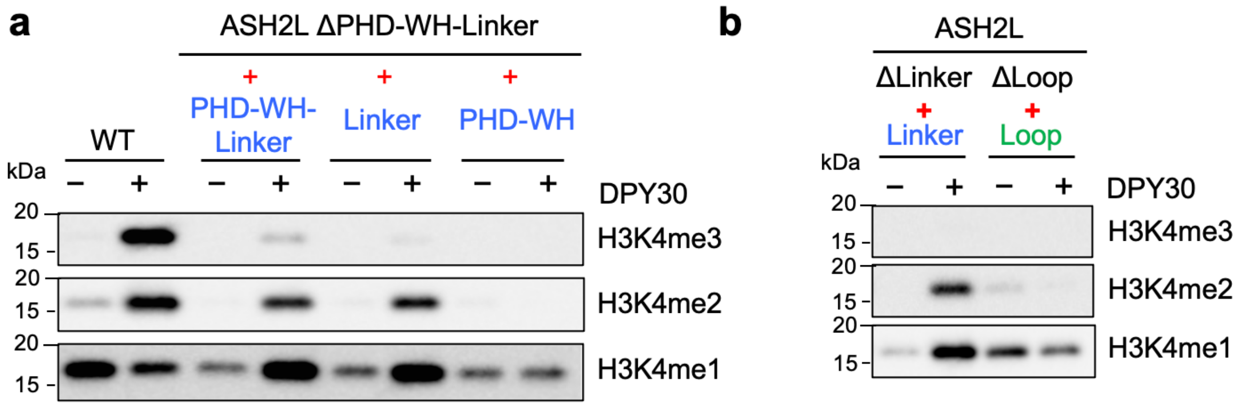


Figure 3.4. ASH2L IDRs can function with DPY30 *in trans*. **a**, *In vitro* HMT assay for the MLL1 core complex with either wild type ASH2L or a mixture of two stoichiometric ASH2L fragments as indicated on top. **b**, Testing the transactivation capability of Linker and Loop IDRs. *In vitro* HMT assay for the MLL1 core complex containing a mixture of Linker and ASH2L ΔLinker polypeptides or Loop and ASH2L ΔLoop polypeptides as indicated on top. For **a** and **b**, *in vitro* HMT assays were performed with or without DPY30. Equal amount of the NCP was used in each reaction and histone methylation was detected by immunoblot using antibodies as indicated on right. Transactivation experiments done by YTL

Furthermore, detailed mapping of ASH2L Linker IDRs (Figure 3.5a) identified three highly conserved linker regions, 247-251, 252-263 and 275-285 (Figure S3.2 in Appendix B), that were critical for DPY30-dependent regulation (Figure 3.5b-d). These results highlight a previously uncharacterized function of ASH2L IDRs in regulating MLL1 activity on the NCP.

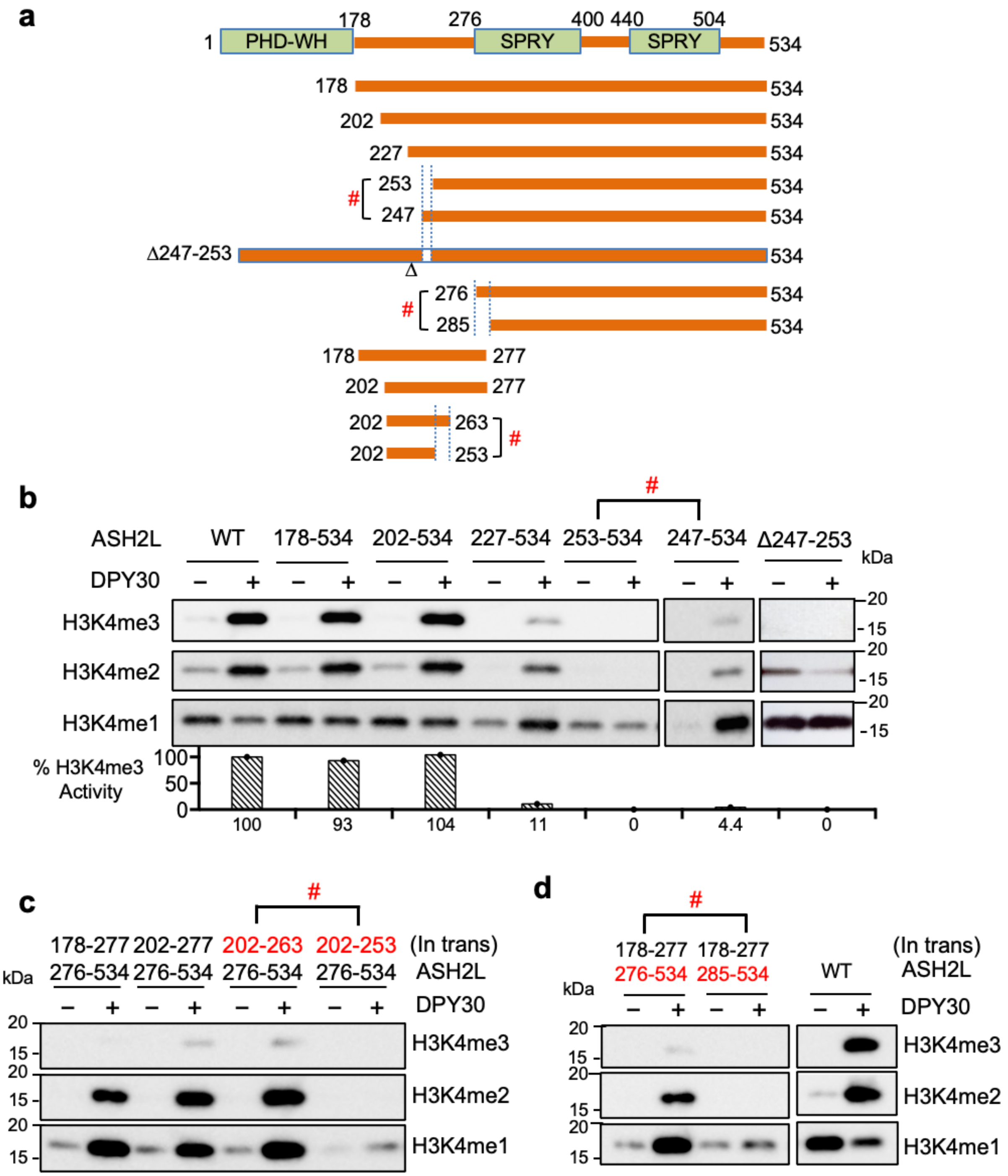


Figure 3.5. Identification of essential ASH2L IDRs in DPY30-mediated stimulation. **a**, Human ASH2L deletion mutants used in **b-d** and transactivation peptides used in **c** and **d.** **b**, Serial deletion to map essential ASH2L Linker IDRs for DPY30 function. *In vitro* HMT assay for the MLL1 core complex containing wild type or various ASH2L mutants as indicated on top. The assays were performed in the presence or absence of DPY30. **c-d**, Trans-activation experiments using two fragments of ASH2L in the *in vitro* HMT assay. The MLL1 core complexes containing a mixture of two stoichiometric ASH2L fragments were used with or without DPY30 as indicated on top**.** #, indicates abolishment of DPY30-dependent activity. Quantification completed using ImageJ [3] with %activity calculated relative to wild-type ASH2L-containing complex. Truncation and transactivation experiments done by YTL

We next examined whether IDRs present in other complex subunits also participate in the DPY30-mediated HMT stimulation. The potential IDRs in the MLL1 core complex include the RbBP5 C-terminus (aa. 382-538) and a segment of the SET domain between the WIN motif and the catalytic domain (aa. 3767-3812) [35]. Sequential C-terminal RbBP5 truncations were tested and none of them abolished DPY30-mediated HMT stimulation (Figure 3.6). Notably, larger deletion of RbBP5 C-terminus lowered the overall HMT activities (Figure 3.6), consistent with previous studies for yeast homolog Swd1 in the SET1 complex [54],[55]. To test the MLL1 SET IDR, MLL$^{SETIL\ (3813\text{-}3969)}$ was used so that the MLL1 complex remains active in the absence of the WIN motif or WDR5 [35]. Removal of the SET IDR did not affect DPY30-dependent stimulation (Figure 3.6). Circumvention of WDR5 in the MLL1$^{SETIL\ (3813\text{-}3969)}$-containing core complex also indicates that WDR5 is dispensable for DPY30-mediated stimulation. These results suggest that ASH2L IDRs are necessary and sufficient for DPY30-dependent HMT stimulation on the NCP.

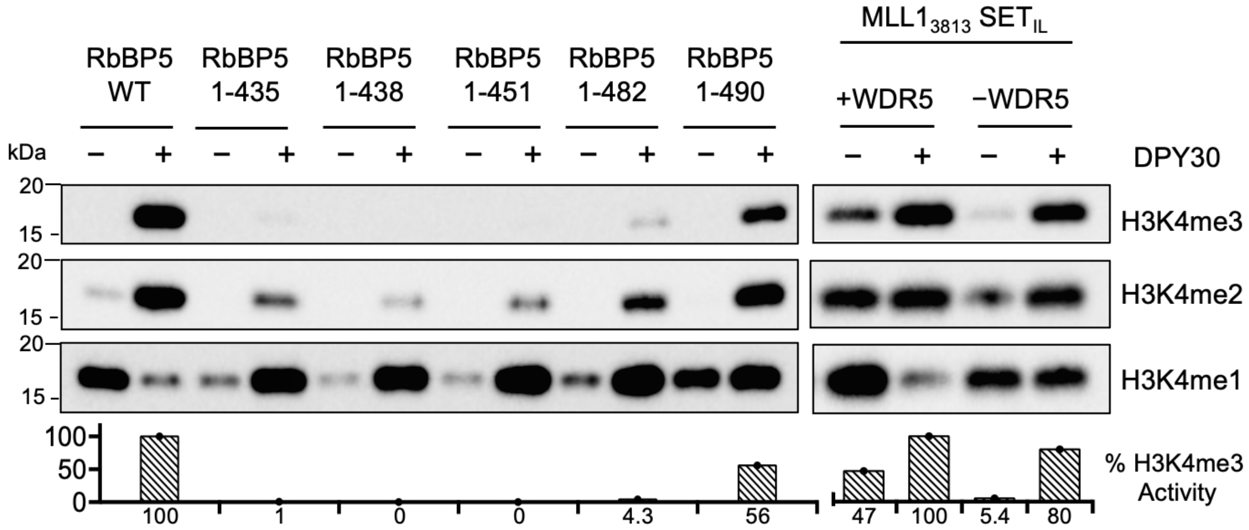


Figure 3.6. IDRs in RbBP5 and MLL1^{SET} are not essential for DPY30-depedent regulation. *In vitro* HMT assay for the MLL1 complex containing RbBP5 truncation fragments (left) or the MLL13813$^{SET\ N3861I/Q3867L}$ (MLL^{SETIL}) with or without WDR5 (right) as indicated on top. RbBP5 has a C-terminal IDR (382-538aa). The assays were performed in the presence or absence of DPY30. Antibodies used in the immunoblots were indicated on right. Quantification of H3K4me3 for reaction containing DPY30 was done using ImageJ [3] and presented as relative %activity to that of MLL1 complex with wild type RbBP5. WBs done by YTL

DPY30 induces widespread changes in ASH2L IDR NMR spectra

To evaluate effects of DPY30 binding on global ASH2L structure and to explore the mechanism by which DPY30 and ASH2L IDRs regulate MLL1 activity, we performed methyl-TROSY NMR on $^{13}CH_3$-labeled Ile-Leu-Val (ILV) ASH2L$^{202-534}$, in the presence of stoichiometric amount of unlabeled RbBP5 peptide (330-363), the minimal region for ASH2L binding (see Methods for details). We identified approximately 65% of the 100 anticipated peaks in $^{13}CH_3$-labeled ILV ASH2L$^{276-534}$ (Figure S3.3a, red in Appendix B). Majority of these peaks were also observed in the $^{13}CH_3$-labeled ILV ASH2L$^{276-534}$ (i.e., without Linker) sample (Figure S3.3a, black in Appendix B). Surprisingly, addition of DPY30 triggered striking and widespread changes in the NMR spectrum, with appearance of many new peaks with significantly dispersed chemical shifts (Figure 3.7 and Figure S3.3b and S3.3c, red peaks, in Appendix B). Chemical shift changes of some apo-state peaks were also observed (Figure 3.7). To further characterize these newly appeared peaks, we carried out residue-specific methyl-assignments by mutagenesis on the ASH2L$^{202-534}$-DPY30 complex (Figure S3.4a-d in Appendix B) [56]. About 60% of total methyl peaks were unambiguously assigned (see Table 3.1 in Appendix B), owing to their dispersed chemical shifts. Interestingly, majority of the DPY30-induced new peaks corresponded to residues in the ASH2L Linker and Loop IDRs (Figure 3.7, blue and orange, respectively). A number of peaks corresponding to residues in the SPRY domain (Figure 3.7, green) were also perturbed (e.g., I274, V287, I300, V322, I488) or newly appeared (e.g., L291, L350). Importantly, deletion of either Linker (blue) or Loop (orange) IDRs in ASH2L (modeled in Figure S3.5 and S3.6 in Appendix B) abolished DPY30-induced changes in NMR spectra (Figure S3.6a-c, right, in Appendix B). The NMR results suggest that DPY30 mainly affects ASH2L IDRs and the DPY30-dependent NMR changes require all ASH2L IDRs.

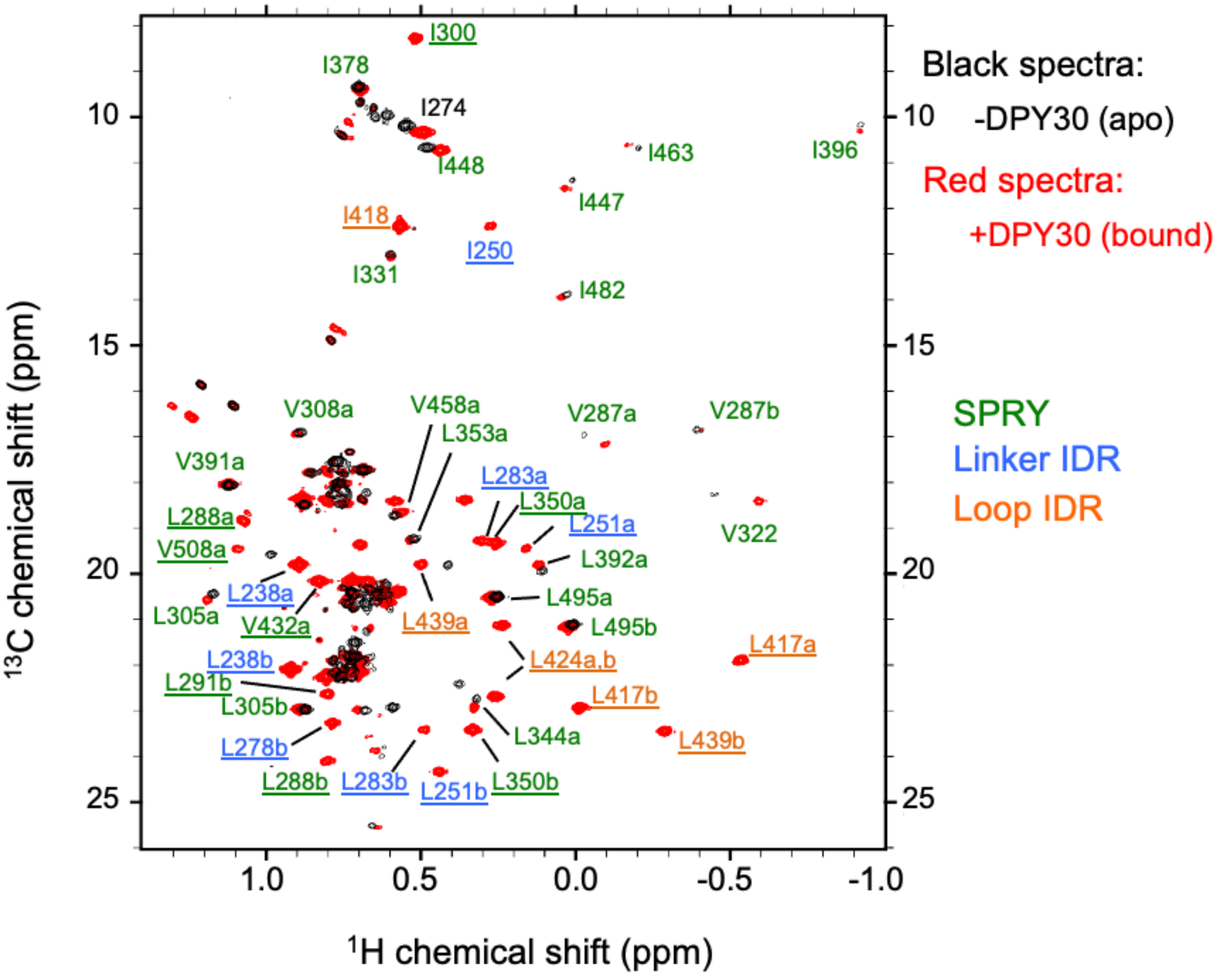


Figure 3.7. ASH2L IDRs undergo significant conformational changes upon DPY30 binding **a**, DPY30 binding induces drastic conformational change in ASH2L. Superimposed methyl-TROSY spectra of [^{2}H, $^{13}CH_3$-ILV] ASH2L$^{202\text{-}534}$ in the absence (black) or presence (red) of DPY30. The labels indicate assigned residues in the DPY30-bound state. Underlined residues are newly appeared peaks upon DPY30 addition. NMR experiments done by YTL

Small-angle X-ray scattering of ASH2L and ASH2L-DPY30 subcomplex

The DPY30-dependent changes of ASH2L IDRs in NMR spectra can be due to alterations of inter-or intra-molecular interactions or stabilization of a particular conformation. To gain more insights into these possibilities, we performed small-angle X-ray scattering (SAXS) experiment for ASH2L, DPY30 and the ASH2L/DPY30 complex. The molecular weight for ASH2L was estimated to be 65 kDa by the SAXS experiment.

Since the combined mass of ASH2L (60.12 kDa) and RbBP5(330-363) (4.07 kDa), which was included in all ASH2L SAXS samples (see Methods), is ~ 64 kDa, ASH2L is likely monomeric in solution. This excludes the possibility that DPY30 functions through resolving intermolecular interactions of ASH2L IDRs. Furthermore, SAXS data shows that pair distance distribution function of ASH2L had a peak around 30 Å and decreased smoothly (Figure S3.7a in Appendix B), suggesting that the structural domains in ASH2L were probably not locked in a rigid configuration. As shown in Figure S3.7a (Appendix B), ASH2L/DPY30 had similar D_{max} (~140 Å) as compared to ASH2L despite a 30% increase in size (Figure S3.7a in Appendix B). It suggests that ASH2L in the DPY30/ASH2L complex is probably in a more compact conformation. Interestingly, analysis using ensemble-optimized method (EOM) [57] identified two distinguishable ASH2L population in both the D_{max} and R_g plots (Figure S3.7b in Appendix B), suggesting that ASH2L is likely in a structural equilibrium between two largely different conformations, with one more extended than the other (Figure S3.7b in Appendix B). We were not able to perform EOM analysis for ASH2L/DPY30 due to method limitation [57]. Taken together, we speculate that DPY30 binding may shift the structural equilibrium of ASH2L and stabilizes ASH2L IDRs in a more compact conformation. This is consistent with DPY30-dependent appearance of NMR peaks with well dispersed chemical shifts (Figure 3.7).

Molecular modeling the DPY30-ASH2L complex

While it is challenging to determine the exact conformation(s) of the dynamic ASH2L IDRs in the apo-state, we were able to build a structural model to visualize ASH2L IDRs in the DPY30-bound state. The molecular model of the human ASH2L-DPY30 is based on the co-crystal structure of the ySET1 complex subunits Bre2-Sdc1 (PDB code: 6CHG)

[54] as well as crystal structures of human ASH2L SPRY domain (without Loop IDR, PDB code: 3TOJ) [53] (Figure 3.8, see Method). When we mapped the residues that showed DPY30-dependent chemical shift in the NMR spectra onto this structural model, close spatial proximity of these residues was apparent (Figure 3.8). They clustered together in both IDRs (Figure S3.5a in Appendix B) and SPRY regions (Figure S3.5b in Appendix B). In this model, ASH2L IDRs, the SPRY domain and SDI adopt a compact triangular structural arrangement upon interacting with DPY30 (Figure 3.8).

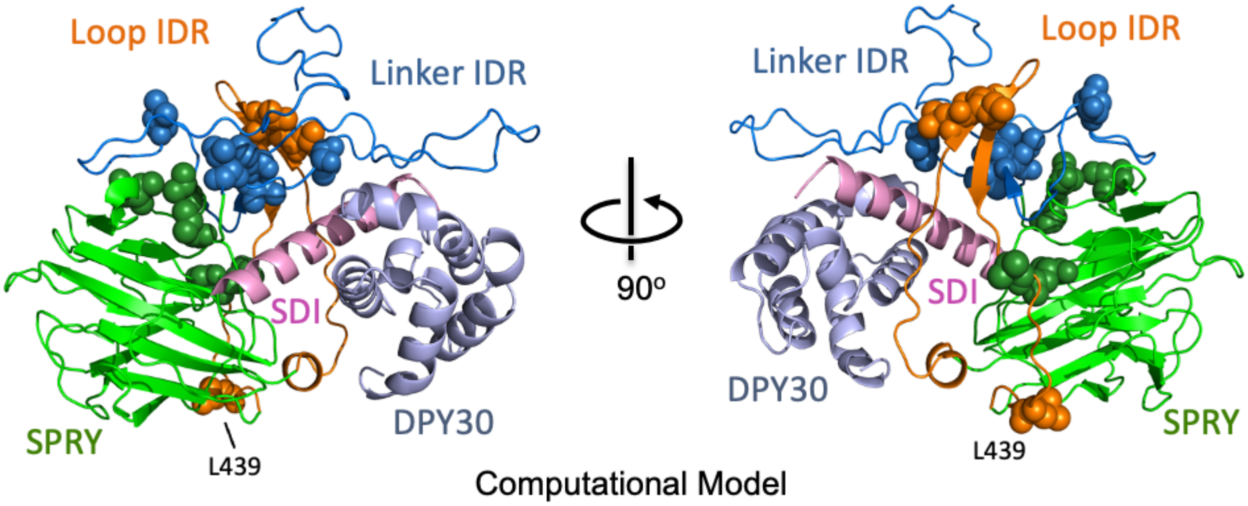


Figure 3.8. Computational model of conformational changes in ASH2L IDRs upon DPY30 binding. Computational model for ASH2L IDRs after DPY30 binding. Underlined residues in a. are presented as spheres. These residues clustered together into a compact structure in this model. For both a and b, SPRY domain is shown in green, Linker IDR is shown in blue, Loop IDR is shown in orange. Model by WZ

ASH2L IDRs form an ordered three-strand β-sheet, comprised of highly conserved residues 247-252 from Linker IDR and residues 416-428 from Loop IDR (Figure S3.8a, red box, in Appendix B). In addition to the β-sheet structure, residues 252-263 and 275-286 of the Linker IDR also adopt a β-sheet-like conformation next to SDI (Figure S3.8a, blue box, in Appendix B), enclosing a binding interface for the α-helical SDI (orange) and DPY30 (Figure S3.8b in Appendix B). Although this is only a computational model, many highlighted structural elements are essential for DPY30-dependent stimulation in the *in* *vitro* HMT assays. Removal of residues 247-253 or 400-440 completely abolished DPY30-dependent MLL1 regulation *in vitro* (Figure 3.5b and 3.3d). Similarly, deletion of residues 252-263 or 275-285 in ASH2L also reduced DPY30-dependent activity (Figure 3.5c and 3.5d) as well as the DPY30-dependent changes in NMR spectrum (Figure S3.6a-c in Appendix B).

DPY30/ASH2L IDRs restrict the rotational dynamics of the MLL1 complex on the NCP

Recently, we and others have solved the cryo-EM structure of the MLL1-NCP complex [2, 7]. It reveals the overall architecture of the five component MLL1 core complex with the NCP. In the MLL1$^{\text{RWSAD}}$-NCP structure, ASH2L binds to the NCP at DNA SHL7 (Figure 3.9a), which together with RbBP5 at SHL1.5, allows MLL1$^{\text{SET}}$ binds above the nucleosome dyad. To understand the molecular mechanism by which DPY30 regulates MLL1 activity on the NCP, we determined the single-particle cryo-EM structure of the human recombinant MLL1$^{\text{RWSA}}$ complex (4-MLL1), containing four of the five core proteins, i.e., $\underline{\text{R}}$bBP5 (aa 1–538); $\underline{\text{W}}$DR5 (aa 23–334); MLL1$^{\underline{\text{S}}\text{ET}}$ (aa 3762–3969); and $\underline{\text{A}}$SH2L$^{\Delta\text{SDI}}$ (aa 1–504), bound to the NCP (4-MLL1-NCP). Overall, a total of 1,288K particles were picked from 6,242 micrographs collected from 300 keV Titan Krios equipped with the K2 summit direct director (Figure S3.9 in Appendix B). After several rounds of heterogeneous refinement using cryoSPARC [58], we isolated four different subclasses of 4-MLL1 bound to the NCP (Class01, 02, 03, and 05). The best behaving particles were further selected from each subset of the 4-MLL1-NCP images after focused refinement and subsequent 3D classification in RELION (Figure S3.9 in Appendix B) [59]. In the end, we obtained three different subclasses of 4-MLL1-NCP structures (Class 01, 02, and 05, Figure 3.9b-d and Figure S3.9 in Appendix B).

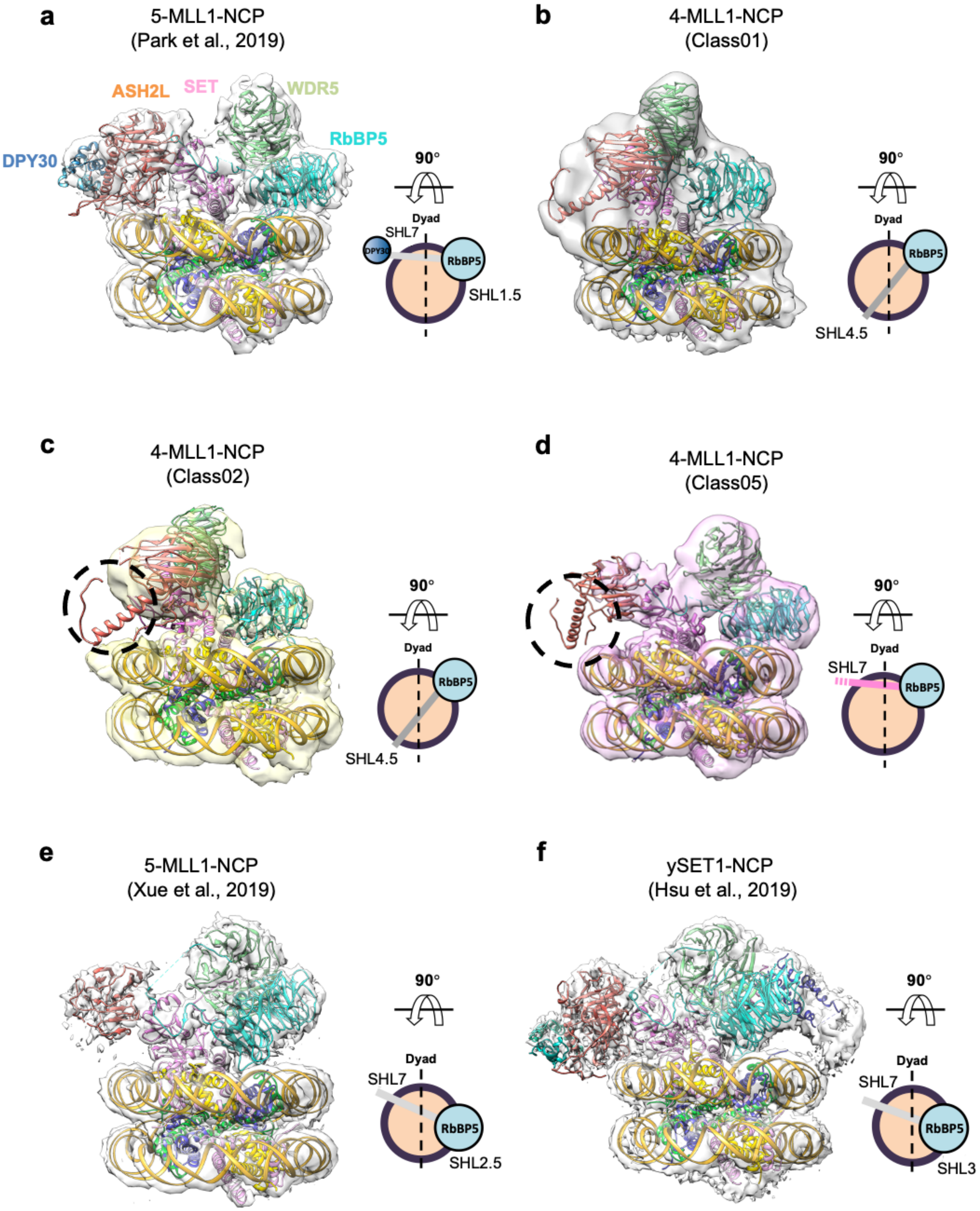


Figure 3.9. Cryo-EM structures of the 4-MLL1-NCP complexes**. a**, Front views of the 5-MLL1^{RWSAD}-NCP (PDB ID: 6PWV and EMDB: EMD-20512) [2]. The 90° top view, on right, shows the relative position of the anchoring points. DNA, yellow; DPY30, blue; ASH2L, orange; WDR5, light green; RbBP5, cyan. The NCP (histone octamer core, orange circle; DNA, black), RbBP5 (cyan circle), and DPY30 (blue circle), and the remaining MLL components were displayed as a gray bar indicating the orientation of the MLL1 complex. **b-d,** Front view of the 4-MLL1-NCP structures. The 90° top view shows its anchoring points on the NCP. The missing EM density of ASH2L IDRs and DPY30 were indicated as black dash-circle in **c-d** and as a dashed line end in the accompanying cartoons. **e-f**, Alternative conformation for the human 5-MLL1-NCP structure [7] and ySET1-NCP [28] as well as their respective top view. ASH2L/Bre2 in both structures interact with the NCP near SHL7 with slight rotational dynamics at the RbBP5/Swd1-NCP contact points. Cryo-EM collected by USC and SHP

The overall resolution of these structures ranged from 4.6 Å to 6.9 Å (Figure S3.10a-c in Appendix B), which were sufficient to dock coordinates of the MLL1 core components and the NCP from our previous MLL1^{RWSAD}-NCP structure (PDB ID: 6PWV) [2]. In comparison to the MLL1^{RWSAD}-NCP complex (or 5-MLL1-NCP, Figure 3.9a) [2], the 4-MLL1-NCP complexes displayed much higher rotational dynamics at the ASH2L-NCP interface (Figure 3.9b and 3.9c). While the majority of the 5-MLL1-NCP complexes anchored on the NCP with RbBP5 and ASH2L at DNA SHL1.5 and 7, respectively, the 4-MLL1-NCP complex adopted multiple modes of interaction. With RbBP5 anchoring near SHL1.5, ASH2L binding sites varied from SHL7 to SHL4.5 among different subclasses (Figure 3.9b and 3.9d). Furthermore, local ASH2L binding dynamics on the NCP also increased significantly in the absence of DPY30, as demonstrated by extremely low or complete loss of ASH2L IDR density in a significant subset of the structures (Figure 3.9c and 3.9d, dashed circle).

The molecular modelling using the iterative template-based fragment assembly refinement (I-TASSER) method [60, 61] showed that ASH2L IDRs make multiple contacts with nucleosomal DNA (Figure 3.10a). In addition to the conserved basic residues ($_{205}$KRK$_{207}$) that contributes to overall MLL1 activity on the NCP [2], DPY30-induced IDR changes may enable ASH2L residues 419-421, which reside on a short loop between the newly formed three-stranded β-sheet, to provide another contact with DNA (Figure 3.10a). Consistent with the modelling, K419A/K421A mutation or deletion of 419-421 significantly reduced or abolished DPY30-dependent regulation of MLL1 activity, respectively (Figure 3.10b). These results suggest that DPY30 probably induces conformational changes in ASH2L to restrict its rotational dynamics on the NCP to promote productive H3K4 methylation (see Discussion).

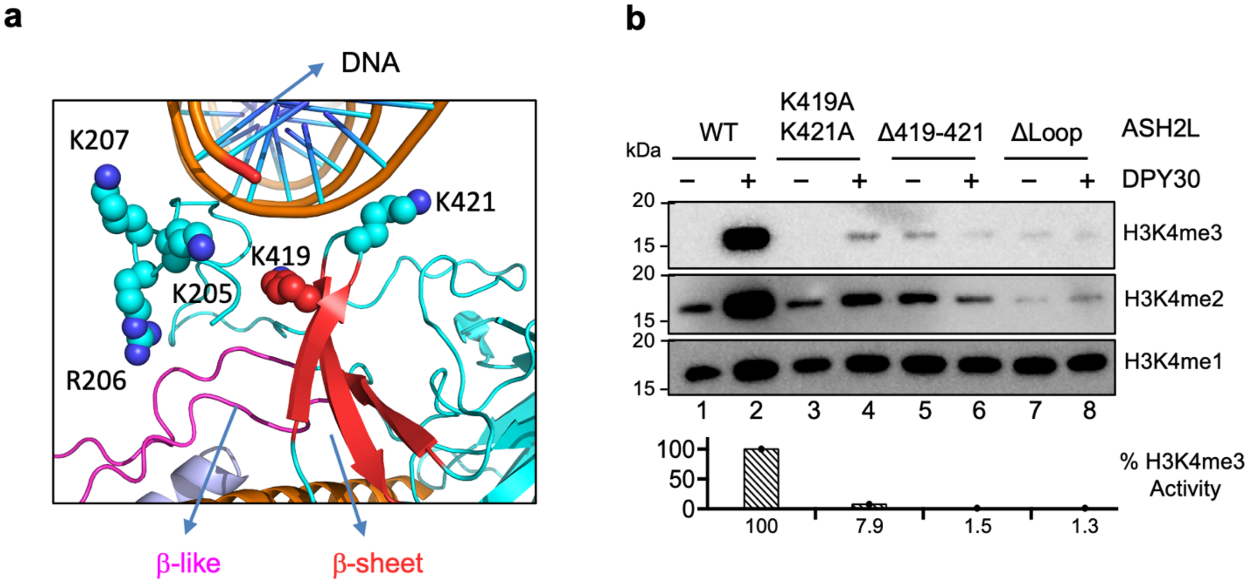
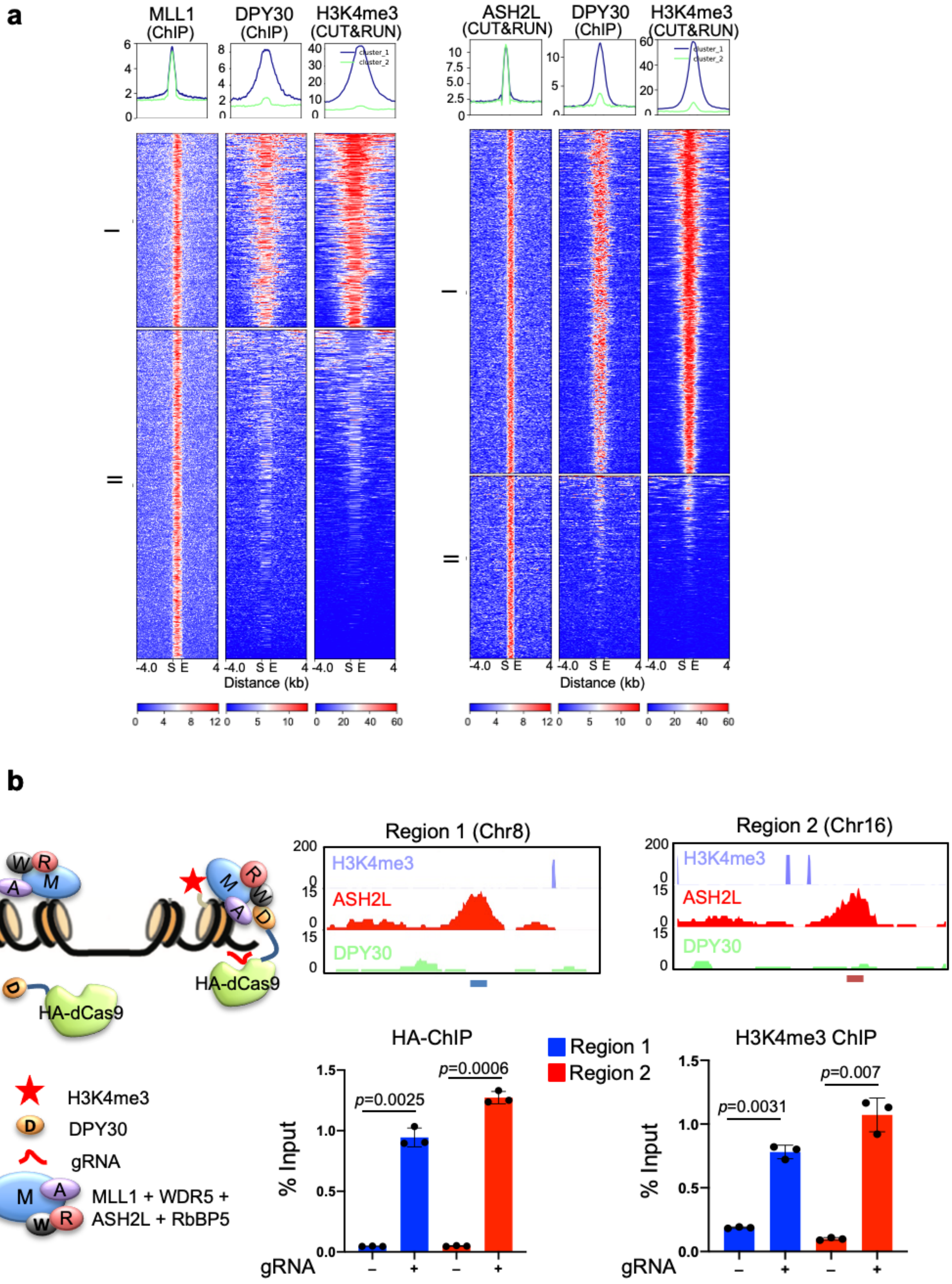


Figure 3.10. Computational model predicts new contacts between ASH2L IDRs and the NCP**. a**, Molecular modeling shows that DPY30-induced conformational change in ASH2L IDRs may enable a short loop ($_{419}$KFK$_{421}$) in the Loop IDR to interact with nucleosomal DNA. Cryo-EM structure of the 5-MLL-NCP structure (PDB ID: 6PWV and EMDB: EMD-20512) [1] was used for the modeling. Red, DPY30-induced β-sheet structure from ASH2L Linker and Loop IDR; pink, β-like structure from ASH2L Linker IDR; orange, DNA (top). In this model, basic ASH2L residues K419 and K421 are positioned near nuclear DNA. **b**, *In vitro* HMT assay for the MLL1 core complex containing wildtype (WT) or mutant ASH2L proteins as indicated on top. The assays were performed in the presence or absence of DPY30. Antibodies for detection of the methylation products were indicated on right. Mutation or deletion of basic residues in Loop IDR, predicted by the model, drastically reduced MLL1 activity on H3K4me3. Quantification for samples containing DPY30 was performed using ImageJ [3] and presented as relative %activity to that of WT ASH2L. Computational model by WZ

DPY30 is essential for establishing de novo *H3K4me3 in E14 embryonic stem cells (ESCs)*

To investigate the function of DPY30 in establishing H3K4me3 in cells, we first examined correlation of DPY30 binding and H3K4me3 at the MLL1 binding sites in E14 ESCs [44, 62]. We identified 4,009 MLL1 peaks in ESCs [62] and among them, 1,070

(26.69%) MLL1 peaks overlapped with those of DPY30 (Figure 3.11a) [44]. Selected loci were shown in Figure S3.11a (Appendix B). Strikingly, H3K4me3 was highly correlated with DPY30 binding at the MLL1 targets (Figure 3.11a). A similar close correlation of DPY30 and H3K4me3 was also found at the 2,431 ASH2L binding sites, 67% of which colocalized with DPY30 at gene regulatory regions in the E14 ESCs (Figure 3.11b and Figure S3.11b and S3.12 in Appendix B). These results showed that MLL1/ASH2L alone were ineffective for depositing H3K4me3 on chromatin. Instead, DPY30 was required for promoting high levels of H3K4me3 on chromatin. Next, we tested whether DPY30 plays a causal role in establishing *de novo* H3K4me3 on chromatin. To this end, we expressed catalytically inactive HA-dCas9 or HA-dCas9-DPY30 in E14 cells and targeted the fusion proteins to randomly selected genomic regions by gRNAs (Figure 3.11b, left). The loci were selected from MLL1/ASH2L joint targets that had no prior DPY30 binding (Figure 3.11b, right top). Upon HA-dCas9-DPY30 recruitment, there was a significant increase of H3K4me3 at these loci (Figure 3.11b, bottom right). In contrast, no increase of H3K4me3 was observed for the no gRNA controls (Figure 3.11b) or in cells expressing HA-dCas9 (Figure S3.11c in Appendix B). These results confirmed that DPY30 is required for *de novo* establishment of H3K4me3 in cells.



a
MLL1 (ChIP)
DPY30 (ChIP)
H3K4me3 (CUT&RUN)
ASH2L (CUT&RUN)
DPY30 (ChIP)
H3K4me3 (CUT&RUN)
cluster_1
cluster_2
I
II
-4.0 S E 4
Distance (kb)
0 4 8 12
0 5 10
0 20 40 60
b
Region 1 (Chr8)
Region 2 (Chr16)
H3K4me3
ASH2L
DPY30
200
0
15
HA-dCas9
W R A M D
H3K4me3
DPY30
gRNA
MLL1 + WDR5 + ASH2L + RbBP5
HA-ChIP
H3K4me3 ChIP
Region 1
Region 2
% Input
p=0.0025
p=0.0006
p=0.0031
p=0.007
0.0
0.5
1.0
1.5
gRNA
–
+
–
+

Figure 3.11. DPY30 regulates *de novo* establishment of H3K4me3 on chromatin. **a**, DPY30 binding is highly correlated with H3K4me3 at the MLL1 binding sites. Heat map for 4,009 MLL1 (left) and 2,431 ASH2L (right) peaks and the corresponding DPY30 and H3K4me3 signals in ESCs. The signal as from merged biological duplicates. MLL1 or ASH2L peaks were clustered with K-means (K=2) using normalized read counts at each peak. Two clusters were highlighted on left. Each row represents a 4 kb region up- and down-stream of the peak center. Peaks were sorted based on normalized read counts in each cluster. **b**, DPY30 is able to establish *de novo* H3K4me3 on chromatin. Left, Experimental design for gRNA-mediated recruitment of dCas9-DPY30. The dCas9-DPY30 is recruited by gRNA to chromatin loci with prior binding of ASH2L and MLL1 and promotes H3K4me3 on chromatin. In the absence of gRNAs, dCas9-DPY30 is not recruited to target loci. W, WDR5; R, RbBP5; A, ASH2L; M, MLL1; D, DPY30. Right top, UCSC browser views of two randomly selected genomic regions are bound by ASH2L, but not DPY30. These regions do not have prior H3K4me3. Regions for gRNAs were highlighted on bottom. Right bottom, ChIP assay for HA-dCas9-DPY30 (left) or H3K4me3 (right) in cells transfected with or without the pooled gRNAs. ChIP signals were normalized against input and presented as %Input. Means and standard deviations (error bars) from at least three independent experiments were presented. Two-sided student *t* test was performed to calculate *p*-value. Bioinformatics by FM; ChIP and C&R by LS; Cas9 experiments by JX

3.4 Discussion

Using the biochemical, structural, and cellular approaches, we have revealed the mechanism by which DPY30 regulates H3K4 methylation activity on chromatin. We show that DPY30 functions through ASH2L IDRs and DPY30-induced changes stabilize ASH2L-NCP interaction and restrict the rotational dynamics of the MLL1 complex on the NCP. Consequently, it promotes productive H3K4 methylation, especially at higher methylation states (i.e., H3K4me3 and H3K4me2). Our study has established a new paradigm by which IDRs, the often-ignored segments in chromatin-interacting proteins, contribute to heterogeneity of the epigenetic landscape in eukaryotic cells.

Previous studies have shown that DPY30 has negligible effects on H3 methylation *in vitro* [41-43], yet its deletion leads to global down regulation of H3K4me3 *in vivo* [44]. Our study shows that DPY30 confers NCP-specific regulation of MLL1 activity by regulating ASH2L-NCP interactions. Using modern computational modeling, with its own limitation and caveats, we show that upon DPY30 binding, ASH2L IDRs converge to adopt a compact structural unit at the MLL1-NCP interface, enabling new contacts with the NCP. In support of the computation model, deletion or mutating ASH2L IDRs greatly impaired

DPY30-dependent methyltransferase activity *in vitro* (Figure 3.5b and Figure 3.10b). The cryo-EM structure of the 4-MLL1-NCP complex shows significant rotational dynamics on the NCP as compared to the 5-MLL1-NCP, 5-MLL3-NCP (Figure 3.9e) or ySET1-NCP complexes (Figure 3.9f) [2, 7, 28, 63]. The 4-MLL1 complex is able to swing across the nucleosome disc with ASH2L binding near SHL4 in a subset of the cryo-EM structures (Figure 3.9b-d). Furthermore, ASH2L also exhibits higher local binding dynamics in the absence of DPY30. We envision that increased rotational dynamics of the 4-MLL1 complex or local ASH2L binding dynamics will destabilize the positioning of the MLL1 SET domain near nucleosome dyad. In this scenario, the MLL1^{SET} domain has to go through multiple spatial arrangements to optimally engage both H3 substrates in the NCP, which negatively affect MLL1 processivity [2]. By limiting rotational dynamics of the MLL1 complex on the NCP, DPY30 as well as ASH2L IDRs promote productive enzyme-substrate engagement and have specific impact on higher methylation states.

Notably, DPY30/ASH2L IDRs regulate all MLL/SET1 family enzymes, regardless of their respective intrinsic activity and processivity (Figure 3.1b). We find that despite its selective impact on global H3K4me3 in cells [44], DPY30 is able to stimulate H3K4me1 by the MLL3 complex *in vitro*. The global reduction of H3K4me3, but not H3K4me1 or H3K4me2, after DPY30 deletion/depletion *in vivo* is probably due to compounding effects of relative abundance and activity of different MLL family enzymes as well as offset of H3K4me1 inhibition by blocking its conversion to higher methylation states. We also would like to point out that DPY30 is able to enhance human SET1 activity on the H2BK120ub-containing NCP (Figure 3.2b). Thus, it can probably cooperate with H2BK120ub in H3K4me3 regulation *in vivo*, which awaits future studies. Compared to

ASH2L, the yeast homolog Bre2 is shorter and devoid of a stretch of basic residues (i.e., $_{205}$KRK$_{207}$) in ASH2L Linker IDRs (Figure S3.2 in Appendix B). However, Bre2 shares most DPY30-induced secondary structures of ASH2L (i.e., β-sheet and β-like structures). It is likely that DPY30-dependent regulation is conserved in the ySET1 complex. The role of DPY30 in countering auto-inhibition [28, 63] of ySET1 remain to be determined.

It is well established that intrinsically disordered proteins (IDPs), or proteins containing extensive IDRs, have unique biophysical properties [64, 65]. The undefined structures in solution enable IDRs to adopt many possible conformations and meaningfully engage in versatile protein-protein interactions [66-68]. As a result, IDRs or IDPs are often found at hubs of protein interaction networks and enable functional diversification and environmental responsiveness during the complex developmental processes [67, 69]. Recent studies also show that IDRs are able to facilitate phase transition and heterochromatin functions in cells [70]. While the exact conformation(s) of apo-state ASH2L IDRs remain to be determined, our study suggests that ASH2L IDRs are probably in a highly dynamic conformational equilibrium and DPY30 binding leads to stabilization of ASH2L IDRs in one of the more structurally organized conformations. The DPY30-induced changes are sufficient to exert locus- and context-specific regulation of H3K4me3 in cells. Our study raises the question of whether ASH2L IDRs can be modulated by other proteins beyond DPY30. We envision proteins that are able to induce perturbations in ASH2L IDRs and/or stabilize ASH2L IDRs could potentially modulate MLL/SET1-NCP interactions, thereby regulating H3K4me activity on chromatin. Aberrant expression of ASH2L has been reported in a wide spectrum of human tumors and contributes to disease progression and prognosis [29, 71-73]. Notably, ASH2L cooperates with activating mutations of Ras in cellular transformation [74], recruits the oncogene MYC to target genes in conjunction with WDR5 [75, 76], and regulates p53 targeting gene expression [77]. Future studies on ASH2L IDR and IDR interacting proteins will provide new insights into regulation of H3K4me3 heterogeneity *in vivo*, and potentially shed light on human pathogenesis.

Finally, histone-modifying enzyme complexes usually contain multiple IDRs in both catalytic and non-catalytic subunits**.** Our survey indicates that IDR content can go up to 70-90% for some histone modifying enzymes (Table 3.3 in Appendix B). Furthermore, 60% of lysine HMTs (HKMTs) contain IDRs of 80 residues or more, whereas only 20% of other annotated proteins have IDRs of similar length [78]. It suggests that IDRs in the histone-modifying enzymes may have especially important regulatory roles, which may constitute a new layer of complexity in epigenetic regulations. Inclusion of the IDRs in enzymes or enzyme complexes may be necessary to discover their regulation to the fullest extent.

3.5 Conclusion

Historic biochemical studies of the MLL1 complex have focused on simple peptidic or recombinant histone H3 substrates where DPY30 had little to no effect on *in vitro* activation [41-43], yet it regulates H3K4me3 *in vivo* [44] and the deletion or knockdown leads of it leads to a global reduction in H3K4me3 in ESCs and hematopoietic stem cells [45, 46]. We extend this foundational understanding by showing DPY30 is indeed essential to establishing *de novo* H3K4me3 in ESCs. We show here that DPY30 is able to biochemically stimulate H3K4me3 formation *in vitro* on a nucleosome substrate. Previous work showed that DPY30 specifically bound to the C-terminal SDI of ASH2L

[48, 52], yet how this impacted the MLL1 complex on the NCP remained unclear. Further, no prior work has fully characterized the extended IDRs of ASH2L structurally or biochemically.

Here, we investigate the role of homodimeric DPY30 binding to ASH2L. We find that DPY30 causes widespread and robust changes to Linker and Loop IDRs by $^{13}CH_3$-labeled ILV methyl-TROSY NMR spectra than in the absence of DPY30, suggesting DPY30 allosterically stabilizes regions of these IDRs. This also caused perturbation of several residues in the well-structured split SPRY domain highlighting the breadth of this effect. Using advanced computational modeling, we showed that the Linker IDR folds on top of itself creating a β-sheet-like structural element near the SDI and, upon deletion of these stretches, significantly perturb the DPY30-mediated changes by NMR. This finding demarcates a previously unknown role for DPY30 in indirectly mediating ASH2L structural compaction by SAXS and broad stabilization of IDR residues upon binding.

Critically, we extended the previous understanding of how MLL1 engages and is stabilized on the nucleosome. Previous work [2, 7] showed that the MLL^{RWSAD} complex engages with ASH2L and RbBP5 at SHL7 and 1.5, respectively, as anchors with the ability to freely rotate between two states as ASH2L as a constant anchor point. In this study, we further expand our understanding of DPY30 in NCP binding. Here, we show that upon the loss of DPY30, the MLL1 complex adopts high rotational dynamics about the nucleosome interface. Instead of ASH2L anchoring at SHL7, RbBP5 anchors the complex down and the ASH2L region freely rotates about the nucleosome. This suggests a critical role for DPY30 in nucleosome binding and stabilization, likely through the stabilization of ASH2L IDRs. Future work would distinguish the precise mechanism for this DPY30-induced anchoring of ASH2L.

3.6 Materials and Methods

Mouse and human ES cell lines

E14tg2a (E14) (ATCC, #30-2002) cell line was used for all cellular experiments. To generate the E14 cell line stably expressing HA-ASH2L, the plasmid expressing ASH2L from the pPiggybac-HA vector as well as plasmids carrying PBase transposase and rTTA element were co-transfected into E14 cells by electroporation. Geneticin was added one day after transfection and selection was carried out for 10 days. Single colonies were picked and screened for stable expression of HA-ASH2L in the presence of Doxycycline.

Protein expression and purification

All MLL1 complex subunits and their mutants were expressed using the pET-28a expression vector with N-terminal 6-histidine and SUMO tag [34]. To make ASH2L mutants for methyl assignments, codon-optimized ASH2L$^{202-534}$ DNA (Integrated DNA Technologies) was used as a template for mutagenesis. Each Ile was changed to Leu, and each Leu or Val was changed to Ile. NEBaseChanger web tool (New England Biolabs) was used to design primers for single residue substitution. Mutant plasmids were constructed using Q5 Site-Directed Mutagenesis Kit (NEB, Cat#E0554S). All proteins were expressed in BL21(DE3) *E. coli* strain in LB media. Cells were grown initially at 37 °C until OD$_{600}$ reached 0.6-0.8 and shifted to 20 °C after IPTG was added at a final concentration of 0.2-0.4 mM. Cells were lysed by sonication and lysates was collected after centrifugation at 32,000 g for 30min at 4 °C. Supernatant was filtered through 0.45

μm syringe filter and purified through a Ni-NTA metal-affinity column (Qiagen and Goldbio). After extensive washing with 20 mM Tris (pH 8.0), 300-500 mM NaCl, 2 mM β-mercaptoethanol and 10 mM imidazole (washing buffer), protein was eluted stepwise at 30, 60, 90, 120, 150, 210, and 300 mM imidazole. SUMO protease was added to the pooled fractions during dialysis at 4 °C overnight. Ni-NTA purification was repeated to remove 6-histidine tag and other bacterial impurities. Proteins were further purified on a HiLoad 16/60 Superdex 75PG or 200PG columns (GE Healthcare). All MLL complex subunits and their mutants are nicely expressed and well-behaved in solution with no noticeable differences in protein stability.

GST-fusion MLL and SET1 proteins

GST-tagged MLL (MLL1^{3745}, MLL2^{2490}, MLL3^{4689}, MLL4^{5319}) and SET1 (SET1A^{1474} and SET1B^{1684}) proteins were expressed using a pGEX-parallel 1 expression vector with N-terminal GST tag and TEV cleavage sequence [36]. Plasmids were transformed and expressed in BL21(DE3) *E. coli* in LB media. Cells were grown until OD_{600} reached 0.6-0.8 when temperature was reduced to 20 ºC and, after temperature equilibration, protein expression was induced using 0.4 mM IPTG and grown for 16 h. Cells were harvested and lysed using sonication and the supernatant was collected by centrifugation at 15,000 rpm, filtered through a 0.45 µm syringe and loaded onto a pre-equilibrated Glutathione Sepharose 4B column (GE Healthcare Life Sciences). After several washes with 20 mM Tris HCl (pH 7.5), 300 mM NaCl, 2 mM DTT, 10% v/v glycerol (GST wash buffer), the protein was eluted off of the column using GST wash buffer with 10 mM reduced glutathione added. Proteins were further purified over a HiLoad 16/60 Superdex 200pg column (GE Healthcare Life Sciences). The purified SET domains remain soluble and stable for the *in vitro* assays.

In vitro histone methyltransferase assay

Mixture of stoichiometric amounts of MLL1 core proteins was used for the *in vitro* HMT assay. Recombinant mono-nucleosome was prepared as described previously [79]. The reaction was carried out in 20 μL of the HMT buffer of 20 mM Tris (pH 8.0), 50 mM NaCl, 5 mM Mg^{2+}, 1 mM DTT and 10 % v/v glycerol as previously described [80]. The reaction was initiated by adding 1 μL of 100 μM *S*-adenosyl-L-methionine and incubated at room temperature for 1hr for the NCP substrates or 4 h for recombinant H3 substrate. The 2x SDS-PAGE sample buffer was added to quench the reaction.

Western blotting

The histones were separated on a 10%-15% polyacrylamide gel and transferred onto polyvinylidene difluoride membrane (Millipore). The membrane was blocked in blocking solution, consisting of 5% milk in 0.1% 1X Tween 20/TBS (TBST), followed by incubation at 4°C overnight with primary antibody in blocking solution. Membranes were washed 3 times in TBST and incubated with the HRP-conjugated anti-mouse/rabbit secondary antibodies at room temperature for 1 h. The membrane was developed using Pierce^{TM} ECL Western Blotting Substrate (Thermo Fisher Scientific, #32106), and images were captured by ChemiDoc^{TM} Touch Imaging System (Bio-Rad). The primary and secondary antibodies included: Rabbit anti-H3K4me1 (Abcam, cat ab8895, 1:20000), Rabbit anti-H3K4me2 (Millipore, cat # 07-030, 1:20000), Rabbit-anti H3K4me3 (Millipore, cat # 07-

473, 1:10000), Rabbit anti-Histone H3 (Abcam, #ab1791, 1:20000) and anti-Rabbit IgG Horseradish Peroxidase-linked whole antibody (GE Healthcare, #NA934, 1:10000).

Preparation of ILV $^{13}CH_3$-labeled ASH2L

The U-[^{2}H] Ileδ1-[$^{13}CH_3$] Leu, Val-[$^{13}CH_3$, $^{13}CH_3$] samples were produced using a previously developed protocol [81] with modifications. Freshly transformed single colony was inoculated into H_2O minimal media containing 6.5 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 120 mg/L $MgSO_4$, 11 mg/L $CaCl_2$, 10 mg/L biotin, 10 mg/L thiamine, 30 mg/L kanamycin, 2 g/L D-glucose and 1 g/L NH_4Cl. Cells were cultured at 37 °C until OD_{600} reaches 0.25 and harvested to remove H_2O media. Then cells were resuspended in D_2O (99.9%, CIL, DLM-4-1000) minimal media containing the same salts in H_2O media in which plain glucose was replaced by D-[^{2}H]-glucose (CIL, DLM-2062). Cells were cultured at 37 °C until OD_{600} reaches 0.7-0.8. The temperature was lowered to 20 °C and 70 mg/L [$^{13}CH_3$, 3,3-^{2}H] α-ketobutyrate (Cambridge Isotope Laboratory, CDLM-7318) and 120 mg/L [3-$^{13}CH_3$, 3,4,4,4-^{2}H] α-ketoisovalerate (CIL, CDLM-73170) were added to the culture. After 1 h, IPTG dissolved in D_2O was added to the final concentration of 0.4 mM. Cells were cultured for another 24 h before harvesting. The labeled proteins were purified using the same protocol as described above. All the NMR samples were buffer exchanged into 25 mM sodium phosphate, pH 6.5, 10 mM NaCl, 0.25 mM d_{10}-dithiothreitol (CIL, DLM-2622), and 1 mM NaN_3 in 99.99% D_2O (Aldrich Cat#151882).

NMR spectroscopy

NMR experiments were carried out on 800 MHz Bruker Ascend spectrometer equipped with pulsed-filed gradient 5 mm inverse triple resonance TXI probe and

SampleCASE with 24 sample slots. IconNMR software was used for automated collection of mutant samples for assignment. All HMQC experiments were acquired at 25 °C. Complex points of 2048 and 256 (^{1}H, ^{13}C) were used for most of experiments except for Ile mutants for assignment, for which 128 complex points in ^{13}C dimension was used. The ^{1}H and ^{13}C carrier frequencies were placed at 4.7 and 17 ppm, respectively. Spectral width was set to 12 and 20 ppm for ^{1}H and ^{13}C dimensions, respectively. A recycle delay of 0.5 s was used with 32-256 scans depending on protein concentration. Residual water was suppressed by the WATERGATE scheme. ^{13}C WALTZ-16 decoupling was employed during acquisition in the direct dimension. All spectra were processed using the NMRPipe program [82]. Gaussian broaden window and sine bell window functions were applied in ^{1}H and ^{13}C dimensions. NMRFAM-Sparky was used to visualize NMR spectra [79].

Small-angle X-ray scattering

All SAXS data were collected at the 18-ID BioCAT Beamline (Biophysics Collaborative Access Team, Advanced Photon Source, Argonne National Laboratory) using the inline SEC-SAXS configuration, in which a flow cell was connected to a ÄKTApure FPLC system (GE Healthcare). About 200 – 500 μL of 1 – 2 mg/mL proteins were injected to a Superdex 200 column (10 x 300 mm, GE Healthcare) pre-equilibrated with 20 mM Tris (pH 7.5), 150 mM NaCl and 1 mM DTT. Flow rate was set to 0.7 mL/min during the data collection. The scattering data was collected every 2s with 1s exposure between 5-24 mL of elution from the SEC column. A short RbBP5 peptide (330-363) was added to the ASH2L samples to reduce aggregation [53, 83]. After data reduction, the strongest scattering data around the sample SEC elution peak were selected for sample scattering. Several data points with minimal scattering near the SEC elution peak were chosen to obtain buffer only scattering. PRIMUS [84] was used for data processing, including averaging scattering data, background subtraction and calculation of the radius of gyration, R_g. and the Porod Volume. The molecular weight was estimated from dividing Porod Volume by 1.6. The pair distribution function was calculated by GNOM [85] in the GUI version of PRIMUS. For EOM analysis, a pool of 10,000 structures of ASH2L with N-terminal PHD-WH and C-terminal SPRY domains connected by the linker and loop was generated by RANCH [57]. GAJOE was used to select an ensemble that best fit the experimental data using a generic algorithm [57].

Molecular Modeling of ASH2L IDRs

Human ASH2L protein consists of two domains, PHD-WH domain and SPRY domain, that have homologous PDB structures, 3S32 (A-chains), 3TOJ, respectively [53, 86]. The crystal structure of yeast Bre2 determined in the COMPASS complex (PDB: 6CHG) [54] contains the Linker and Loop IDRs. The three-dimensional (3D) model for the full-length human ASH2L protein (including PHD-WH domain, Linker-IDR and Loop-IDR regions and SPRY domain) was built by C-I-TASSER [87] using homologous PDB structures above. C-I-TASSER is a recently proposed protein structure prediction pipeline based on the classic I-TASSER protocol [87] with newly developed residue-residue contact predictors [88, 89]. LOMETS [90] threading is performed to align the query sequence to template structures from PDB database to extract continuous fragments. These fragments are used as initial models to assemble into full-length structure by a replica-exchange Monte Carlo (REMC) simulation guided by a composite force field consisting of deep learning-predicted contacts, template-derived distance restraints, and knowledge-based energy terms calculated by statistics of PDB database. The REMC simulation produces a variety of “decoy” conformations, which are then clustered by pairwise structure similarity [91]. The centroid of the largest cluster is refined at the atomic level by FG-MD [92] to obtain the final C-I-TASSER 3D model. The first model generated by C-I-TASSER was selected as the ASH2L model for following analysis. The estimated TM-score of the entire model was 0.67 ± 0.13, indicating that it was a high-confidence model [93]. We removed the PHD-WH domain from the model during the cryo-EM fitting and refinement steps, since there is no density map collected for the PHD-WH domain.

Cryo-EM sample preparation and data collection

The GraFix method [94] was applied to the MLL1^{RWSA}-NCP complex to prepare for application to a cryo-EM grid. In brief, 30 μM of MLL1^{RWSA} was incubated with 10 μM NCP and 0.5 mM *S*-adenosyl-L-homocysteine for 30 min at 4 °C in the GraFix buffer (50 mM HEPES, pH 7.5, 50 mM NaCl, 1 mM $MgCl_2$, and 1 mM TCEP). The sample was centrifuged at 48,000 rpm at 4 °C for 3 h after applying onto a centrifuge tube containing a gradient solution of 0-60% glycerol and 0-0.2% glutaraldehyde. After centrifugation, the crosslinked sample was quenched with 1 M Tris-HCl, pH 7.5. To remove glycerol from the GraFix buffer, we performed further buffer exchange using a centrifugal concentrator (Sartorius Vivaspin 500).

The sample at ~1 mg/ml was applied onto a glow discharged Quantifoil R1.2/1.3 grid (Electron Microscopy Sciences) at 4 °C with 100% humidity. The loaded grid was plunge-frozen in liquid ethane after 4 sec blotting and 30 sec waiting using a Mark IV Vitrobot (Thermo Fisher Scientific). The cryo-EM data was collected using Titan Krios (Thermo Fisher Scientific) operating at 300 keV with the K2 Summit direct electron detector. The movie data were recorded in a counting mode at a 29,000X magnification and the pixel size of 1.01 Å/pixel, with a defocus range between -1.5 to -2.5 μm. A dose rate of 1.28 electrons/Å^2/frame with a total 50 frame per 8 sec was applied for data collection, resulting in a total dose of 64 electrons per Å^2. A total of 6,242 movies were collected.

Cryo-EM data processing and model refinement

Micrograph movies were aligned with whole-frame and local drift correction using MotionCorr2 [95], and CTF was estimated with CTFFIND4.1 [96]. Micrographs with higher than 4.5 Å of the estimated resolution were further selected, which resulted in 6,137 micrographs. A total of 1,287,771 particles were picked using Warp [97]. The particles were extracted in RELION [59] and imported into cryoSPARC [58] for 2D classification. After excluding bad particles, a total of 1,194,542 particles were subjected to the first round of *ab initio* 3D classification into five classes (Figure S3.9 in Appendix B). Two of five classes were subjected to the second round of *ab initio* 3D classification into five subclasses, and the subsequent heterogeneous refinement was performed. Four of the five subclasses displayed a well-defined map of the MLL and nucleosome complex after the heterogeneous refinement. They were exported for 3D classification. The focused 3D classification was performed at the MLL1^{RWSA} region without alignment (35 cycles, T=4, binary mask: 10 pixels/soft mask: 10 pixels). The Class03 was excluded because it displayed a structurally heterogeneous and unresolvable EM density even after the focused 3D classification. The best behaving class selected from Class01 (13,086 particles), Class02 (27,730 particles), and Class05 (23,236 particles) was subjected to the 3D auto refinement and further post-processed to a resolution of 6.9, 4.6, and 6.0 Å, respectively. Each final cryo-EM map was locally filtered to avoid over-estimation (Figure

S3.10d in Appendix B). The resolution of all structures was estimated by RELION with Fourier shell correlation (FSC) at the criteria of 0.143.

For the model building, the rigid-body fitting was performed for each class using Chimera [98]. The cryo-EM structure of MLL1^{RWSAD}-NCP (PDB ID: 6PWV) [2] was used for the rigid-body fitting for each individual class. For the model refinement, each class was subjected to the real-space refinement using PHENIX [99], and model validations were performed by MolProbity [100]. Statistics for data collection, refinement, and validation were summarized in Table 3.2 (Appendix B).

ESC culture and transfection

E14 ESCs were grown in the KnockOut™ DMEM medium containing 15% FBS, 2 mM glutamine, 1X non-essential amino acids, 0.1 mM 2-mercaptoethanol and 10^3 U ml^{-1} LIF (Millipore, #ESG1107), unless otherwise indicated. E14 cells were routinely tested for negative mycoplasma contamination using the LookOut® Mycoplasma PCR Detection Kit (Sigma Aldrich, #MP0035) according to the manufacturer's instructions. For expressing dCas9 fusion proteins, E14 ESCs were transfected with pcDNA3-dCas9-HA and pcDNA3-dCas9-DPY30-HA plasmids using Fugene 6 (Promega, Cat# E2691) for 2 days and then selected with G418 (400µg/ml, Gibco, Cat# 10131-035) for 5 days. After selection, the cells were split and transfected with a pool of three pSPgRNA-gRNAs for selected genomic loci. The pSPgRNA-gRNAs were co-transfected with a pBase vector (1:10) that confers puromycin resistance. After 2 days of puromycin selection (1.5µg/ml, Gibco, Cat# A11138-03), the cells were subject to ChIP using anti-HA antibody (Cell Signaling Technology, cat# 3724) and anti-H3K4me3 antibody (Millipore, Cat# 07-473), respectively. ChIP-qPCRs were performed to detect the enrichment of H3K4me3 and HA in each location.

CUT&RUN

CUT&RUN was performed according to the protocol described previously [101]. HA-ASH2L E14 and the parental E14 cell lines were cultured in presence of 1 μg/mL Doxycycline for 2 days. Biological duplicates were performed for HA-ASH2L and H3K4me3. For each experiment, 1x10^6 cells were harvested, washed with wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1X protease inhibitor cocktail) and incubated with Concanavalin A-coated beads (Bangs Laboratories, Inc. #BP531) for 15 min with rotation. Bead-bound cells were resuspended in solution (digitonin/wash buffer) and incubated with anti-HA (Cell Signaling, #3724) or anti-H3K4me3 (Millipore, #07-473) antibodies overnight at 4°C. The beads were washed with digitonin/wash buffer three times before adding protein A-MNase (0.5 ng/μL) and incubating for 1 hr. at 4°C. Following three washes, bound protein A-MNase was activated on ice for 30 min by addition of 3 mM $CaCl_2$. The reaction was quenched with equal volume of 2X stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.02% digitonin (EMD Millipore #300410), 50 μg/mL RNase A (QIAGEN #19101), 50 μg/mL glycogen (Roche #10901393001), 2 pg/mL Drosophila spike-in DNA at 37°C for 30 min. The proteins were removed by incubating with 0.1% SDS and 0.15 mg/mL Proteinase K (Roche 3115879001) at 65°C for 2 h. DNA fragments were purified by phenol-chloroform and ethanol precipitation and subjected to library preparation. The sequencing was performed at University of Michigan Advance DNA Sequencing Core.

ChIP analysis and quantitative real-time PCR (qPCR)

The ChIP experiment was performed as previously described [80]. Specifically, E14 cells expressing dCas9 fusion proteins were transfected with or without pooled gRNAs (4~5 gRNAs for each selected region) prior to the experiment. Cells were crosslinked with 1% paraformaldehyde at room temperature for 10 min and quenched by adding 250 mM glycine. After two washes with cold 1X PBS, cells were lysed, and the chromatin was sonicated for 3 times for 20 min each using Diagenode Bioruptor 300 for 3 rounds of 20 cycles with 30” on/off per cycle. Supernatant of the sonicated lysate was diluted with 5 volumes of ChIP dilution buffer (16.7 mM Tris-HCl pH 7.5, 12 mM EDTA, 1.1% Triton X-100, 167 mM NaCl, 0.01% SDS) and incubated with anti-H3K4me3 or anti-HA antibodies at 4 °C overnight. The immune complexes were purified on 30 µl of protein G magnetic beads (Invitrogen, Cat# 10003D) for 2 hr. at 4 °C, followed by three washes with low stringency buffer (50 mM HEPES pH 7.9, 5 mM EDTA pH 8.0, 1% NP-40, 0.2% DOC, 1X PBS) and high stringency buffer (50 mM HEPES pH 7.9, 5 mM EDTA pH 8.0, 1% NP-40, 0.7% DOC, 500 mM LiCl) as well as 2 times washes with Last Wash Buffer (5X TE pH 8.0, 0.3% NP-40). The beads were eluted twice with elution buffer (100 mM NaHCO3, 1% SDS) and reverse-crosslinked at 65 °C overnight. The samples were incubated with RNase A at 37 °C for 30 min, followed by incubation with proteinase K (20 mg/ml) at 45 °C for 1 h. DNA was recovered by phenol-chloroform extraction and ethanol precipitation. Real-time PCR was carried out using Radiant Green 2X QPCR mix (Alkali Scientific, Cat# QS1050) on Bio-Rad Real-time PCR machine. Sequence of gRNAs used in this study listed in Table 3.4 in Appendix B.

3.7 Statistical Analysis and Reproducibility

Bioinformatics Analyses

ChIP-seq data mapping and normalization

ChIP-seq dataset for DPY30 and MLL1 were downloaded from GEO GSE26136 and GEO GSE107406, respectively. Paired-end sequencing reads were trimmed with trim_galore to remove adaptor sequences. We kept reads that were 20 bp or longer after trimming and paired between the mates. All ChIP-seq data were mapped to the mouse mm10 genome by using Bowtie2 (v2-2.2.4) [102] with parameters "-q --phred33 --very-sensitive -p 10". Duplicated reads were removed using SAMtools (v1.5) [103]. The bigwig files for IP/input ratio were generated from BAM files by using deepTools3 (v3.2.1) [104] with command "bamCompare -b1 ChIP-bam -b2 Input-bam --ignoreDuplicates --minMappingQuality 30 --normalizeUsing RPKM --binSize 1 --operation ratio --scaleFactorsMethod None -p 20". BAM files for mapping results were merged using SAMtools and converted to BED format using BEDTools [105]. Peaks were called from bed files using MACS (v 1.4.2) [106] with parameters "-w -S -p 0.00001 -g mm". The input signal was used as the control for peak calling. Heatmap of ChIP-seq signals were visualized using deepTools3.

CUT&RUN peak calling and visualization

HA or H3K4me3 CUT&RUN from two independent biological replicates were initially analyzed in parallel. Paired-end sequencing reads were processed as described above. The resulting alignments, recorded in BAM file, were sorted, indexed, and marked for duplicates with SAMtools [103]. The analysis showed good correlation and signal-noise ratio from replicates. The BAM files for mapping results from the replicates were used for further analysis. The overlapping peaks were merged as the union of all using SAMtools and converted to BED format using BEDTools [105]. Fragments with size less than 120 bp were retained [107] by using subcommand “alignmentSieve” in deepTools3 [104]. Peaks were called from bed files using MACS (v 1.4.2) [106] with parameters “-w -S -p 0.00001 -g mm”. The bigwig files for visualization were generated from MACS. Heatmap of CUT&RUN signals were visualized using subcommand “computeMatrix” and “plotHeatmap” in deepTools3.

Statistical Analysis

Statistical analysis was performed by two-tailed Student’s *t*-test using GraphPad Prism 7.0 software. Data were presented as standard error of the mean (SEM). *p* value of less than 0.05 was considered statistically significant; *p<0.05, **p<0.01, ***p<0.001.

3.8 References

1. Rubin, A.J., et al., *Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks.* Cell, 2019. **176**(1-2): p. 361-376 e17.

2. Park, S.H., et al., *Cryo-EM structure of the human MLL1 core complex bound to the nucleosome.* Nature Communications, 2019. **10**(1): p. 5540.

3. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-5.

4. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.

5. Calo, E. and J. Wysocka, *Modification of enhancer chromatin: what, how, and why?* Mol Cell, 2013. **49**(5): p. 825-37.

6. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.

7. Xue, H., et al., *Structural basis of nucleosome recognition and modification by MLL methyltransferases.* Nature, 2019. **573**(7774): p. 445-449.

8. Vermeulen, M., et al., *Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4.* Cell, 2007. **131**(1): p. 58-69.

9. Tang, Z., et al., *SET1 and p300 act synergistically, through coupled histone modifications, in transcriptional activation by p53.* Cell, 2013. **154**(2): p. 297-310.

10. Lauberth, S.M., et al., *H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation.* Cell, 2013. **152**(5): p. 1021-36.

11. Ruthenburg, A.J., et al., *Multivalent engagement of chromatin modifications by linked binding modules.* Nat Rev Mol Cell Biol, 2007. **8**(12): p. 983-94.

12. Wysocka, J., et al., *A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling.* Nature, 2006. **442**(7098): p. 86-90.

13. Taverna, S.D., et al., *How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers.* Nat Struct Mol Biol, 2007. **14**(11): p. 1025-1040.

14. Phillips, J.E. and V.G. Corces, *CTCF: master weaver of the genome.* Cell, 2009. **137**(7): p. 1194-211.

15. Tang, Z., et al., *CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription.* Cell, 2015. **163**(7): p. 1611-27.

16. Yan, J., et al., *Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers.* Cell Res, 2018. **28**(2): p. 204-220.

17. Sims, R.J., 3rd, et al., *Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing.* Molecular cell, 2007. **28**(4): p. 665-676.

18. Khan, D.H., et al., *Dynamic Histone Acetylation of H3K4me3 Nucleosome Regulates MCL1 Pre-mRNA Splicing.* Journal of Cellular Physiology, 2016. **231**(10): p. 2196-2204.

19. Ng, S.B., et al., *Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.* Nature genetics, 2010. **42**(9): p. 790-793.

20. Paulussen, A.D., et al., *MLL2 mutation spectrum in 45 patients with Kabuki syndrome.* Hum Mutat, 2011. **32**(2): p. E2018-25.

21. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature, 2011. **472**(7341): p. 120-4.

22. Micale, L., et al., *Mutation spectrum of MLL2 in a cohort of Kabuki syndrome patients.* Orphanet journal of rare diseases, 2011. **6**: p. 38-38.

23. Hannibal, M.C., et al., *Spectrum of MLL2 (ALR) mutations in 110 cases of Kabuki syndrome.* Am J Med Genet A, 2011. **155a**(7): p. 1511-6.

24. Kluijt, I., et al., *Kabuki syndrome–report of six cases and review of the literature with emphasis on ocular features.* Ophthalmic genetics, 2000. **21**(1): p. 51-61.

25. Jones, W.D., et al., *De novo mutations in MLL cause Wiedemann-Steiner syndrome.* The American Journal of Human Genetics, 2012. **91**(2): p. 358-364.

26. Mendelsohn, B.A., et al., *Advanced bone age in a girl with Wiedemann-Steiner syndrome and an exonic deletion in KMT2A (MLL).* Am J Med Genet A, 2014. **164a**(8): p. 2079-83.

27. Strom, S.P., et al., *De Novo variants in the KMT2A (MLL) gene causing atypical Wiedemann-Steiner syndrome in two unrelated individuals identified by clinical exome sequencing.* BMC medical genetics, 2014. **15**: p. 49-49.

28. Hsu, P.L., et al., *Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS.* Molecular Cell, 2019. **76**(5): p. 712-723.e4.

29. Rao, R.C. and Y. Dou, *Hijacked in cancer: the KMT2 (MLL) family of methyltransferases.* Nat Rev Cancer, 2015. **15**(6): p. 334-46.

30. Cho, Y.W., et al., *PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex.* J Biol Chem, 2007. **282**(28): p. 20395-406.

31. Cosgrove, M.S. and A. Patel, *Mixed lineage leukemia: a structure-function perspective of the MLL1 protein.* FEBS J, 2010. **277**(8): p. 1832-42.

32. Dou, Y., et al., *Regulation of MLL1 H3K4 methyltransferase activity by its core components.* Nat Struct Mol Biol, 2006. **13**(8): p. 713-9.

33. Wu, L., et al., *ASH2L regulates ubiquitylation signaling to MLL: trans-regulation of H3 K4 methylation in higher eukaryotes.* Mol Cell, 2013. **49**(6): p. 1108-20.

34. Cao, F., et al., *An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain.* PLoS One, 2010. **5**(11): p. e14102.

35. Li, Y., et al., *Structural basis for activity regulation of MLL family methyltransferases.* Nature, 2016. **530**(7591): p. 447-52.

36. Patel, A., et al., *A Conserved Arginine-containing Motif Crucial for the Assembly and Enzymatic Activity of the Mixed Lineage Leukemia Protein-1 Core Complex.* Journal of Biological Chemistry, 2008. **283**(47): p. 32162-32175.

37. Cao, F., et al., *Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia.* Mol Cell, 2014. **53**(2): p. 247-61.

38. Vedadi, M., et al., *Targeting human SET1/MLL family of proteins.* Protein science : a publication of the Protein Society, 2017. **26**(4): p. 662-676.

39. Han, J., et al., *The internal interaction in RBBP5 regulates assembly and activity of MLL1 methyltransferase complex.* Nucleic Acids Research, 2019. **47**(19): p. 10426-10438.

40. Kaustov, L., et al., *The MLL1 trimeric catalytic complex is a dynamic conformational ensemble stabilized by multiple weak interactions.* Nucleic acids research, 2019. **47**(17): p. 9433-9447.

41. Patel, A., et al., *On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex.* J Biol Chem, 2009. **284**(36): p. 24242-56.

42. Haddad, J.F., et al., *Structural Analysis of the Ash2L/Dpy-30 Complex Reveals a Heterogeneity in H3K4 Methylation.* Structure, 2018.

43. Shinsky, S.A. and M.S. Cosgrove, *Unique Role of the WD-40 Repeat Protein 5 (WDR5) Subunit within the Mixed Lineage Leukemia 3 (MLL3) Histone Methyltransferase Complex.* J Biol Chem, 2015. **290**(43): p. 25819-33.

44. Jiang, H., et al., *Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains.* Cell, 2011. **144**(4): p. 513-25.

45. Yang, Z., et al., *The DPY30 subunit in SET1/MLL complexes regulates the proliferation and differentiation of hematopoietic progenitor cells.* Blood, 2014. **124**(13): p. 2025-33.

46. Yang, Z., et al., *Dpy30 is critical for maintaining the identity and function of adult hematopoietic stem cells.* The Journal of experimental medicine, 2016. **213**(11): p. 2349-2364.

47. Shah, K.K., et al., *Specific inhibition of DPY30 activity by ASH2L-derived peptides suppresses blood cancer cell growth.* Experimental Cell Research, 2019. **382**(2): p. 111485.

48. Tremblay, V., et al., *Molecular basis for DPY-30 association to COMPASS-like and NURF complexes.* Structure, 2014. **22**(12): p. 1821-1830.

49. Chen, Y., et al., *Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding.* EMBO Rep, 2011. **12**(8): p. 797-803.

50. Ikegawa, S., et al., *Cloning and characterization of ASH2L and Ash2l, human and mouse homologs of the Drosophila ash2 gene.* Cytogenet Cell Genet, 1999. **84**(3-4): p. 167-72.

51. Roguev, A., et al., *The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4.* The EMBO Journal, 2001. **20**(24): p. 7137-7148.

52. South, P.F., et al., *A Conserved Interaction between the SDI Domain of Bre2 and the Dpy-30 Domain of Sdc1 Is Required for Histone Methylation and Gene Expression.* Journal of Biological Chemistry, 2010. **285**(1): p. 595-607.

53. Chen, Y., et al., *Structure of the SPRY domain of human Ash2L and its interactions with RbBP5 and DPY30.* Cell Res, 2012. **22**(3): p. 598-602.

54. Hsu, P.L., et al., *Crystal Structure of the COMPASS H3K4 Methyltransferase Catalytic Module.* Cell, 2018. **174**(5): p. 1106-1116 e9.

55. Mersman, D.P., et al., *Charge-based interaction conserved within histone H3 lysine 4 (H3K4) methyltransferase complexes is needed for protein stability, histone methylation, and gene expression.* J Biol Chem, 2012. **287**(4): p. 2652-65.

56. Amero, C., et al., *A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies.* J Biomol NMR, 2011. **50**(3): p. 229-36.

57. Bernadó, P., et al., *Structural characterization of flexible proteins using small-angle X-ray scattering.* J Am Chem Soc, 2007. **129**(17): p. 5656-64.

58. Punjani, A., et al., *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination.* Nat Methods, 2017. **14**(3): p. 290-296.

59. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3.* Elife, 2018. **7**.

60. Zhang, Y., *I-TASSER server for protein 3D structure prediction.* BMC Bioinformatics, 2008. **9**: p. 40.

61. Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: a unified platform for automated protein structure and function prediction.* Nat Protoc, 2010. **5**(4): p. 725-38.

62. Zhang, H., et al., *MLL1 Inhibition and Vitamin D Signaling Cooperate to Facilitate the Expanded Pluripotency State.* Cell Reports, 2019. **29**(9): p. 2659-2671.e6.

63. Worden, E.J., X. Zhang, and C. Wolberger, *Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome.* eLife, 2020. **9**: p. e53199.

64. Haynes, C., et al., *Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes.* PLoS Comput Biol, 2006. **2**(8): p. e100.

65. Kim, P.M., et al., *The role of disorder in interaction networks: a structural analysis.* Mol Syst Biol, 2008. **4**: p. 179.

66. Oldfield, C.J. and A.K. Dunker, *Intrinsically disordered proteins and intrinsically disordered protein regions.* Annu Rev Biochem, 2014. **83**: p. 553-84.

67. Wright, P.E. and H.J. Dyson, *Intrinsically disordered proteins in cellular signalling and regulation.* Nat Rev Mol Cell Biol, 2015. **16**(1): p. 18-29.

68. van der Lee, R., et al., *Classification of intrinsically disordered regions and proteins.* Chemical reviews, 2014. **114**(13): p. 6589-6631.

69. Dunker, A.K., et al., *Flexible nets. The roles of intrinsic disorder in protein interaction networks.* FEBS J, 2005. **272**(20): p. 5129-48.

70. Gibson, B.A., et al., *Organization of Chromatin by Intrinsic and Regulated Phase Separation.* Cell, 2019. **179**(2): p. 470-484 e21.

71. Bochynska, A., J. Luscher-Firzlaff, and B. Luscher, *Modes of Interaction of KMT2 Histone H3 Lysine 4 Methyltransferase/COMPASS Complexes with Chromatin.* Cells, 2018. **7**(3).

72. Butler, J.S., et al., *Low expression of ASH2L protein correlates with a favorable outcome in acute myeloid leukemia.* Leuk Lymphoma, 2017. **58**(5): p. 1207-1218.

73. Magerl, C., et al., *H3K4 dimethylation in hepatocellular carcinoma is rare compared with other hepatobiliary and gastrointestinal carcinomas and correlates with expression of the methylase Ash2 and the demethylase LSD1.* Hum Pathol, 2010. **41**(2): p. 181-9.

74. Luscher-Firzlaff, J., et al., *The human trithorax protein hASH2 functions as an oncoprotein.* Cancer Res, 2008. **68**(3): p. 749-58.

75. Ullius, A., et al., *The interaction of MYC with the trithorax protein ASH2L promotes gene transcription by regulating H3K27 modification.* Nucleic Acids Res, 2014. **42**(11): p. 6901-20.

76. Thomas, L.R., et al., *Interaction with WDR5 promotes target gene recognition and tumorigenesis by MYC.* Mol Cell, 2015. **58**(3): p. 440-52.

77. Mungamuri, S.K., et al., *Ash2L enables P53-dependent apoptosis by favoring stable transcription pre-initiation complex formation on its pro-apoptotic target promoters.* Oncogene, 2015. **34**(19): p. 2461-70.

78. Lazar, T., et al., *Intrinsic protein disorder in histone lysine methylation.* Biology Direct, 2016. **11**(1): p. 30.

79. Lee, Y.T., et al., *One-pot refolding of core histones from bacterial inclusion bodies allows rapid reconstitution of histone octamer.* Protein Expr Purif, 2015. **110**: p. 89-94.

80. Dou, Y., et al., *Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF.* Cell, 2005. **121**(6): p. 873-85.

81. Tugarinov, V., V. Kanelis, and L.E. Kay, *Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy.* Nat Protoc, 2006. **1**(2): p. 749-54.

82. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.

83. Zhang, Y., et al., *Evolving Catalytic Properties of the MLL Family SET Domain.* Structure, 2015. **23**(10): p. 1921-1933.

84. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis.* Journal of Applied Crystallography, 2003. **36**(5): p. 1277-1282.

85. Svergun, D., *Determination of the regularization parameter in indirect-transform methods using perceptual criteria.* Journal of Applied Crystallography, 1992. **25**(4): p. 495-503.

86. Sarvan, S., et al., *Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain.* Nat Struct Mol Biol, 2011. **18**(7): p. 857-9.

87. Zheng, W., et al., *I-TASSER gateway: A protein structure and function prediction server powered by XSEDE.* Future Gener Comput Syst, 2019. **99**: p. 73-85.

88. Li, Y., et al., *ResPRE: high-accuracy protein contact prediction by coupling precision matrix with deep residual neural networks.* Bioinformatics, 2019. **35**(22): p. 4647-4655.

89. Li, Y., et al., *Ensembling multiple raw coevolutionary features with deep residual neural networks for contact-map prediction in CASP13.* Proteins: Structure, Function, and Bioinformatics, 2019. **87**(12): p. 1082-1091.

90. Zheng, W., et al., *LOMETS2: improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins.* Nucleic Acids Res, 2019. **47**(W1): p. W429-W436.

91. Zhang, Y. and J. Skolnick, *SPICKER: a clustering approach to identify near-native protein folds.* J Comput Chem, 2004. **25**(6): p. 865-71.

92. Zhang, J., Y. Liang, and Y. Zhang, *Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling.* Structure, 2011. **19**(12): p. 1784-95.

93. Xu, J. and Y. Zhang, *How significant is a protein structure similarity with TM-score = 0.5?* Bioinformatics, 2010. **26**(7): p. 889-95.

94. Kastner, B., et al., *GraFix: sample preparation for single-particle electron cryomicroscopy.* Nat Methods, 2008. **5**(1): p. 53-5.

95. Zheng, S.Q., et al., *MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy.* Nat Methods, 2017. **14**(4): p. 331-332.

96. Rohou, A. and N. Grigorieff, *CTFFIND4: Fast and accurate defocus estimation from electron micrographs.* J Struct Biol, 2015. **192**(2): p. 216-21.

97. Tegunov, D. and P. Cramer, *Real-time cryo-electron microscopy data preprocessing with Warp.* Nature Methods, 2019. **16**(11): p. 1146-1152.

98. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

99. Afonine, P.V., et al., *Real-space refinement in PHENIX for cryo-EM and crystallography.* Acta Crystallogr D Struct Biol, 2018. **74**(Pt 6): p. 531-544.

100. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.

101. Skene, P.J., J.G. Henikoff, and S. Henikoff, *Targeted in situ genome-wide profiling with high efficiency for low cell numbers.* Nat Protoc, 2018. **13**(5): p. 1006-1019.

102. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.

103. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-2079.

104. Ramirez, F., et al., *deepTools2: a next generation web server for deep-sequencing data analysis.* Nucleic Acids Res, 2016. **44**(W1): p. W160-5.

105. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-842.

106. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.

107. Skene, P.J. and S. Henikoff, *An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites.* Elife, 2017. **6**.

CHAPTER 4.
Regulation of MLL1 Methyltransferase Activity in Two Distinct Nucleosome Binding Modes[4]

4.1 Abstract

Cryo-EM structures of the KMT2A/MLL1 core complex bound on the nucleosome core particles (NCP) suggest unusual rotational dynamics of the MLL1 complex approaching its physiological substrate. However, the functional implication of such dynamics remains unclear. Here we show that the MLL1 core complex also show high rotational dynamics bound on the NCP carrying the catalytically inert histone H3 lysine 4 to methionine (K4M) mutation. There are two major binding modes of the MLL1 complex on the NCP^{K4M}. Importantly, disruption of only one of the binding modes compromised the overall MLL1 activity in an NCP-specific manner. We propose that the MLL1 core complex probably exists in an equilibrium of poised and active binding modes. The high rotational dynamics of the MLL1 complex on the NCP is a feature that can be exploited for loci-specific regulation of H3K4 methylation in higher eukaryotes.

[4] The contents of this chapter were adapted from a published co-first authored manuscript: **Ayoub, A.***, Park, S. H.*, Lee, Y.-T., Cho, U.-S., and Dou, Y. L. "Regulation of MLL1 Methyltransferase Activity in Two Distinct Nucleosome Binding Modes," *Biochemistry,* doi: https://doi.org/10.1021/acs.biochem.1c00603 (published online in December 2021).

4.2 Introduction

Mixed Lineage Leukemia 1 (MLL1/KMT2A) belongs to the MLL/SET1 family of histone H3 lysine 4 (H3K4) methyltransferases. Heterozygous mutations of the MLL/KMT2 family enzymes are found in multiple human congenital diseases [2]. MLL/KMT2s are also among the most frequently mutated genes in cancer [2]. MLL1 is able to deposit mono-, di- and tri-methylation of H3K4 (H3K4me1/2/3), which demarcates active transcription at gene promoters and distal enhancers and recruits basal transcriptional machinery as well as other chromatin remodeling complexes [3-6]. H3K4me1/2/3 in higher eukaryotic cells has distinct distribution patterns. Interestingly, binding by the MLL family enzymes do not always correlate with H3K4me levels in the genome [7-9]. MLL1-mediated H3K4me1/2 are enriched at distal regulatory regions while H3K4me3 is enriched mostly at gene promoters, despite recruitment of the MLL1 complex to both genomic regions [10, 11]. These observations raise question of how H3K4me heterogeneity is achieved in cells.

Through a highly conserved C-terminal catalytic SET domain, MLL1 interacts with four evolutionarily conserved proteins, i.e., RbBP5, WDR5, ASH2L, and DPY30 (or MLL1^{RWSAD}, MLL1.com) to form a minimal core complex [12-14]. This conserved core complex is essential for efficient catalysis of all levels of H3K4 methylation [14-16], with similar activity to that of the holo-complex [14]. Recently, two groups, including ours, reported the cryo-EM structures of the MLL1 core complex bound to the nucleosome core particles (NCPs) [17, 18]. Interestingly, the two MLL1-NCP structures vary significantly in terms of how the MLL1 complex orients on the NCP. In Xue et al. [18], the MLL1 SET domain resides across the nucleosome disc with RbBP5 and ASH2L binding at the NCP superhelical location (SHL) 2.5 and 7, respectively [18]. This orientation on the NCP is similar to that of the yeast SET1 (ySET1) complex [19, 20], which is constitutively-repressed in the absence of histone H2B K123 ubiquitylation (H2B K120 ubiquitylation in human) [21, 22]. In Park et al., we show that the MLL1 complex adopts a configuration with ASH2L and RbBP5 binding at SHL7 and 1.5, respectively [17]. In this configuration, the MLL1 SET domain sits above the nucleosome dyad, with near symmetric access to both histone H3 tails for optimal processivity [17]. This configuration was also captured as a minor population in Xue et al. [18]. The surprisingly divergent configurations of the MLL1 complex on the NCP raise the question of whether MLL1-NCP interactions are indeed highly dynamic and whether disruption of any of the two binding modes leads to inactivation of the MLL1 complex on the NCP.

Here we solved the cryo-EM structure of the MLL1 complex on the NCP carrying the H3 K4M mutation, which stabilizes the MLL1-NCP interactions. We show that the MLL1 complex exhibits high rotational dynamics on the NCP^{K4M}. More importantly, we show that not all of the MLL1-NCP binding modes are equally active on the NCP. Disruption of only the near dyad MLL1-NCP binding mode leads to compromised overall activity of MLL1 *in vitro*. Our result suggests that the MLL1 complex may bind to the NCP in an equilibrium between active ('dyad') and poised ('cross surface') conformations. We envision that the dynamic feature of the MLL1 complex could be exploited as a regulatory mechanism for higher methylation state of H3K4 in higher eukaryotes.

4.3 Results

MLL1 orients in two distinct modes on the NCP interface

Previous studies show that the H3K4M mutant is able to trap the SET domain in a bound state [23, 24]. To examine the functional significance of two MLL1-NCP binding modes, we decided to use the NCP containing histone H3 lysine 4 to methionine (H3K4M) mutation (NCP^{K4M}) to stabilize the MLL1 interaction (Figure 4.1a). The gel shift assay shows that the MLL1 complex bound more tightly to the NCP^{K4M} as compared to the wild-type NCP (Figure 4.1b).

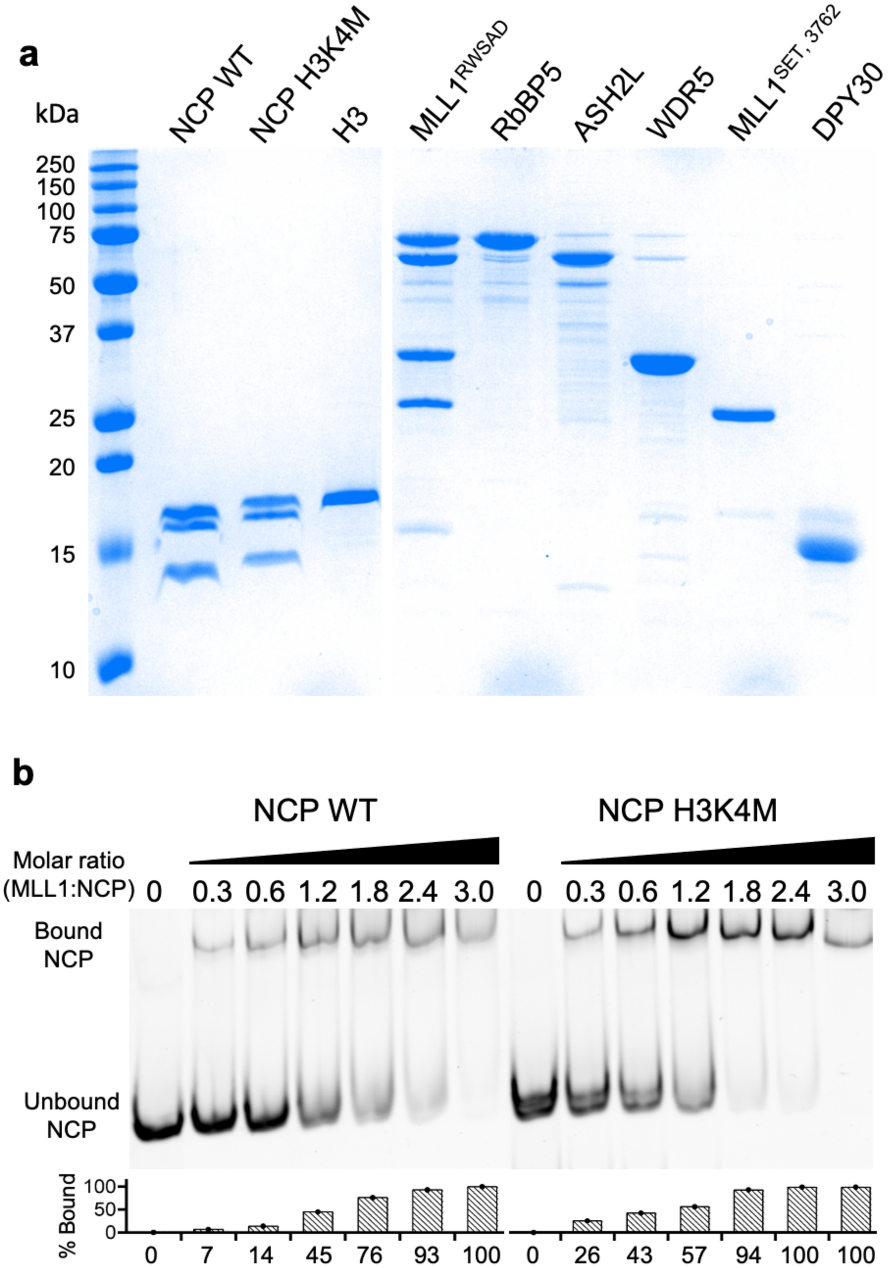
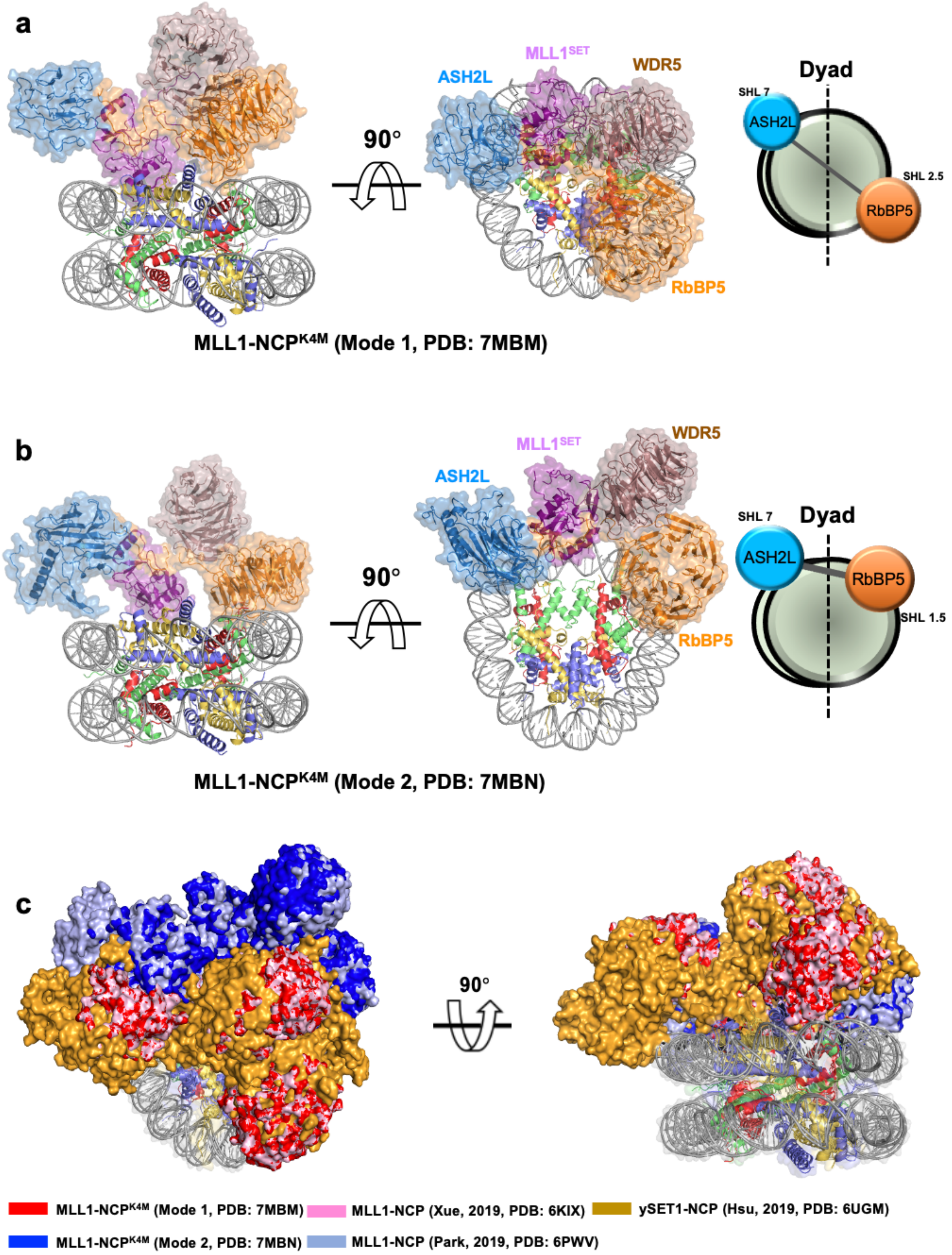


Figure 4.1. H3K4M modulates heterogeneity in NCP binding. **a**, Coomassie brilliant blue staining of substrates and proteins used in this study including wild type NCP, H3K4M NCP, recombinant histone H3, the MLL1 core complex and each individual component, indicated above their respective lanes. **b**, Electrophoretic mobility shift assay comparing relative affinity for wild type NCP (left) and H3K4M NCP (right) for MLL1^{RWSAD} complex. Values represent increasing molar ratio of MLL1.com relative to NCP.

MLL1-NCP^{K4M} binding modes overlay with previous cryo-EM structures

Next, we resolved the single particle cryo-EM structure of the MLL1 complex bound to the NCP^{K4M}. We obtained a total of 808,836 particles of the MLL1-NCP^{K4M} complex from 2,377 micrographic images. It captured two major populations of the MLL1-NCP^{K4M} complex from 3D classification. The structure of these two populations, i.e., mode 1 (29.2 %) and mode 2 (48.5 %), were determined at 4.76 Å and 4.02 Å resolution, respectively (Figure S4.1, S4.2, S4.3 and Table 4.1 in Appendix C). In MLL1-NCP$^{mode\ 1}$, the MLL1 complex binds diagonally across the nucleosome disc, with interactions driven by RbBP5/NCP$^{SHL2.5}$ and MLL1^{SET}/H3/H2A contacts (Figure 4.2a and S4.1a in Appendix C). This structure is highly similar to the previously published MLL1-NCP^{WT} structure by Xue et al. [18] and shares the overall configuration with the ySET1-NCP structures (Figure 4.2c and Figure S4.4b and S4.4c in Appendix C) [19, 20]. In MLL1-NCP$^{mode\ 2}$, the MLL1 complex binds at the edge of the NCP via the RbBP5/NCP$^{SHL1.5}$ and ASH2L/NCP^{SHL7} anchors (Figure 4.2b and S4.1b in Appendix C). This binding mode overlays well with our previously published MLL1-NCP^{WT} structure (Figure 4.2c and Figure S4.4a in Appendix C) [17, 18]. These results confirm that the MLL1 complex displays significant rotational dynamics on both NCP^{WT} and NCP^{K4M}.



MLL1-NCP^{K4M} (Mode 1, PDB: 7MBM)

MLL1-NCP^{K4M} (Mode 2, PDB: 7MBN)

🟥 MLL1-NCP^{K4M} (Mode 1, PDB: 7MBM) ▨ MLL1-NCP (Xue, 2019, PDB: 6KIX) ▨ ySET1-NCP (Hsu, 2019, PDB: 6UGM)

🟦 MLL1-NCP^{K4M} (Mode 2, PDB: 7MBN) ▨ MLL1-NCP (Park, 2019, PDB: 6PWV)

Figure 4.2. The cryo-EM structures of MLL1-NCP^{K4M}. **a**, Front and top view of the MLL1- NCP$^{K4M, mode1}$ structure (PDB: 7MBM, EMDB: 23738). Right, cartoon model to show orientation of the MLL1 complex on the NCP. ASH2L (blue) and RbBP5 (orange) anchor the complex at SHL7 and SHL2.5 of the NCP, respectively. **b**, Front and top view of the MLL1-NCP$^{K4M, mode2}$ structure (PDB: 7MBN, EMDB: 23739). Right, cartoon model to show orientation of the MLL1 complex on the NCP. ASH2L and RbBP5 anchor the complex at SHL7 and SHL1.5 of the NCP, respectively. In both a and b, dyad axis is shown dashed line. **c**, Top (left) and front (right) views of aligned cryo-EM structures of MLL1^{RWSAD}-NCP from Xue et al. (pink, PDB: 6KIX, EMDB: EMD-0694), MLL1^{RWSAD}-NCP from Park et al. (pale blue, PDB: 6PWV, EMDB: EMD-20512), ySET1-NCP from Hsu et al., 2019 (PDB: 6UGM, EMDB: EMD- 20765, goldenrod), MLL1-NCP$^{K4M, mode 1}$ (red) and MLL1-NCP$^{K4M, mode 2}$ (blue). Cryo-EM data collected and processed by SHP

MLL1-NCP mode 1 and 2 contain unique NCP interaction motifs

These two distinct MLL1-NCP^{K4M} binding modes have major differences at the MLL1-NCP interaction interface. In MLL1-NCP$^{mode 2}$, an arginine quartet (R220, R251, R272, and R294; Quad-R), an A-loop ($_{236}$DGEPE$_{240}$), and an I-loop ($_{193}$TGTSNT$_{198}$) in RbBP5 are important for the NCP^{K4M} interaction (Figure 4.3a and S4.4a in Appendix C). They make close contact with the phosphate backbone of DNA as well as core histones H3 and H4, consistent with that of MLL1-NCP^{WT} in Park et al. (Figure S4.5b in Appendix C) [17]. In MLL1-NCP$^{mode 1}$, while ASH2L retains binding near SHL7, RbBP5 rotates clockwise to SHL2.5 (Figure 4.3b). This orientation aligns well with both previously published structures of MLL1^{RWSAD}-K120UbNCP [18] and ySET-K120UbNCP [19] (Figure S4.5a and S4.5c in Appendix C) This rotation partially reorients the Quad-R motif and I-loop, thereby breaking most of their interactions shown in mode 2 (Figure 4.3b and S4.4b). Both R272 and R294 of Quad-R are detached from the DNA backbone interactions. Instead, a small loop ($_{294}$RGE$_{296}$) between WD40 blades 5 & 6 of RbBP5, and R294 in particular, makes a productive charged-charged interaction with E74 of H4 α2 helix (Figure 4.3b and S4.4b, Loop 2). The I-loop is completely displaced from the core histones to reside above DNA at SHL2.5. Since the I-loop lacks positively charged residues, it is unlikely to productively interact with the DNA phosphate backbone. Instead of Quad-R and A-/I-loops, RbBP5 in MLL1-NCP$^{mode 1}$ mainly interacts with α3 and αC helices of H2B through a highly conserved amphipathic loop, $_{248}$LVNR$_{251}$ (Figure 4.3b, Loop 1).

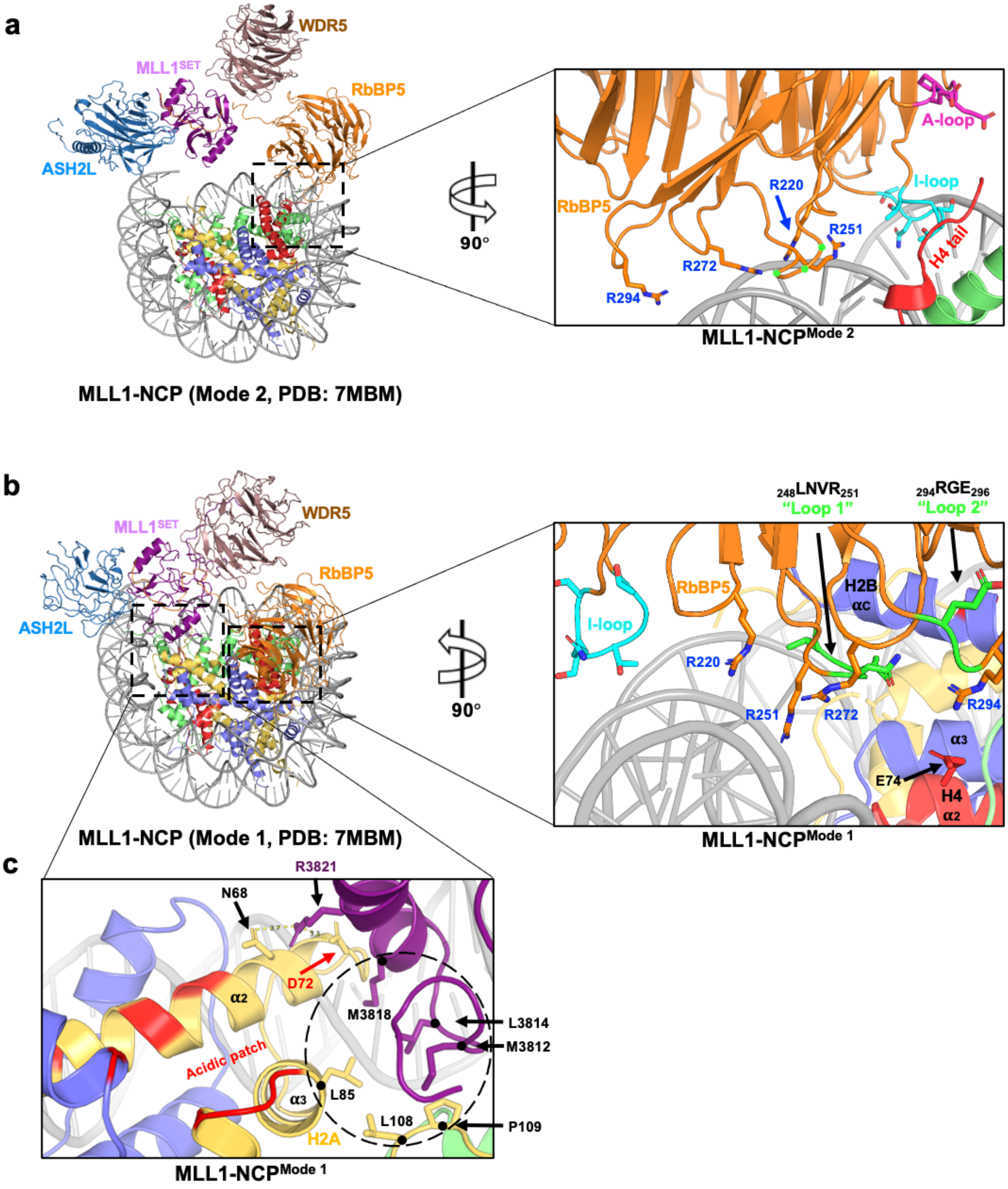


Figure 4.3. Unique motifs in the MLL1-NCP$^{K4M,\ mode\ 1}$ and MLL1-NCP$^{K4M,\ mode\ 2}$ structures**.** **a**, Left, top view of the MLL1-NCP$^{K4M,\ mode\ 2}$ structure. Right, inset shows Quad-R (blue), I-loop (cyan), and A-loop (pink) that are engaged in the NCP interactions. **b**, Left, top view of the MLL1-NCP$^{K4M,\ mode\ 1}$ structure. Right, inset shows that Quad-R (blue) and I-loop (cyan) disengage from the NCP in this configuration. New interactions involving RbBP5 $_{248}$LVNR$_{251}$ (Loop 1, green) are highlighted. **c**, Inset from mode 1 that shows the hydrophobic interface (dashed circle) between MLL1 SET-N 3806-3821 (purple) and the α2, α3 and C-terminal helices (yellow) of H2A. Salt bridge and polar contacts between R3821 (purple) of MLL1 and residues (N68 and D72) from H2A are shown.

The second major difference between two binding modes of the MLL1-NCP^{K4M} complex is the position of MLL1^{SET}. In MLL1-NCP$^{mode\ 1}$, the MLL1 complex rotates clockwise, leading to extensive interactions between MLL1^{SET} and the nucleosome disc. An extended helical patch (3806-21, including M3812, L3814, and M3818) in MLL1^{SET} make hydrophobic interactions with the α2 (N73) and α3 (L85) helices and C-terminus (L108 and P109) of H2A (Figure 4.3c, dashed circled, and Figure S4.4c). There is also a productive electrostatic interaction between an arginine anchor (R3821) of SET-N and D72 of the α2 helix of H2A (Figure 4.3c, red, and S4.4c), a common feature in many protein-NCP complexes [7, 25-28]. In contrast, the SET domain in the MLL1-NCP$^{mode\ 2}$ sits above the nucleosome dyad without making significant contacts with the NCP.

Mode 1 and Mode 2 motifs differentially affect MLL1 activation on NCP

Divergent MLL1-NCP binding modes raise the question of whether they function to activate MLL1 on NCP. We previously showed that mutating key residues in the I-loop, A-loop, or arginine quartet that are important for the MLL1-NCP$^{mode\ 2}$ significantly reduced H3K4 methylation activity in an NCP-specific manner [17]. However, mutation of key residues in MLL1-NCP$^{mode\ 1}$ has not been functionally tested in the histone methyltransferase (HMT) assay [18]. To examine whether these residues are functionally important, we used the *in vitro* HMT assay to assess whether disruption of unique interactions in MLL1-NCP$^{mode\ 1}$ affects MLL1 catalysis on the NCP and recombinant H3. As shown in Figure 4.4a and 4.4b, mutating the highly conserved $_{248}$LVNR$_{251}$ amphipathic loop (Loop 1) or L248/V249 in RbBP5 to alanine (L248A/V249A, LVNR→4A) did not affect MLL1 activity on either the NCP (Figure 4.4a) or recombinant H3 (Figure 4.4b), suggesting that they probably do not contribute to MLL1 activity *in vitro*. Time course analyses also did not show decrease of the MLL1 methyltransferase activity for the RbBP5 LVNR→4A mutant at earlier time points (Figure 4.4c).

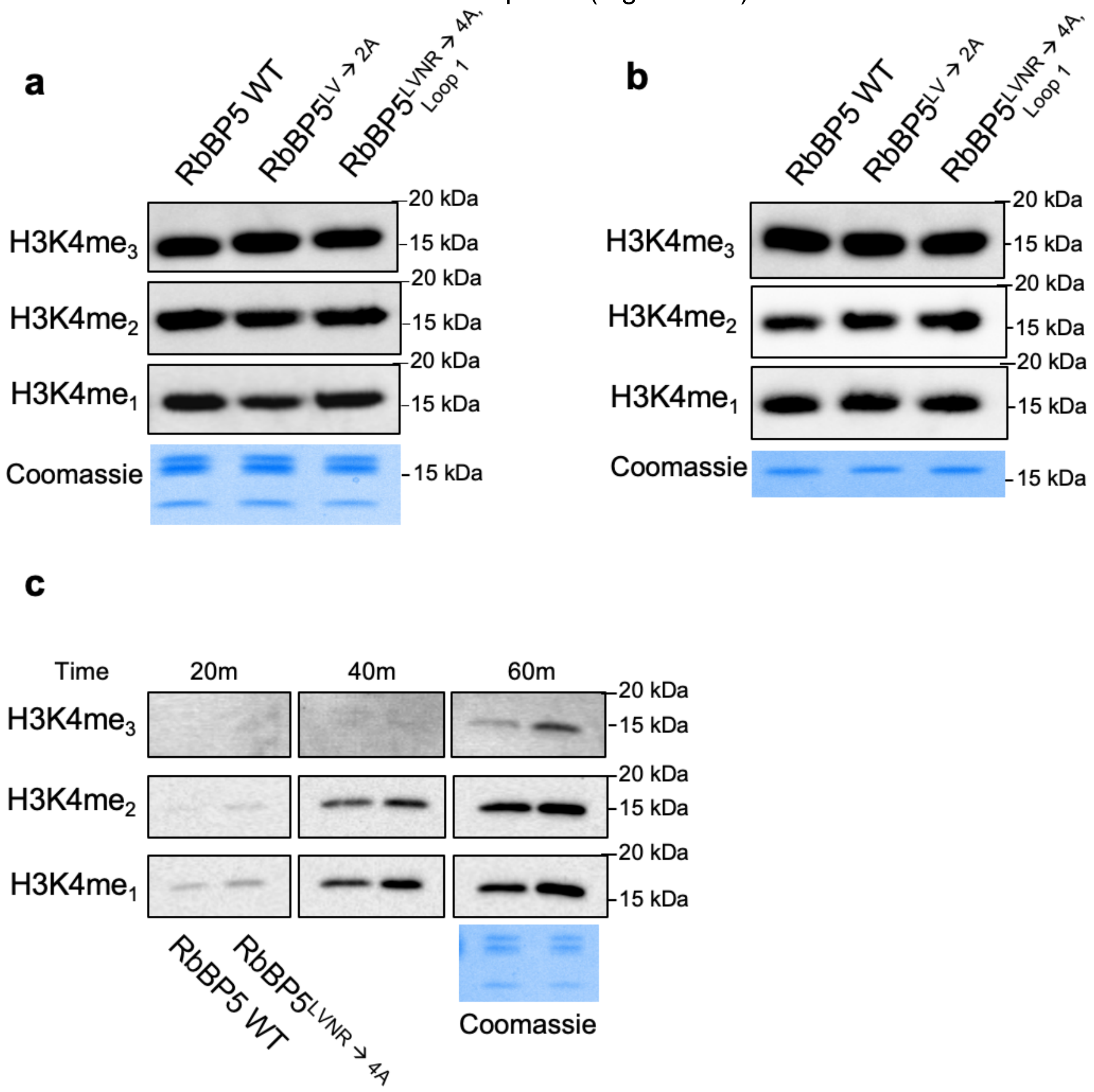


Figure 4.4. MLL1-NCP$^{\text{K4M, mode 1}}$ RbBP5 interaction motifs are dispensable for MLL1 activation**. a-c**, *In vitro* histone methyltransferase assay using **a**, RbBP5 and RbBP5 mutants with NCP as substrate; **b**, RbBP5 and RbBP5 mutants with recombinant H3 as substrate; **c**, a representative methylation time-course for WT MLL1 complex or with RbBP5$^{\text{LVNR → 4A}}$. Antibodies used in the immunoblot were indicated on left. Histone H3 (**b**) and nucleosome (**a** and **c**) loading controls were provided below respective western blots.

Next, we deleted or introduced alanine mutations to key residues of the helical patch (3806-14) in the MLL1^{SET} domain (Figure 4.3c). As shown in Figure 4.5a, deletion of 3806-14 of MLL1^{SET} reduced H3K4 methylation on the NCP (lane 3). Similarly, mutating M3812/L3814 to alanine also significantly reduced H3K4 methylation on the NCP (lane 4), which was partially ameliorated by the triple alanine mutant (3812/14/18A, lane 5) for unknown reasons. However, these MLL1^{SET} mutants showed similarly reduced methyltransferase activity on recombinant H3 (Figure 4.5b), suggesting that they probably function by affecting intrinsic activity of MLL1^{SET}, instead of disrupting MLL1-NCP interactions.

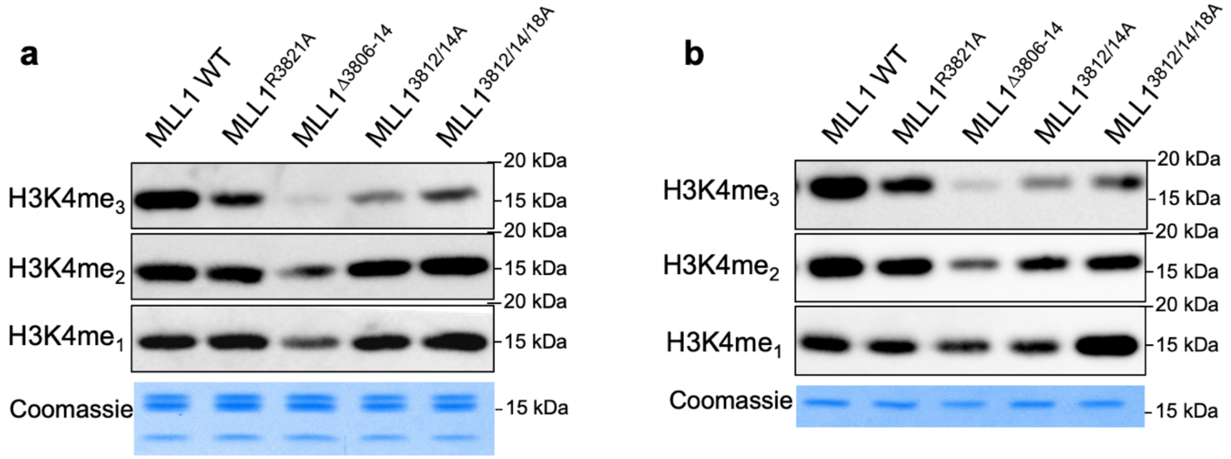


Figure 4.5. MLL1-NCP$^{K4M,\ mode\ 1}$ MLL1^{SET} interaction motifs are dispensable for MLL1 activation on the NCP**.** **a** and **b**, *In vitro* histone methyltransferase assay using **a**, MLL1^{SET} and MLL1^{SET} mutants with NCP as substrate; **b**, MLL1^{SET} and MLL1^{SET} mutants with recombinant H3 as substrate. Antibodies used in the immunoblots were indicated on left. Nucleosome (**a**) and histone H3 (**b**) Coomassie loading controls were provided below respective western blots.

To rule out potential redundant interactions by RbBP5 and MLL1^{SET} in the MLL1-NCP$^{mode\ 1}$, we combined mutations from RbBP5 (i.e., LVNR→4A) and the MLL1^{SET} domain (e.g., helical patch or arginine anchor) and tested them in the *in vitro* HMT assay. We did not observe any additive effect of the combinatorial mutations on the NCP (Figure

4.6a) or recombinant H3 (Figure 4.6b). Finally, we found that mutating the H2A acidic patch (NCP^{APM}) also only modestly affected MLL1 activity on the NCP (Figure 4.6c).

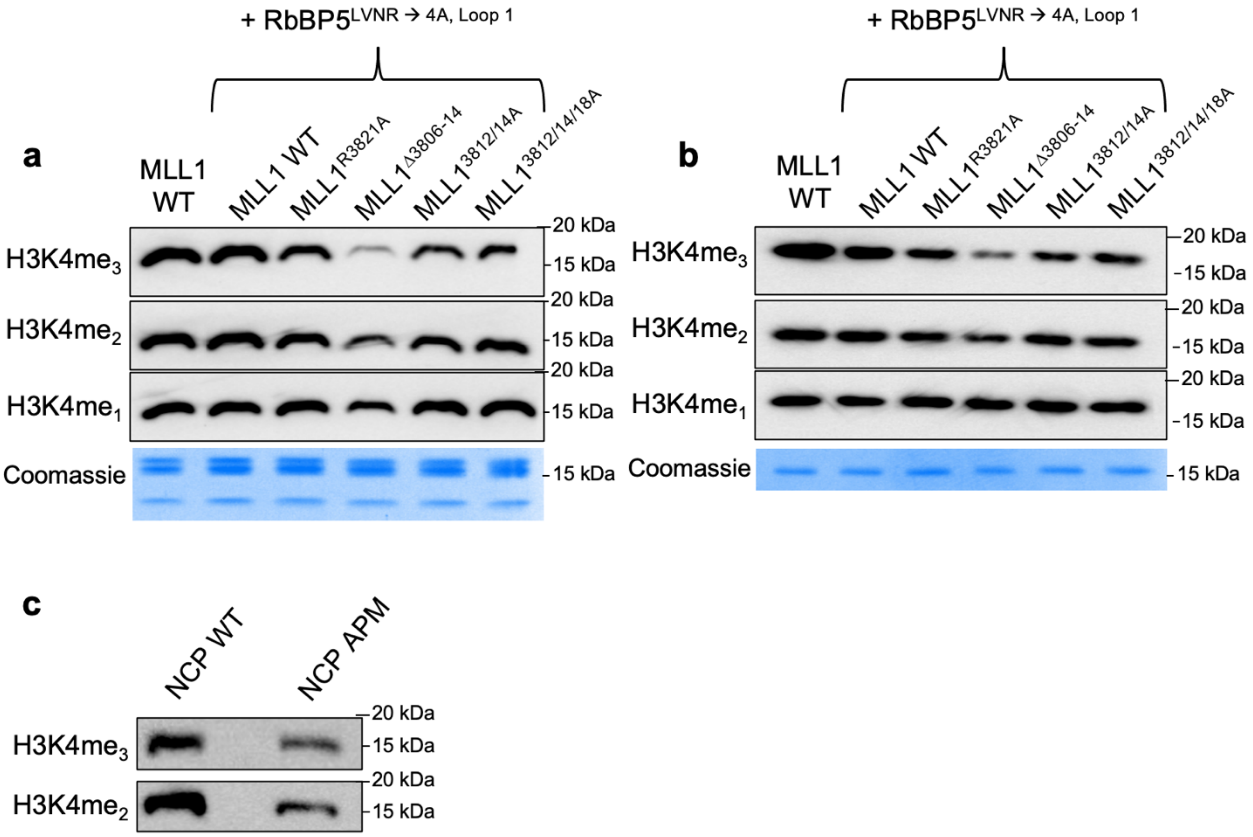


Figure 4.6. Combined MLL1-NCP$^{K4M,\ mode\ 1}$ RbBP5$^{LVNR \rightarrow 4A}$ and MLL1^{SET} interaction motifs do not show combined effects for MLL1 activity**.** **a** and **b**, *In vitro* histone methyltransferase assay using **a**, RbBP5$^{LVNR \rightarrow 4A}$ combined with either wild-type MLL1^{SET} and MLL1^{SET} mutants with NCP as substrate; **b**, RbBP5$^{LVNR \rightarrow 4A}$ combined with either wild-type MLL1^{SET} and MLL1^{SET} mutants with recombinant H3 as substrate. **c**, NCP^{WT} or acidic patch mutant (NCP^{APM}), were used in the reaction. Antibodies used in the immunoblot were indicated on left. Antibodies used in the immunoblots were indicated on left. Nucleosome (**a**) and histone H3 (**b**) Coomassie loading controls were provided below respective western blots. APM WB by YTL

Consistent with previous results [10], we found that MLL1^{SET} alone by electrophoretic mobility shift assay (EMSA) does not bind NCP alone (Figure S4.6a in Appendix C) and ASH2L/RbBP5 are sufficient and necessary for binding NCP (Figure S4.6b in Appendix C). By EMSA, it showed that these mutants cause only modest reduction in NCP binding (Figure 4.7). Taken together, our results show that disruption of unique MLL1-NCP$^{mode\ 1}$

interactions do not significantly affect the activity of the MLL1 complex in an NCP-specific manner *in vitro*.

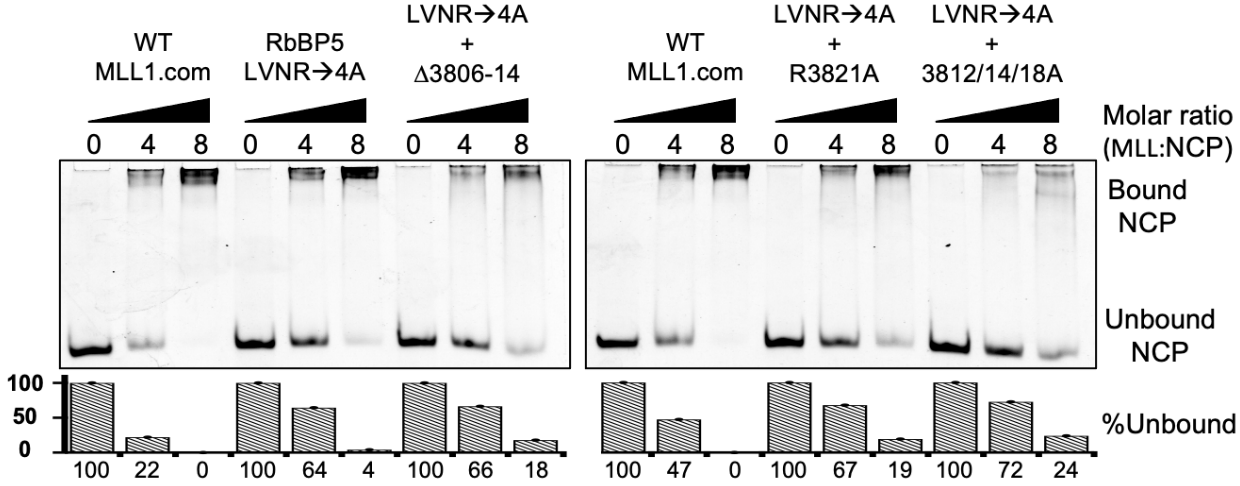


Figure 4.7. Combined RbBP5$^{LVNR \rightarrow 4A}$ and MLL1^{SET} mutants modestly affect NCP binding. **a**, For full MLL1 complexes containing wild-type, RbBP5$^{LVNR \rightarrow 4A}$, and MLL1^{SET} mutant combinations as indicated above. For **a**, **b**, and **c** values represent increasing molar ratio of proteins on top relative to NCP. Gels were quantified using ImageJ [1] and representative bar graphs were produced in GraphPad Prism.

4.4 Discussion

Here, we report two major binding modes of the MLL1-NCP^{K4M} core complex *in vitro*. Our study suggests that the MLL1 complex is highly dynamic on the NCP, adopting at least two major conformations with ASH2L as an anchor. We show that disruption of only one of the binding modes (i.e., MLL1-NCP$^{mode\ 2}$) reduces the MLL1 activity on the NCP. In this configuration, MLL1^{SET} resides above the nucleosome dyad, granting near symmetrical access to both H3 tails for maximal processivity [17]. Mutation of key residues at the MLL1-NCP interaction interface significantly reduces MLL1 activity, especially at higher methylation states. Thus, we believe this binding mode likely represents an active conformation for the MLL1 complex on the NCP. In contrast, despite a larger interaction interface, disruption of the unique interactions between the MLL1 and the NCP in MLL1-NCP$^{mode\ 1}$ does not reduce overall activity of the MLL1 complex. We propose that the MLL1-NCP$^{\text{mode 1}}$ may represent a ‘poised’ state, which enables the binding of MLL1 to the NCP in a less productive conformation. Indeed, mutation of amino acids unique to MLL1-NCP$^{\text{mode 1}}$ binding, including RbBP5$^{\text{LVNR}\rightarrow\text{4A, “Loop 1”}}$ and MLL$^{\text{SET},\Delta 3806\text{-}14}$ or MLL$^{\text{SET,3812/14/18A}}$, modestly affect the NCP binding via EMSA. The finding that MLL1-NCP$^{\text{mode 1}}$ may be in a ‘poised’ state is surprising since it resembles the active conformation of MLL1-NCP$^{\text{H2BK120ub}}$ complex [18]. H2BK120ub is known to stimulate the activity of the mammalian KMT2 family of enzymes [29, 30]. However, how H2BK120ub regulates MLL1 activity on the NCP remains unclear. The Ub moiety in the cryo-EM structure of the MLL1-NCP$^{\text{H2BK120ub}}$ complex is barely detectable [18] and its interaction with the NCP remains to be defined at the molecular level. We would like to point out that the *in vitro* HMT assay measures overall activity from ensemble of different MLL1-NCP binding modes. Thus, it is possible that the MLL1-NCP$^{\text{mode 1}}$ binding is active, but it has less contribution to overall activity as compared to the MLL1-NCP$^{\text{mode 2}}$.

The MLL1-NCP$^{\text{mode 1}}$ is similar to the configuration of the ySET1-NCP complex [19, 20], which is intrinsically inactive on an unmodified NCP [19]. Just like MLL1-NCP$^{\text{mode 1}}$, the ySET1 complex binds across the nucleosome disc with the highly conserved $_{271}$IINR$_{274}$ loop of Swd1, homologous to the $_{248}$LVNR$_{251}$ amphipathic loop (Loop 1) in RbBP5, anchoring on DNA near SHL2.5. There are also extensive contacts between a unique arginine rich motif (ARM) in ySET1 and the acidic patch on the NCP [19, 20]. Unlike RbBP5, mutation of both I271/I272 or $_{271}$IINR$_{274}$ loop in Swd1 (yeast RbBP5 homolog) leads to complete loss of higher methylation of H3K4 in yeast [19, 31] while ARM is essential for rendering ySET1 inactive on an unmodified NCP [19, 20]. In contrast to MLL family enzymes in higher eukaryotes, ySET1 complex does not display rotational dynamics *in vitro* [19, 20]. This could be due to additional NCP interactions provided by both extended SET domain acidic patch interactions and the yeast specific Spp1 protein in the ySET1 complex. Removal of Spp1 in the ySET1 complex derepresses SET1 on the NCP [19]. Whether removal of Spp1 increases rotational dynamics of the ySET1 complex remains an open question for future investigations.

Our study highlights significant rotational dynamics of the MLL1 complex on the NCP, which is unique among histone methyltransferase complexes [25, 27, 32, 33]. The MLL1 complex is able to rotate clockwise or counterclockwise on the NCP with either ASH2L (this study) and RbBP5 [10] as anchors, respectively. Rotation of the MLL1 complex on the NCP allows the MLL1^{SET} domain to move away from the nucleosome dyad and thereby reduces H3K4me3 activity without changing MLL1 binding on chromatin. The dynamics that allow the MLL1 complex to shift between the active (dyad) and poised (cross-surface) states likely enable loci-specific regulation of H3K4me3 in cells. Indeed, we recently showed that DPY30, a core component of the MLL1 complex, is able to regulate MLL1 tri-methylation activity by stabilizing the MLL1 complex in binding mode 2 via intrinsically disordered regions (IDRs) of ASH2L [10]. In the absence of DPY30, the MLL1^{SET} moves away from the nucleosome dyad and significantly reduces H3K4me3 activity [10]. It would be interesting to examine whether other proteins, especially those in the transcriptional machinery or in chromatin-remodeling complexes, are able to regulate MLL1 activity via modulating RbBP5 anchoring and shifting its binding equilibriums between active vs. poised states. Taken together, our studies reveal a new paradigm for regulation of MLL1-mediated H3K4 methylation through restricting or modulating the engagement of the MLL1 complex on the NCP. Since high rotational dynamics are unique to the MLL1 complex [25, 27, 32, 33], it may reflect a functional or regulatory necessity of loci-specific regulation of H3K4 methylation states and/or H3K4me heterogeneity in higher eukaryotes.

Finally, we envision that the rotational dynamics of the MLL1 core complex on the NCP can also be regulated by proteins that engage the linker DNA or adjacent nucleosomes. Given that both the PRC2 and Rpd3S complexes require linker DNA for optimal chromatin binding and catalysis [25-27], it is possible that the nucleosome template assembled with linker DNA or an oligo-nucleosome array may reduce rotational dynamics of the MLL1 complex *in vitro*. The winged-helix motif in ASH2L is able to interact with DNA in a sequence-independent manner [28, 34], which potentially allows for additional interactions with linker DNA. It would be interesting to investigate the structure of the MLL1 complex on di-nucleosomes or oligo-nucleosomes in the future.

4.5 Conclusion

Cryo-EM structures of the MLL1, MLL3 and ySET1 complexes with the NCP have shed light on the underlying mechanisms of these methyltransferase complexes on chromatin [17-20]. These studies, for the first time, highlight the divergent regulation of the MLL/SET1 family enzymes on chromatin. Two laboratories (including ours) reported the cryo-EM structures of the MLL1-NCP complex [17, 18]. Both structures show a dynamic interaction between the MLL1 core complex and the NCP. We show that the MLL1 core complex binds at the edge of the NCP (MLL1-NCP$^{\text{mode 2}}$) via RbBP5 and ASH2L at DNA superhelical location (SHL) 1.5 and 7, respectively [17]. This positions

$MLL1^{SET}$ at the nucleosome dyad, facilitating near symmetrical access of both H3 tails to the catalytic site. The RbBP5-NCP interface constitutes a conserved Quad-R motif in RbBP5 that interacts with DNA at SHL1.5 as well as an I-loop emanating from the RbBP5 WD40 repeats that interact with histone H4. By comparison, the ASH2L-NCP interface is more dynamic and involves a highly conserved ${}_{205}KRK_{207}$ motif at N-terminus of ASH2L [17]. Mutation of RbBP5 (e.g., I-loop or Quad-R) or ASH2L (e.g., ${}_{205}KRK_{207}$ or ${}_{419}KFK_{421}$) significantly affect MLL1 catalysis in an NCP-specific manner, suggesting an active orientation for optimized catalysis. This binding conformation is also observed by Xue and colleagues [18].

Interestingly, Xue and colleagues have reported a second conformation of the MLL1-NCP interaction in a majority of their cryo-EM particles ($MLL1\text{-}NCP^{mode\ 1}$) [18]. In this case, $MLL1^{SET}$ binds across the nucleosome disc, in close proximity to the C-terminal helical region of H2A while RbBP5 binds to DNA SHL2 on the NCP. In this conformation, an arginine anchor (R3821) in $MLL1^{SET}$ contacts asparagine 72 (D72) of H2A. Several hydrophobic residues in RbBP5 (Leu248 and Val249 of ${}_{248}LNVR_{251}$ of Loop 1) are close to αC of H2B. Electrostatic contacts between RbBP5 (E296 of ${}_{294}RGE_{296}$ of Loop 2) and K79 of H3 is also observed [18]. Importance of these residues in MLL1 activity and binding on the NCP was not previously reported [18].

Here, we show that mutating residues involved in the $MLL1\text{-}NCP^{mode\ 1}$ does not affect overall MLL1 activity on the NCP. Despite this, combinatorial mutations of RbBP5 (i.e., $RbBP5^{LVNR \rightarrow 4A}$) and $MLL1^{SET}$ (e.g., $MLL1^{SET,\ \Delta 3806\text{-}14}$ or $MLL1^{SET,\ 3812/14/18A}$) modestly affected both MLL1 catalysis and NCP binding competency, suggesting $MLL1\text{-}NCP^{mode\ 1}$ likely plays a role in NCP recognition. Though it is likely that the MLL1 complex binds to the NCP in multiple conformations that have distinct intrinsic activities and this distinct state, these orientations consist of the majority of particles in these studies. This suggests to us that, despite the dynamic Ub orientations of MLL1-NCP$^{\text{mode 1}}$, this has a critical role in NCP recognition and MLL1 positioning, allowing a “poised” catalytic state. This seems further clarified by MLL1 functioning as a non-processive enzyme [16], in which successive methylation events occur with intermediate steps of H3 acquisition and release. The highly dynamic interactions between the MLL1 complex and chromatin allows for loci-specific fine-tuning of H3K4me level and state specificity *in vivo*. It would be important to establish which conformation represents the functionally active conformation in cells and whether transition between different binding conformation is regulated.

4.6 Materials and Methods

Protein expression and purification

MLL1 complex subunits (MLL1$^{\text{SET, 3762-3969}}$, ASH2L$^{1\text{-}534}$, RbBP5$^{1\text{-}538}$, WDR5$^{23\text{-}334}$, and DPY30$^{1\text{-}99}$) and mutants were expressed using the pET-28a expression vector with N-terminal SUMO- and His$_6$-tags [15]. Deletion and point mutation plasmids were constructed using overlapping PCR and confirmed by sequencing. All proteins were expressed in BL21 (DE3) *E. coli* strain in LB media. Cells were grown to OD$_{600}$ at 37 °C until 0.6-0.8, and expression was induced by adding 0.4 mM IPTG. After 16-18 hours at 20 °C, cells were lysed by sonication and soluble lysate was collected by centrifugation at 32,000 *x* g for 30 minutes at 4 °C. After filtration through a 0.45 μm syringe, the soluble fraction was loaded and purified through a Ni-NTA metal-affinity column (Goldbio). After several column volumes of wash buffer (20 mM Tris, pH 8.0, 300-500 mM NaCl, 2 mM β-mercaptoethanol, 10% v/v glycerol, and 10 mM imidazole), proteins were eluted with a stepwise imidazole gradient of wash buffer with 30, 60, 90, 120, 150, 210, and 300 mM imidazole. Fractions containing protein of interest were pooled and dialyzed overnight at 4 ºC in presence of SUMO-tagged ULP1. Negative Ni-NTA purification was repeated to remove SUMO- and His_6-tag, ULP1 and other bacterial impurities by collecting protein in flowthrough. Proteins were further purified on a HiLoad 16/600 Superdex 75pg or 200pg gel-filtration columns (GE Healthcare). To obtain stoichiometric $\text{MLL1}^{\text{RWSAD}}$ complex, this complex was purified by combining equimolar amounts of MLL1^{SET}, ASH2L, RbBP5, WDR5, and excess DPY30 and purified by HiLoad 16/600 Superdex 200pg gel-filtration column.

Histone preparation and nucleosome reconstitution

For histone purifications, full-length *Xenopus laevis* histones H2A, H2B, H3, H3K4M, and H4 were expressed and purified according to one-pot protocol [35]. Briefly, histone constructs were transformed into BL21 (DE3), except for H4, which used C41 (DE3). After growing at 37 ºC to OD_{600} of 0.6-0.8, protein expression was induced with 0.4 mM IPTG for 3 h (H2A, H2B, H3, and H3K4M) or 2 h (H4). Equimolar amounts of histones were combined and isolated from inclusion bodies and subject to octamer refolding [35]. Octamer concentration was determined using UV at 280 nm. Reconstitution of nucleosome was conducted by combining 147 bp Widom 601 DNA and octamer in 1:1 molar ratio in high salt buffer using standard linear salt gradient method [36] overnight at 4 ºC. Low salt buffer (20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT) was added via a peristaltic pump at ~ 1 ml/min. Nucleosomes were then further dialyzed into long-term storage buffer (20 mM cacodylic acid, pH 6.0, 1 mM EDTA) overnight at 4 ºC.

In vitro *histone methyltransferase assay*

For *in vitro* HMT assay, 0.3 µM MLL1 complex was mixed with *S*-adenosyl-L-methionine and either NCP (1 µM) or recombinant H3 (0.1 µM) in 20 µL of HMT buffer (20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT and 10% v/v glycerol) as previously described [37]. The reactions were incubated at room temperature for 1hr and quenched by adding 20 µL 2x SDS-PAGE loading buffer.

Western blotting

The histones were separated on a 15% polyacrylamide gel and transferred onto polyvinylidene difluoride membrane (PVDF, Millipore). The membrane was blocked in blocking solution, consisting of 5% milk in TBS buffer with 0.1% Tween-20 (TBST) and then incubated for 2 hours at room temperature with primary antibody. After washing 3 times with TBST, the membrane was incubated with the HRP-conjugated anti-rabbit secondary antibodies at room temperature for 1 h and developed using Pierce^{TM} ECL Western Blotting Substrate (Thermo Fisher Scientific, #32106). The images were captured on ChemiDoc^{TM} Touch Imaging System (Bio-Rad). The primary and secondary antibodies used in this study include: rabbit anti-H3K4me1 (Abcam, cat ab8895, 1:20000), rabbit anti-H3K4me2 (Millipore, cat # 07-030, 1:20000), rabbit-anti H3K4me3 (Millipore, cat # 07-473, 1:10000), and anti-Rabbit IgG Horseradish Peroxidase (HRP)-linked whole antibody (GE Healthcare, #NA934, 1:10000).

Electrophoretic mobility shift assay

EMSA was conducted by incubating 0.1 μM NCP and with increasing concentration of the MLL1^{RWSAD} complex in 10 μL buffer. The protein/NCP mixture was then loaded onto a 6% 0.5X TBE gel and run for 1.5 hours at 150 V on ice. After the run, the gel was stained with 1:20,000 diluted ethidium bromide for 10 minutes, washed for 10 minutes in ddH2O and visualized by UV transillumination on Bio-Rad ChemiDoc Imaging System. Quantification was done using ImageJ software [1] and graphs generated using GraphPad Prism.

Cryo-EM sample preparation and data collection

The MLL1-NCP samples were prepared using the GraFix method [38] as previously described [17]. In brief, 30 μM of MLL1 complex was incubated with 10 μM NCP and 0.5 mM *S*-adenosyl-L-homocysteine for 30 min at 4 °C in the GraFix buffer (50 mM HEPES, pH 7.5, 50 mM NaCl, 1 mM $MgCl_2$, and 1 mM TCEP). The sample was centrifuged at 48,000 rpm at 4 °C for 3 h through a gradient 0-60% glycerol and 0-0.2% glutaraldehyde. After centrifugation, the crosslinked sample was quenched with 1 M Tris-HCl, pH 7.5 and glycerol was removed by buffer exchange. The sample at 1 mg/ml was applied onto Quantifoil R1.2/1.3 grids (Electron Microscopy Sciences) and the grid was plunge-frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific) at 4 °C, 100% humidity and 4 sec blotting time. The cryo-EM data was collected on the 300 keV FEI Titan Krios (Thermo Fisher Scientific) equipped with the Gatan K2 summit direct electron detector at a magnification of 130,000x in a counted mode. Each micrograph was imaged at the pixel size of 1.06 Å/pixel with the defocus range of -1.0 to -2.5 μm. A dose rate of 1.34 electrons/Å/frame with a total 40 frames per 8 sec was applied, resulting in an accumulated dose of 53.4 elections per Å^2. A total of 2,377 movies were collected.

Cryo-EM data processing and model refinement

Micrographic movie stacks were subject to MotionCorr2 [39] for whole-frame and local drift correction, and Contrast transfer function (CTF) was performed using CTFFIND4.1 [40]. Micrographs with lower than 4.5 Å of the estimated resolution were excluded, which resulted in 2,377 micrographs. Particle picking was performed using Warp [41], and Warp picked a total of 808,836 particles The particles were extracted in RELION [42] and imported into cryoSPARC [43] for 2D classification. A total of 768,708 particles, after excluding bad classes from the 2D classification, were subjected to the first round of *ab initio* 3D classification into three classes (Figure S4.2 in Appendix C). One of the three classes showed a clear density for the MLL complex and the NCP. This class of particles was then used for the second round of *ab initio* 3D classification into three subclasses. Two of the three classes seemed to maintain a well-defined cryo-EM map of the MLL1 and NCP complex, but the binding patterns of MLL1 toward the NCP were distinguishable. Therefore, two subclasses were subjected for subsequent heterogeneous refinement independently. Two particle sets were further exported to RELION for additional 3D classification. For each class, the focused 3D classification was performed at the MLL1^{RWSAD} region without alignment (35 cycles, T=4, binary mask: 10 pixels/soft mask: 10 pixels). The best behaving class selected from Class 01 (30,847 particles, Mode 2) and Class 03 (30,322 particles, Mode 1) were subjected to 3D auto refinement. After CTF-refinement and particle polishing, each class was further refined and post-processed to a resolution of 4.03 and 4.76 Å, respectively (Figure S4.3a and S4.3b in Appendix C). The resolution of all structures was estimated by RELION with Fourier shell correlation (FSC) at the criteria of 0.143. To build the atomic model of the MLL1-NCP$^{K4M, \text{mode } 1}$ and MLL1-NCP$^{K4M, \text{mode } 2}$, respectively, the structures of MLL1-NCP

(PDB ID: 6KIX and 6PWV) were used for rigid-body fitting. The real-space refinement using PHENIX [44] was performed after the rigid-body fitting. For the model validation, MolProbity [45] was used, and the map and model FSC curves were calculated using phenix.mtriage in the PHENIX program package (Figure S4.3c and S4.3d in Appendix C). Statistics of data collection, refinement, and validation are summarized in Table 4.1 in Appendix C.

4.7 References

1. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-5.

2. Rao, R.C. and Y. Dou, *Hijacked in cancer: the KMT2 (MLL) family of methyltransferases.* Nat Rev Cancer, 2015. **15**(6): p. 334-46.

3. Sha, L., et al., *Insights on the regulation of the MLL/SET1 family histone methyltransferases.* Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 2020. **1863**(7): p. 194561.

4. Lauberth, S.M., et al., *H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation.* Cell, 2013. **152**(5): p. 1021-36.

5. Tang, Z., et al., *SET1 and p300 act synergistically, through coupled histone modifications, in transcriptional activation by p53.* Cell, 2013. **154**(2): p. 297-310.

6. Vermeulen, M., et al., *Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4.* Cell, 2007. **131**(1): p. 58-69.

7. Wang, P., et al., *Global Analysis of H3K4 Methylation Defines MLL Family Member Targets and Points to a Role for MLL1-Mediated H3K4 Methylation in the Regulation of Transcriptional Initiation by RNA Polymerase II.* Molecular and Cellular Biology, 2009. **29**(22): p. 6074.

8. Haddad, J.F., et al., *Structural Analysis of the Ash2L/Dpy-30 Complex Reveals a Heterogeneity in H3K4 Methylation.* Structure, 2018.

9. Lee, J.-E., et al., *H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation.* eLife, 2013. **2**: p. e01503.

10. Lee, Y.T., et al., *Mechanism for DPY30 and ASH2L intrinsically disordered regions to modulate the MLL/SET1 activity on chromatin.* Nature Communications, 2021. **12**(1).

11. Zhang, H., et al., *MLL1 Inhibition Reprograms Epiblast Stem Cells to Naive Pluripotency.* Cell Stem Cell, 2016. **18**(4): p. 481-94.

12. Cho, Y.W., et al., *PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex.* J Biol Chem, 2007. **282**(28): p. 20395-406.

13. Cosgrove, M.S. and A. Patel, *Mixed lineage leukemia: a structure-function perspective of the MLL1 protein.* FEBS J, 2010. **277**(8): p. 1832-42.

14. Dou, Y., et al., *Regulation of MLL1 H3K4 methyltransferase activity by its core components.* Nat Struct Mol Biol, 2006. **13**(8): p. 713-9.

15. Cao, F., et al., *An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain.* PLoS One, 2010. **5**(11): p. e14102.

16. Patel, A., et al., *On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex.* J Biol Chem, 2009. **284**(36): p. 24242-56.

17. Park, S.H., et al., *Cryo-EM structure of the human MLL1 core complex bound to the nucleosome.* Nature Communications, 2019. **10**(1): p. 5540.

18. Xue, H., et al., *Structural basis of nucleosome recognition and modification by MLL methyltransferases.* Nature, 2019. **573**(7774): p. 445-449.

19. Hsu, P.L., et al., *Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS.* Molecular Cell, 2019. **76**(5): p. 712-723.e4.

20. Worden, E.J., X. Zhang, and C. Wolberger, *Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome.* eLife, 2020. **9**: p. e53199.

21. Dover, J., et al., *Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6.* J Biol Chem, 2002. **277**(32): p. 28368-71.

22. Sun, Z.W. and C.D. Allis, *Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast.* Nature, 2002. **418**(6893): p. 104-8.

23. Jiao, L. and X. Liu, *Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2.* Science, 2015. **350**(6258): p. aac4383.

24. Justin, N., et al., *Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2.* Nat Commun, 2016. **7**: p. 11316.

25. Finogenova, K., et al., *Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3.* eLife, 2020. **9**: p. e61964.

26. Lee, C.-H., J. Wu, and B. Li, *Chromatin Remodelers Fine-Tune H3K36me-Directed Deacetylation of Neighbor Nucleosomes by Rpd3S.* Molecular Cell, 2013. **52**(2): p. 255-263.

27. Poepsel, S., V. Kasinath, and E. Nogales, *Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes.* Nat Struct Mol Biol, 2018. **25**(2): p. 154-162.

28. Chen, Y., et al., *Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding.* EMBO Rep, 2011. **12**(8): p. 797-803.

29. Wu, L., et al., *ASH2L regulates ubiquitylation signaling to MLL: trans-regulation of H3 K4 methylation in higher eukaryotes.* Mol Cell, 2013. **49**(6): p. 1108-20.

30. Kwon, M., et al., *H2B ubiquitylation enhances H3K4 methylation activities of human KMT2 family complexes.* Nucleic Acids Research, 2020. **48**(10): p. 5442-5456.

31. Qu, Q., et al., *Structure and Conformational Dynamics of a COMPASS Histone H3K4 Methyltransferase Complex.* Cell, 2018. **174**(5): p. 1117-1126 e12.

32. McGinty, R.K., R.C. Henrici, and S. Tan, *Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome.* Nature, 2014. **514**(7524): p. 591-596.

33. Anderson, C.J., et al., *Structural Basis for Recognition of Ubiquitylated Nucleosome by Dot1L Methyltransferase.* Cell Rep, 2019. **26**(7): p. 1681-1690 e5.

34. Sarvan, S., et al., *Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain.* Nat Struct Mol Biol, 2011. **18**(7): p. 857-9.

35. Lee, Y.T., et al., *One-pot refolding of core histones from bacterial inclusion bodies allows rapid reconstitution of histone octamer.* Protein Expr Purif, 2015. **110**: p. 89-94.

36. Luger, K., T.J. Rechsteiner, and T.J. Richmond, *Preparation of nucleosome core particle from recombinant histones.* Methods Enzymol, 1999. **304**: p. 3-19.

37. Dou, Y., et al., *Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF.* Cell, 2005. **121**(6): p. 873-85.

38. Kastner, B., et al., *GraFix: sample preparation for single-particle electron cryomicroscopy.* Nat Methods, 2008. **5**(1): p. 53-5.

39. Zheng, S.Q., et al., *MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy.* Nat Methods, 2017. **14**(4): p. 331-332.

40. Rohou, A. and N. Grigorieff, *CTFFIND4: Fast and accurate defocus estimation from electron micrographs.* J Struct Biol, 2015. **192**(2): p. 216-21.

41. Tegunov, D. and P. Cramer, *Real-time cryo-electron microscopy data preprocessing with Warp.* Nature Methods, 2019. **16**(11): p. 1146-1152.

42. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3.* Elife, 2018. **7**.

43. Punjani, A., et al., *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination.* Nat Methods, 2017. **14**(3): p. 290-296.

44. Liebschner, D., et al., *Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix.* Acta Crystallogr D Struct Biol, 2019. **75**(Pt 10): p. 861-877.

45. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.

CHAPTER 5.
Conclusions and Future Directions

5.1 Conclusions

The major goal of this dissertation was to dissect the mechanisms involving DPY30-mediated activation of MLL1^{SET} on a nucleosome core particle (NCP) substrate. MLL1^{SET} functions through a conserved complex of protein partners to efficiently catalyze mono-, di- and trimethylation of histone H3 lysine 4. Despite DPY30 having negligible effect in the system, previous biochemical and structural studies primarily focused on histone H3 as a substrate. Yet, somewhat paradoxically, it has drastic effects on global H3K4me3 *in vivo* and in embryonic stem cells (ESCs) and hematopoietic stem cells [1-6]. Further, no previous structural work involving MLL1 with the nucleosome had been undertaken, suggesting a gap in knowledge to be filled. Therefore, this system affords a possible insight into both how MLL/SET family methyltransferases engage chromatin, but also reveal the mechanism behind the how DPY30 regulates H3K4me3 on the nucleosome.

Here, we used detailed and expansive structural and biochemical techniques to show how MLL1 engages the NCP. We later showed the role of DPY30 in stabilizing MLL1 on NCP and its mechanism for activating MLL family members on the NCP. We finally show MLL1 engages in dynamic states on the NCP that have functionally unique interaction motifs. We can use the accumulation of these studies to investigate divergent mechanisms of regulation in higher eukaryotic organisms.

MLL1 Engages the NCP through Essential RbBP5 and ASH2L Interactions

In chapter two, we use structural and biochemical techniques to show how MLL1 engages with an unmodified nucleosome. Specifically, we show how complete MLL1^{RWSAD} complex engages the nucleosome. We show that MLL1^{RWSAD} aligns at the nucleosome dyad allowing for near symmetrical access for each histone H3 tail to the MLL1^{SET} active site. MLL1^{RWSAD} uses several motifs in RbBP5 and ASH2L to engage the NCP. At SHL1.5, RbBP5 uses a Quad-R motif engaging the DNA backbone; an inserting loop (I-loop) that meshes with histones H3 and H4; and an anchoring loop (A-loop) that engages the histone H4 tail to anchor the complex. At SHL7, ASH2L Linker ($_{205}$KRK$_{207}$) and Loop ($_{419}$KFK$_{421}$) IDRs form the other anchor point on the NCP. Loss of these motifs universally attenuated MLL1 activity, particularly higher methylation states, in an NCP-specific manner to varying degrees.

Despite identifying a second conformation, which was published at the same time as ours [7], this orientation lacked a consistent DPY30 density. Further, this orientation differentiated strongly from the yeast homolog (ySET1) whose NCP structure showed consistent orientation across the nucleosome disc [8, 9]. The latter three structures were acquired on a H2BK120ub-modified NCP. Given H2BK120ub is not a catalytic prerequisite for MLL1 (unlike ySET1) and the molecular details for how MLL1 is activated by H2BK120ub remain unknown, the role of H2BK120ub in MLL family catalytic regulation is an open question to be dissected in future work.

DPY30 Functions through ASH2L IDRS and is Integral in MLL1 Stability on the NCP

In chapter three, we sought to reveal the paradoxical finding that although DPY30 does not affect catalysis on an H3 substrate [1-3], it is capable of regulating global H3K4me3 *in vivo* and in embryonic stem cells (ESCs) and hematopoietic stem cells [4-6]. Our studies of DPY30 to reveal the unique mechanism behind DPY30-induced activation amongst MLL/SET family showing that it globally enhances MLL/SET family catalysis on an NCP substrate. We explore how the specific binding of DPY30 to ASH2L Sdc1/DPY30 interaction (SDI) motif [10, 11] affects ASH2L intrinsically disordered regions (IDRs) using methyl-TROSY NMR studies. Specifically, we show that, upon DPY30 binding, widespread peak dispersion and stabilization occurs. Using mutagenesis, we show these peaks occur primarily in ASH2L Linker (202-286) and Loop (400-440) IDRs.

Through these IDRs, DPY30 stabilizes the ASH2L interaction at SHL7. We use cryo-EM to show that, without DPY30, MLL1 freely rotates with RbBP5 as an anchor. Further, in some of these densities, the ASH2L density is missing, suggesting that DPY30 maintains ASH2L stability within the MLL1 complex. Finally, we show that DPY30 is essential for establishing *de novo* H3K4me3 in cells. Taken together, these studies finally revealed the underlying mechanisms for 1) DPY30-mediated activation of MLL1 on the NCP *in vitro,* 2) a novel role for DPY30 in MLL1 complex stability on the NCP, and 3) that DPY30 in necessary to establish H3K4me3 in cells.

Distinct MLL1-NCP Interaction Motifs have Unique Roles in NCP Binding

In chapter four, we first use a catalytically inactive mutant containing H3 K-to-M at position four to capture discrete MLL1-NCP populations. We then fully investigate the alternative binding mode (mode 1) identified but not explored fully when originally published [7, 12]. We show that this alternate binding orientation (mode 1) rotates from the dyad orientation (mode 2) to allow for new interaction motifs. This orientation allows for new motifs to engage the NCP at the histone core and DNA specifically the MLL1^{SET} domain and a highly conserved amphipathic loop from RbBP5.

We show the orientation in mode 1 overlays well with prior structures of the ancient homolog ySET-NCP^{H2BK120ub} [8]. This occurs through a large rotation at the ASH2L/Bre2 anchor allowed RbBP5/Swd1 to rotate to SHL2.5. Despite this the densities of Ub orient very differently in each structure. Specifically, for the ySET1 structure, Ub orients very closely and makes direct contact at the SET1-Swd1 C-terminal interface [8, 9]. This allows direct relief of autoinhibition of the SET1-specific arginine rich motif (ARM) [8, 9]. This agrees with a previous biochemical study identifying ARM of ySET1 and Spp1 as essential factors for this regulation [13]. Similar to DOT1L [14], H2BK120ub-depdendent stimulation occurs intra-nucleosomally and its trans-tail regulation is modest in the absence of Spp1. Conversely, lacking an ARM motif, the Ub moiety in the MLL1/3-NCP^{H2BK120ub} structures makes highly dynamic interactions with RbBP5, freely rotating at a location distant from the MLL1/3^{SET} active site, suggesting it may not be an essential prerequisite as it is for ySET1 [15, 16].

We further show that, unlike in the motifs essential to mode 2, mode 1 motifs do not affect MLL1 catalysis in an NCP-specific manner, despite modestly affecting overall methylation on the NCP and recombinant H3. Specifically, we show, unlike the primordial homolog ySET1, several highly conserved motifs that strongly attenuate ySET1 activity, little effect is seen in MLL1. This includes an arginine anchor motif (R3821) engaging the acidic patch that, when mutated, has no effect despite the acidic patch being essential to ySET1 [17]. Additionally, mode 1 allows for a highly conserved amphipathic loop ($_{248}$LVNR$_{251}$ in RbBP5, $_{271}$IINR$_{274}$ in yeast Swd1) to contact core histones. These are essential contacts that, when mutated in yeast (i.e., II→AA or IINR→A4), completely attenuating all higher methylation. However, these mutations in RbBP5 (i.e., LV→AA or LNVR→4A) are completely benign in MLL1 catalysis on NCP and H3. Further, we show that combinatorial mutations (e.g., RbBP5$^{LVNR\rightarrow4A}$ + MLL1$^{SET,\ R3821A}$) do not show additive effects in repressing MLL1 methylation on NCP or H3. Concomitant with their modest reduction of MLL1 catalysis, these combined mutants similarly affected MLL1 binding by electrophoretic mobility shift assay (EMSA).

Taken together, we show here that MLL1-NCP$^{mode\ 1}$ and MLL1-NCP$^{mode\ 2}$ show divergent effects on MLL1 binding and catalysis on the NCP. We show that while MLL1-NCP$^{mode\ 1}$ overlays well with ySET-NCP^{H2BK120ub}, the motifs unique to this mode do not affect MLL1 catalysis in an NCP-specific manner. Moreover, highly conserved motifs between ySET1 and MLL1 complexes show ySET1-specific effects suggesting a foundation for unique eukaryotic regulation mechanisms.

5.2 Future Directions

The conclusions drawn from this dissertation reveal the manners in which MLL1 engages chromatin with dynamics distinct from its ancient homolog, ySET1. Further, it implicates DPY30 in several new roles including stabilizing ASH2L IDRs upon binding, thereby stabilizing ASH2L within the MLL1 complex, decreasing MLL1 rotation dynamics and promoting higher H3K4 methylation states. Finally, this distinct regulatory mechanism for MLL1 appears to allow distinct “poised” versus “active” states of NCP engagement unique to mammalian MLL. However, future efforts can clarify several points of contention to underscore novel mechanisms of eukaryotic transcriptional regulation.

First, regulation of H3K4me3 by H2B ubiquitylation (H2BK120ub) is one of the best described histone cross-talks *in vivo*. In yeast, H2BK123ub (H2BK120ub in humans), a highly prevalent mark [15, 16], is a prerequisite for global H3K4me3 [15, 16, 18]. The H2BK120ub mediated trans-tail regulation occurs intra-nucleosomally, similar to that of DOT1L [14]. Beyond this, however, Spp1 (CFP1 in humans) has been shown to directly regulate ySET1 activity [13]. Previous studies show that removal of Spp1 leads to decreased ySET1 activity on the NCP [13] and 40% reduction of global H3K4me *in vivo* [19]. The discrepancy could be due to other components (i.e., Swd2 and Shg1) that significantly repress ySET1 activity on the H2BK120ub-NCP but are not included in the structure studies. Decoupling this complexity in both yeast Spp1 and human CFP1 could go a long way in distinguishing their regulatory mechanisms.

Alternatively, the inclusion of the extended nSET region in the recent EM model (726-C) and previous reports, particularly within the nSET domain, may function redundantly with Spp1 [13], and may account for this. Indeed, Kim and colleagues found in a minimal SET and post-SET-containing complex ($SET1C^{938}$), the nSET would not bind with $SET1C^{938}$ until Spp1 was added [13]. This strongly suggests a functional prerequisite for Spp1, not only H2BK120ub, in the nSET-SET1C catalytic and post-SET domain interaction. In agreement with nSET-Spp1 redundancy, a SET1 complex containing an extended nSET domain ($SET1C^{762}$) now lacking Spp1 was completely incapable of methylating an H2BK120Ub NCP and only after the deletion was methylation restored. Indeed, this observation extended to nSET-containing human SET1A complex

(SET1AC^{1421}) wherein only upon deletion of CFP1 (mammalian Spp1 homolog) was methylation on an H2BK120Ub NCP restored. Given that in fission yeast, Spp1 is a requirement for robust H3K4 methylation, it suggests the role of Spp1, along with Swd2 and Shg1, warrant future studies [20]. In humans, CFP1 seems to be important for activation of the hSET1 complex on NCP^{H2BK120ub} [13], contrary to the role of Spp1 in yeast [8]. We also showed that SET1A can be stimulated independently and cooperatively by DPY30 and H2BK120ub. It would be important to examine the function of CFP1 on both modified and unmodified NCP and its mechanism of action. Structural studies similar to what has been done [7, 12] with SET1A or SET1B with modified and unmodified NCPs and with or without CFP1 may be the first step in decoupling these divergent effects.

Given the unique autoinhibitory role of ARM in ySET1 and lack of conservation of ARM in MLL homologs, it raises the question of whether H2BK120ub regulates other MLL family enzymes? Unlike yeast, decoupling of H2BK120ub and H3K4me3 has been widely described in mammals and *Tetrahymena thermophila* [21, 22]. Furthermore, it has been shown that the MLL1 complex has very high methylation activity on the unmodified NCP [2, 12, 23]. H2BK120ub does not regulate activity of the MLL3 complex and has only a modest effect on the MLL1 activity *in vitro* [23]. The cryo-EM structure of the MLL1/3 core complexes bound to NCP^{H2BK120ub} (MLL1 EMDB:9999, PDB: 6KIV; MLL3 EMDB: 0693, PDB: 6KIW) seems to support a different role of H2BK120ub in regulation of the MLL1/3 activity. Although the MLL1/3 complexes overlay well with that of ySET1- NCP^{H2BK120ub}, key interactions between ySET1-NCP^{H2BK120ub} are not conserved in the MLL1/3-NCP^{H2BK120ub} complexes [7, 8]. The N-terminal region of the MLL1^{SET} domain interact with neither ubiquitin nor the "acidic patch" on the NCP. Instead, ubiquitin module exhibits dynamic binding to multiple different surfaces near RbBP5 and its density is not readily visible in the coulomb potential map under normal contour levels [7]. These results suggest that H2BK120ub is probably less important for regulating the MLL complexes in higher eukaryotes. Alternatively, it may regulate MLL activities through proteins that are not fully characterized in the structure. Given the conflicting biochemical and structural results, it would be an important study to define the precise molecular mechanisms by which H2BK120ub activates MLL family members. In particular, are there essential motifs in MLL complexes required? Are they unique from those required for DPY30-dependent activation? Answering these questions would allow for major strides in understanding how MLL1 responds to contextual requirements for H3K4me3 in cells.

Finally, despite the critical findings in these recent structural papers [7, 12], the N-terminal PHD-WH density of ASH2L is completely absent from these structures. Previous work showed that this domain is capable of crystallization and freely binds non-specifically [24], but we do not see this density binding intra-nucleosomally. In fact, given our recent finding that the PHD-WH does not contribute to DPY30-dependent activation of MLL1, this might suggest the substrate used was insufficient [25]. Particularly with the structure of PRC2 bound to a hetero-dinucleosome [26], it is possible that an optimized substrate may resolve this. It is possible that by using a dinucleosome construct, it may reveal new mechanisms for an otherwise dispensable domain of ASH2L based on currently used biochemical substrates.

5.3 References

1. Haddad, J.F., et al., *Structural Analysis of the Ash2L/Dpy-30 Complex Reveals a Heterogeneity in H3K4 Methylation.* Structure, 2018.

2. Patel, A., et al., *On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex.* J Biol Chem, 2009. **284**(36): p. 24242-56.

3. Shinsky, S.A. and M.S. Cosgrove, *Unique Role of the WD-40 Repeat Protein 5 (WDR5) Subunit within the Mixed Lineage Leukemia 3 (MLL3) Histone Methyltransferase Complex.* J Biol Chem, 2015. **290**(43): p. 25819-33.

4. Jiang, H., et al., *Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains.* Cell, 2011. **144**(4): p. 513-25.

5. Yang, Z., et al., *The DPY30 subunit in SET1/MLL complexes regulates the proliferation and differentiation of hematopoietic progenitor cells.* Blood, 2014. **124**(13): p. 2025-33.

6. Yang, Z., et al., *Dpy30 is critical for maintaining the identity and function of adult hematopoietic stem cells.* The Journal of experimental medicine, 2016. **213**(11): p. 2349-2364.

7. Xue, H., et al., *Structural basis of nucleosome recognition and modification by MLL methyltransferases.* Nature, 2019. **573**(7774): p. 445-449.

8. Hsu, P.L., et al., *Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS.* Molecular Cell, 2019. **76**(5): p. 712-723.e4.

9. Worden, E.J., X. Zhang, and C. Wolberger, *Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome.* eLife, 2020. **9**: p. e53199.

10. Tremblay, V., et al., *Molecular basis for DPY-30 association to COMPASS-like and NURF complexes.* Structure, 2014. **22**(12): p. 1821-1830.

11. South, P.F., et al., *A Conserved Interaction between the SDI Domain of Bre2 and the Dpy-30 Domain of Sdc1 Is Required for Histone Methylation and Gene Expression.* Journal of Biological Chemistry, 2010. **285**(1): p. 595-607.

12. Park, S.H., et al., *Cryo-EM structure of the human MLL1 core complex bound to the nucleosome.* Nature Communications, 2019. **10**(1): p. 5540.

13. Kim, J., et al., *The n-SET domain of Set1 regulates H2B ubiquitylation-dependent H3K4 methylation.* Mol Cell, 2013. **49**(6): p. 1121-33.

14. McGinty, R.K., et al., *Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation.* Nature, 2008. **453**(7196): p. 812-6.

15. Dover, J., et al., *Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6.* J Biol Chem, 2002. **277**(32): p. 28368-71.

16. Sun, Z.W. and C.D. Allis, *Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast.* Nature, 2002. **418**(6893): p. 104-8.

17. Nakanishi, S., et al., *A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation.* Nat Struct Mol Biol, 2008. **15**(8): p. 881-8.

18. Zheng, S., J.J. Wyrick, and J.C. Reese, *Novel trans-tail regulation of H2B ubiquitylation and H3K4 methylation by the N terminus of histone H2A.* Mol Cell Biol, 2010. **30**(14): p. 3635-45.

19. Dehé, P.-M., et al., *Protein Interactions within the Set1 Complex and Their Roles in the Regulation of Histone 3 Lysine 4 Methylation.* Journal of Biological Chemistry, 2006. **281**: p. 35404-35412.

20. Roguev, A., et al., *High conservation of the Set1/Rad6 axis of histone 3 lysine 4 methylation in budding and fission yeasts.* J Biol Chem, 2003. **278**(10): p. 8487-93.

21. Vethantham, V., et al., *Dynamic Loss of H2B Ubiquitylation without Corresponding Changes in H3K4 Trimethylation during Myogenic Differentiation.* Molecular and Cellular Biology, 2012. **32**(6): p. 1044-1055.

22. Wang, Z., B. Cui, and M.A. Gorovsky, *Histone H2B Ubiquitylation Is Not Required for Histone H3 Methylation at Lysine 4 in Tetrahymena.* Journal of Biological Chemistry, 2009. **284**(50): p. 34870-34879.

23. Wu, L., et al., *ASH2L regulates ubiquitylation signaling to MLL: trans-regulation of H3 K4 methylation in higher eukaryotes.* Mol Cell, 2013. **49**(6): p. 1108-20.

24. Chen, Y., et al., *Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding.* EMBO Rep, 2011. **12**(8): p. 797-803.

25. Lee, Y.T., et al., *Mechanism for DPY30 and ASH2L intrinsically disordered regions to modulate the MLL/SET1 activity on chromatin.* Nature Communications, 2021. **12**(1).

26. Poepsel, S., V. Kasinath, and E. Nogales, *Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes.* Nat Struct Mol Biol, 2018. **25**(2): p. 154-162.

APPENDIX A.
Supplementary Information for Chapter 2

This appendix includes supplementary information and figures for Chapter 2 that were used as quality control or clarity in this study.

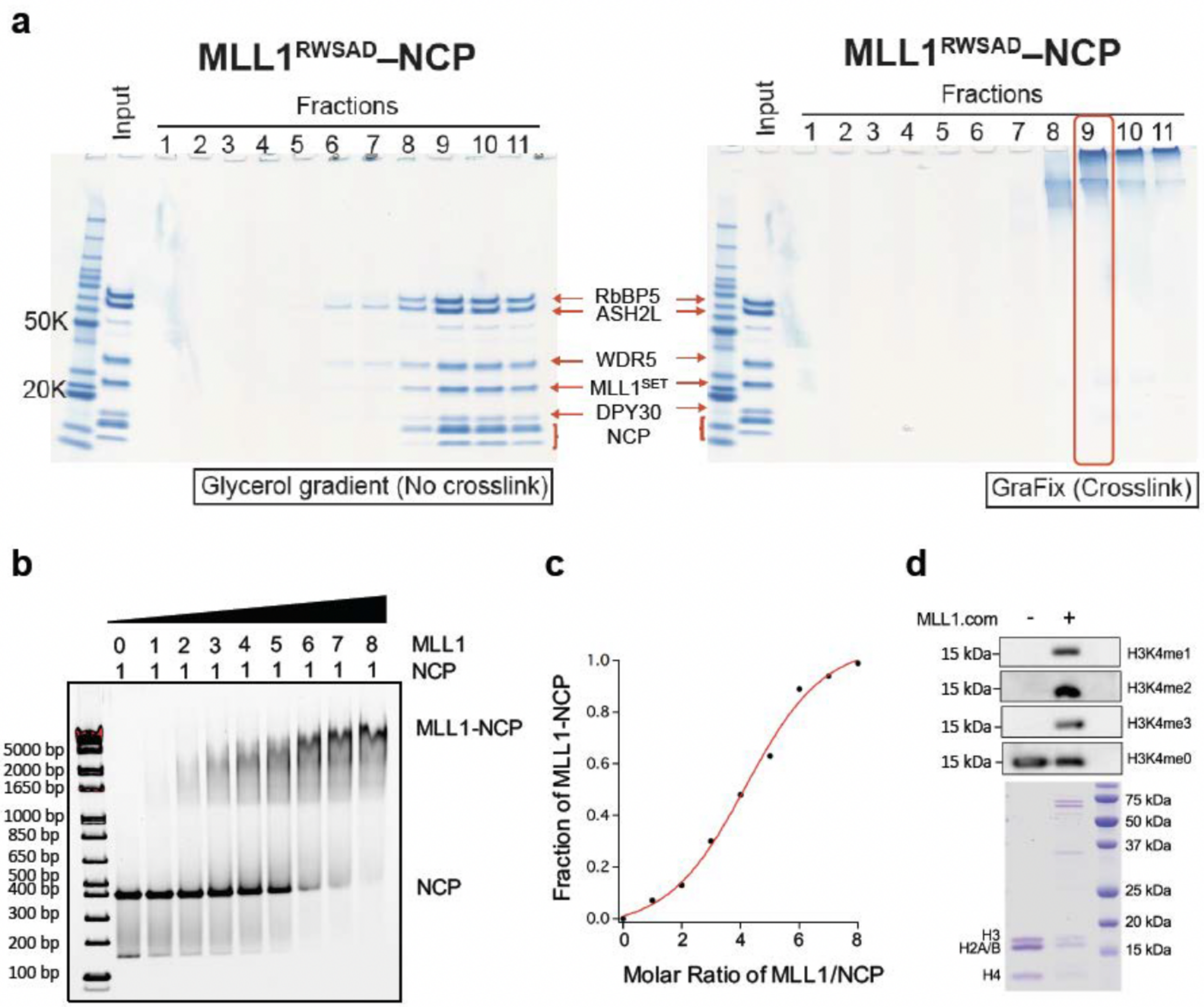


Figure S2.1. Preparation of the MLL1^{RWSAD}-NCP complex. **a**, MLL1^{RWSAD} complex and the NCP were incubated and isolated by glycerol gradient (0-60%) with (right) or without (left) the addition of glutaraldehyde crosslinker. The fractions (1-11) were analyzed by SDS-PAGE. Individual components of the MLL1 core complex and the NCP were indicated. Fraction #9 from the GraFix sample (red box) was used for structural analysis. **b**, Electrophoretic mobility shift assay (EMSA) for the MLL1 core complex and the NCP. The molar ratio of MLL1 vs. NCP was indicated on top. **c**, Relative quantification of bound NCP (in **b**) by ImageJ [4] was presented. The experiment was repeated twice (not shown) to conform concentration dependent binding of MLL1 to the NCP. **d**, *In vitro* histone methyltransferase assay (HMT) for the NCP incubated with or without the MLL1 complex. No signal was detected in the immunoblot for the H3K4me1, H3K4me2 or H3K4me3 antibodies, which confirmed their respective specificity for the modified H3 in the NCP. The antibody against unmodified H3K4 was used as a control. Coomassie gel on the bottom was provided to show components used in each reaction. GraFix and crosslinking by SHP; EMSA by YTL

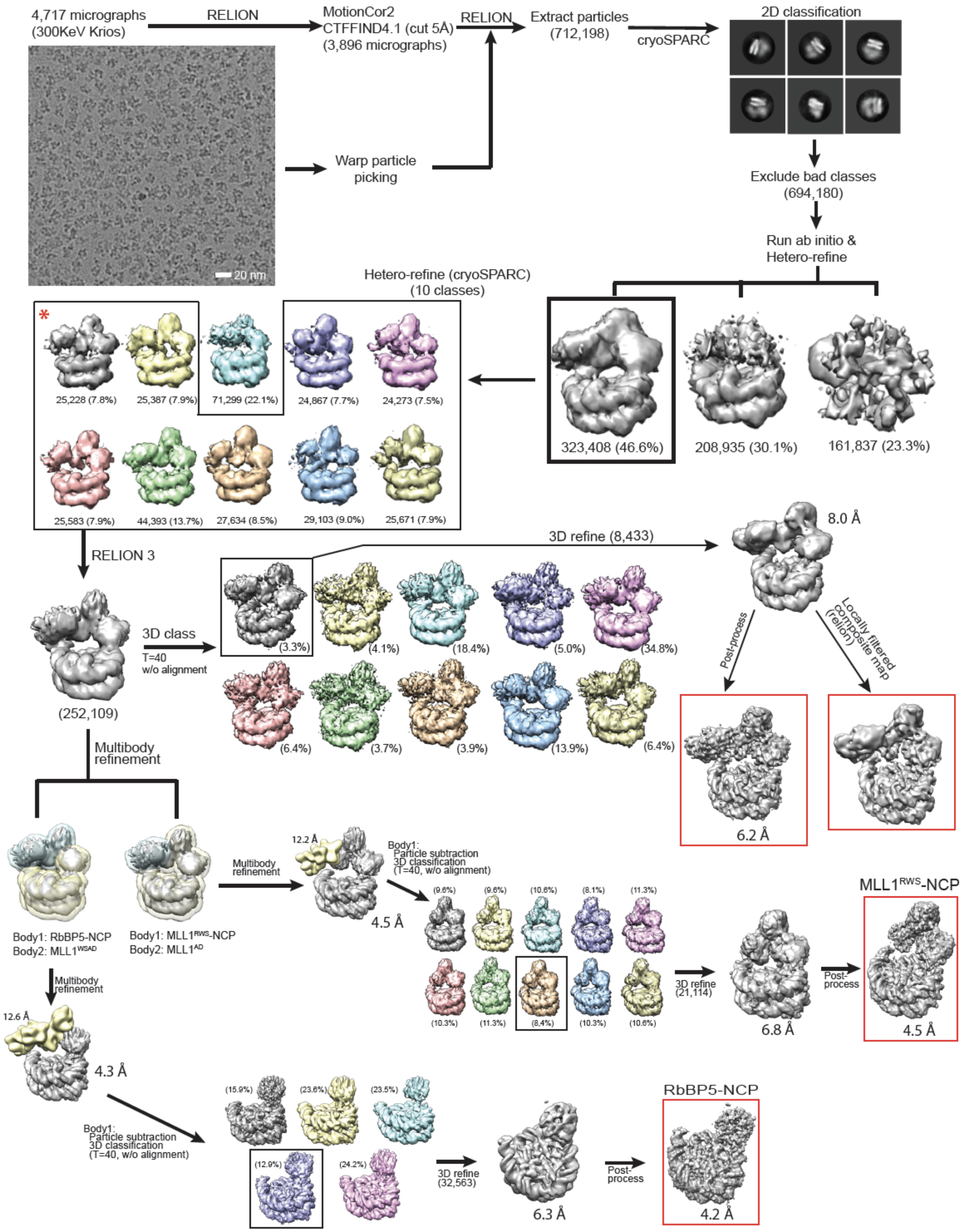

4,717 micrographs
(300KeV Krios)
RELION
MotionCor2
CTFFIND4.1 (cut 5Å)
(3,896 micrographs)
RELION
Extract particles
(712,198)
cryoSPARC
2D classification
20 nm
Warp particle
picking
Exclude bad classes
(694,180)
Run ab initio &
Hetero-refine
Hetero-refine (cryoSPARC)
(10 classes)
323,408 (46.6%)
208,935 (30.1%)
161,837 (23.3%)
*
25,228 (7.8%)
25,387 (7.9%)
71,299 (22.1%)
24,867 (7.7%)
24,273 (7.5%)
25,583 (7.9%)
44,393 (13.7%)
27,634 (8.5%)
29,103 (9.0%)
25,671 (7.9%)
RELION 3
3D refine (8,433)
8.0 Å
3D class
T=40
w/o alignment
(252,109)
(3.3%)
(4.1%)
(18.4%)
(5.0%)
(34.8%)
(6.4%)
(3.7%)
(3.9%)
(13.9%)
(6.4%)
Post-process
Locally filtered
composite map
(relion)
6.2 Å
Multibody
refinement
Body1: RbBP5-NCP
Body2: MLL1WSAD
Body1: MLL1RWS-NCP
Body2: MLL1AD
Multibody
refinement
12.2 Å
4.5 Å
Body1:
Particle subtraction
3D classification
(T=40, w/o alignment)
(9.6%)
(9.6%)
(10.6%)
(8.1%)
(11.3%)
(10.3%)
(11.3%)
(8.4%)
(10.3%)
(10.6%)
3D refine
(21,114)
6.8 Å
Post-
process
MLL1RWS-NCP
4.5 Å
Multibody
refinement
12.6 Å
4.3 Å
Body1:
Particle subtraction
3D classification
(T=40, w/o alignment)
(15.9%)
(23.6%)
(23.5%)
(12.9%)
(24.2%)
3D refine
(32,563)
6.3 Å
Post-
process
RbBP5-NCP
4.2 Å

Figure S2.2. Cryo-EM data processing for the MLL1^{RWSAD}-NCP. Representative micrograph image (Titan Krios 300 keV) and 2D classifications of the MLL1^{RWSAD}-NCP complex were shown. The particle numbers for each classification step as well as the estimated resolution of overall and selected subcomplexes (red box) were shown at the bottom. The detailed data processing of MLL1^{RWSAD}-NCP, MLL1^{RWS}-NCP and RbBP5-NCP were described in the Materials and Methods section of Chapter 2. Cryo-EM data collection and processing done by USC and SHP

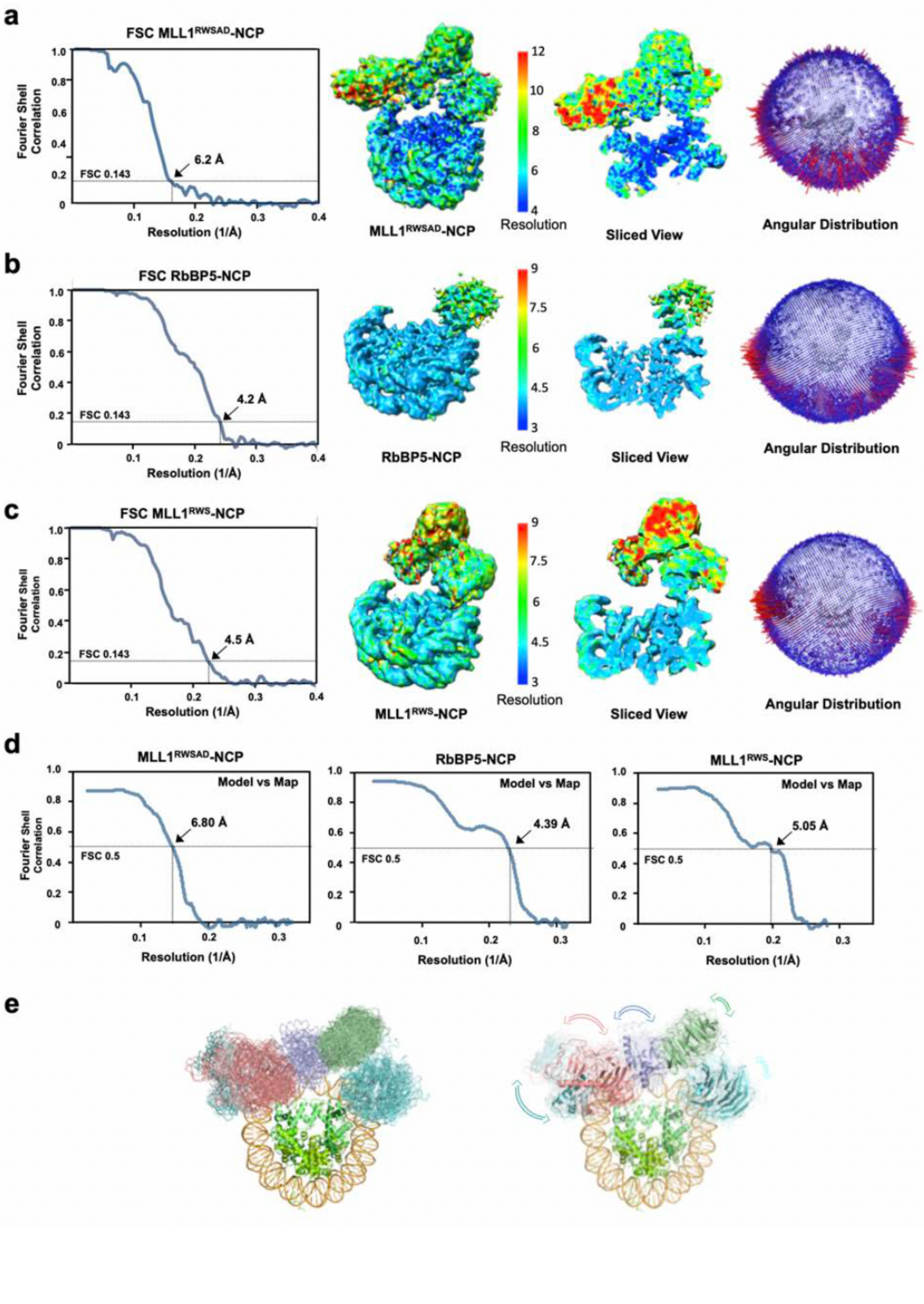

a
FSC MLL1^{RWSAD}-NCP
Fourier Shell Correlation
1.0
0.8
0.6
0.4
0.2
0
FSC 0.143
6.2 Å
0.1
0.2
0.3
0.4
Resolution (1/Å)
MLL1^{RWSAD}-NCP
12
10
8
6
4
Resolution
Sliced View
Angular Distribution
b
FSC RbBP5-NCP
Fourier Shell Correlation
1.0
0.8
0.6
0.4
0.2
0
FSC 0.143
4.2 Å
0.1
0.2
0.3
0.4
Resolution (1/Å)
RbBP5-NCP
9
7.5
6
4.5
3
Resolution
Sliced View
Angular Distribution
c
FSC MLL1^{RWS}-NCP
Fourier Shell Correlation
1.0
0.8
0.6
0.4
0.2
0
FSC 0.143
4.5 Å
0.1
0.2
0.3
0.4
Resolution (1/Å)
MLL1^{RWS}-NCP
9
7.5
6
4.5
3
Resolution
Sliced View
Angular Distribution
d
MLL1^{RWSAD}-NCP
Model vs Map
Fourier Shell Correlation
1.0
0.8
0.6
0.4
0.2
0
FSC 0.5
6.80 Å
0.1
0.2
0.3
Resolution (1/Å)
RbBP5-NCP
Model vs Map
1.0
0.8
0.6
0.4
0.2
0
FSC 0.5
4.39 Å
0.1
0.2
0.3
Resolution (1/Å)
MLL1^{RWS}-NCP
Model vs Map
1.0
0.8
0.6
0.4
0.2
0
FSC 0.5
5.05 Å
0.1
0.2
0.3
Resolution (1/Å)
e

Figure S2.3. Cryo-EM map validation of the MLL1^{RWSAD}-, RbBP5-, and MLL1^{RWS}-NCP complexes. **a**-**c**, Fourier Shell Correlation (FSC) curves for **a**, MLL1^{RWSAD}-NCP, **b**, RbBP5-NCP, and **c**, MLL1^{RWS}-NCP were shown on the left and the corresponding local resolution assessments by RESMAP [1] were shown in the middle. The final resolution was determined using FSC = 0.143 criterion, which was shown by the arrowhead on the FSC curve. Angular distribution plots were shown on the right. **d**, Model-map FSC curves for MLL1^{RWSAD}-NCP, RbBP5-NCP, and MLL1^{RWS}-NCP calculated by phenix-mtriage [2]. The resolution was indicated using FSC = 0.5 criterion, which was shown by the arrowhead on the FSC curve. **e**, Rigid-body fitting of the MLL1^{RWSAD} domains into 9 cryo-EM maps of MLL1^{RWSAD}-NCP, which were shown as hetero-refine subclasses in Figure S2.2 (highlighted by a red asterisk [*]). Left, nine coordinates of the MLL1^{RWSAD} core complex were overlaid and displayed. Right, the degree of relative movement of each MLL1^{RWSAD} domain within nine coordinates was indicated by arrows. The length and orientation of each arrow indicated the degree of dynamics and moving direction of each domain. Cryo-EM data collection and processing done by USC and SHP

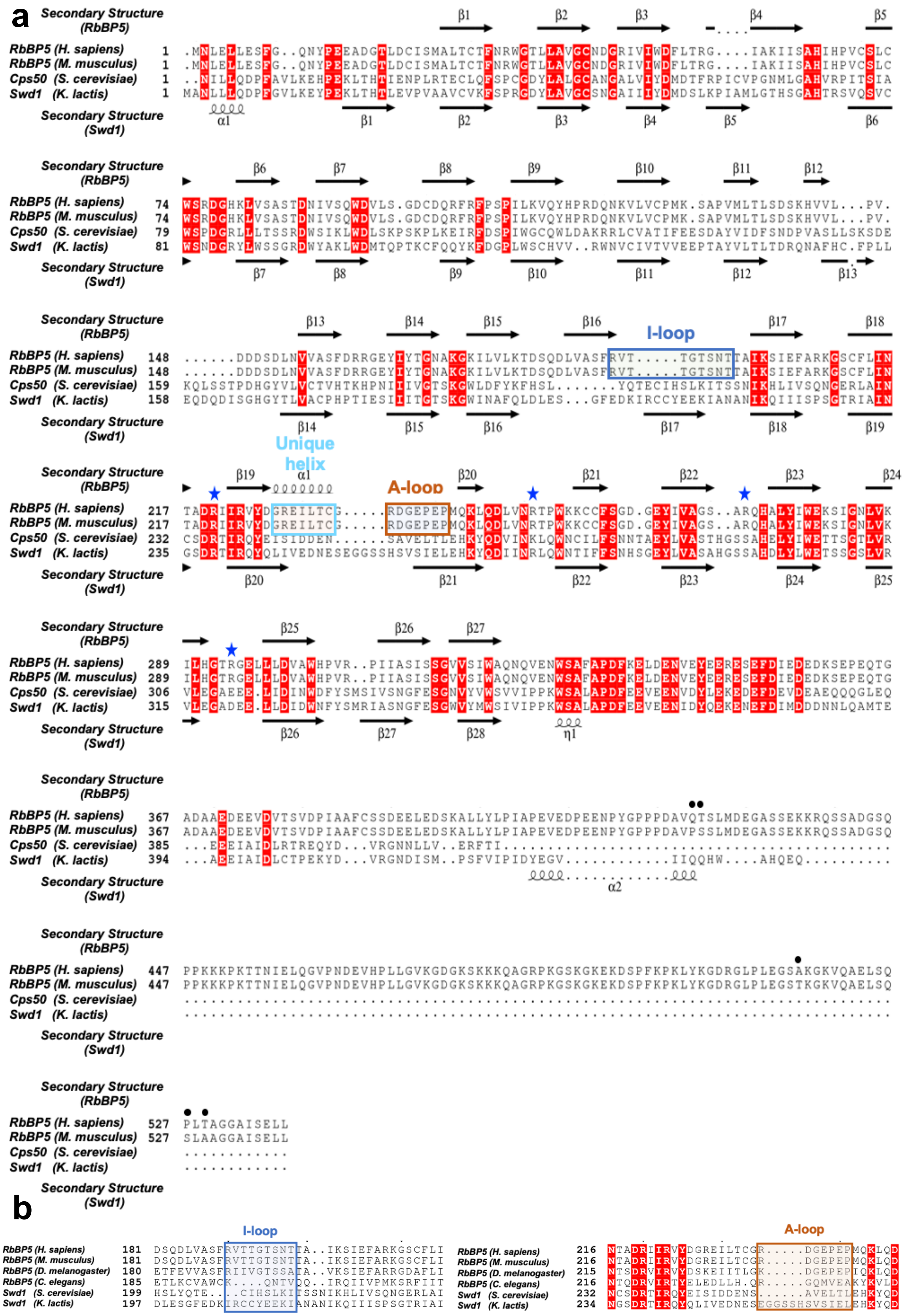

a
Secondary Structure (RbBP5)
β1 β2 β3 β4 β5
RbBP5 (H. sapiens) 1 .MNLELLESFG..QNYPEEADGTLDCISMALTCTFNRWGTLLAVGCNDGRIVIWDFLTRG....IAKIISAHIHPVCSLC
RbBP5 (M. musculus) 1 .MNLELLESFG..QNYPEEADGTLDCISMALTCTFNRWGTLLAVGCNDGRIVIWDFLTRG....IAKIISAHIHPVCSLC
Cps50 (S. cerevisiae) 1 ..NILLQDPFAVLKEHPEKLTHTIENPLRTECLQFSPCGDYLALGCANGALVIYDMDTFRPICVPGNMLGAHVRPITSIA
Swd1 (K. lactis) 1 MANLLLQDPFGVLKEYPEKLTHTLEVPVAAVCVKFSPRGDYLAVGCSNGAIIIYDMDSLKPIAMLGTHSGAHTRSVQSVC
Secondary Structure (Swd1)
α1 β1 β2 β3 β4 β5 β6
Secondary Structure (RbBP5)
β6 β7 β8 β9 β10 β11 β12
RbBP5 (H. sapiens) 74 WSRDGHKLVSASTDNIVSQWDVLS.GDCDQRFRFPSPILKVQYHPRDQNKVLVCPMK.SAPVMLTLSDSKHVVL...PV.
RbBP5 (M. musculus) 74 WSRDGHKLVSASTDNIVSQWDVLS.GDCDQRFRFPSPILKVQYHPRDQNKVLVCPMK.SAPVMLTLSDSKHVVL...PV.
Cps50 (S. cerevisiae) 79 WSPDGRLLLTSSRDWSIKLWDLSKPSKPLKEIRFDSPIWGCQWLDAKRRLCVATIFEESDAYVIDFSNDPVASLLSKSDE
Swd1 (K. lactis) 81 WSNDGRYLWSSGRDWYAKLWDMTQPTKCFQQYKFDGPLWSCHVV..RWNVCIVTVVEEPTAYVLTLTDRQNAFHC.FPLL
Secondary Structure (Swd1)
β7 β8 β9 β10 β11 β12 β13
Secondary Structure (RbBP5)
β13 β14 β15 β16 I-loop β17 β18
RbBP5 (H. sapiens) 148 ......DDDSDLNVVASFDRRGEYIYTGNAKGKILVLKTDSQDLVASFRVT.....TGTSNTTAIKSIEFARKGSCFLIN
RbBP5 (M. musculus) 148 ......DDDSDLNVVASFDRRGEYIYTGNAKGKILVLKTDSQDLVASFRVT.....TGTSNTTAIKSIEFARKGSCFLIN
Cps50 (S. cerevisiae) 159 KQLSSTPDHGYVLVCTVHTKHPNIIIVGTSKGWLDFYKFHSL.......YQTECIHSLKITSSNIKHLIVSQNGERLAIN
Swd1 (K. lactis) 158 EQDQDISGHGYTLVACPHPTIESIIITGTSKGWINAFQLDLES...GFEDKIRCCYEEKIANANIKQIIISPSGTRIAIN
Secondary Structure (Swd1)
β14 β15 β16 β17 β18 β19
Unique helix
Secondary Structure (RbBP5)
β19 α1 A-loop β20 β21 β22 β23 β24
RbBP5 (H. sapiens) 217 TADRIIRVYDGREILTCG.....RDGEPEPMQKLQDLVNRTPWKKCCFSGD.GEYIVAGS..ARQHALYIWEKSIGNLVK
RbBP5 (M. musculus) 217 TADRIIRVYDGREILTCG.....RDGEPEPMQKLQDLVNRTPWKKCCFSGD.GEYIVAGS..ARQHALYIWEKSIGNLVK
Cps50 (S. cerevisiae) 232 CSDRTIRQYEISIDDEN......SAVELTLEHKYQDVINKLQWNCILFSNNTAEYLVASTHGSSAHELYIWETTSGTLVR
Swd1 (K. lactis) 235 GSDRTIRQYQLIVEDNESEGGSSHSVSIELEHKYQDIINRLQWNTIFFSNHSGEYLVASAHGSSAHDLYLWETSSGSLVR
Secondary Structure (Swd1)
β20 β21 β22 β23 β24 β25
Secondary Structure (RbBP5)
β25 β26 β27
RbBP5 (H. sapiens) 289 ILHGTRGELLLDVAWHPVR..PIIASISSGVVSIWAQNQVENWSAFAPDFKELDENVEYEERESEFDIEDEDKSEPEQTG
RbBP5 (M. musculus) 289 ILHGTRGELLLDVAWHPVR..PIIASISSGVVSIWAQNQVENWSAFAPDFKELDENVEYEERESEFDIEDEDKSEPEQTG
Cps50 (S. cerevisiae) 306 VLEGAEEE.LIDINWDFYSMSIVSNGFESGNVYVWSVVIPPKWSALAPDFEEVEENVDYLEKEDEFDEVDEAEQQQGLEQ
Swd1 (K. lactis) 315 VLEGADEE.LLDIDWNFYSMRIASNGFESGWVYMWSIVIPPKWSALAPDFEEVEENIDYQEKENEFDIMDDDNNLQAMTE
Secondary Structure (Swd1)
β26 β27 β28 η1
Secondary Structure (RbBP5)
●●
RbBP5 (H. sapiens) 367 ADAAEDEEVDVTSVDPIAAFCSSDEELEDSKALLYLPIAPEVEDPEENPYGPPPDAVQTSLMDEGASSEKKRQSSADGSQ
RbBP5 (M. musculus) 367 ADAAEDEEVDVTSVDPIAAFCSSDEELEDSKALLYLPIAPEVEDPEENPYGPPPDAVPSSLMDEGASSEKKRQSSADGSQ
Cps50 (S. cerevisiae) 385 ...EEEIAIDLRTREQYD...VRGNNLLV..ERFTI...
Swd1 (K. lactis) 394 ...AEEIAIDLCTPEKYD...VRGNDISM..PSFVIPIDYEGV............IIQQHW....AHQEQ..........
Secondary Structure (Swd1)
α2
Secondary Structure (RbBP5)
●
RbBP5 (H. sapiens) 447 PPKKKPKTTNIELQGVPNDEVHPLLGVKGDGKSKKKQAGRPKGSKGKEKDSPFKPKLYKGDRGLPLEGSAKGKVQAELSQ
RbBP5 (M. musculus) 447 PPKKKPKTTNIELQGVPNDEVHPLLGVKGDGKSKKKQAGRPKGSKGKEKDSPFKPKLYKGDRGLPLEGSTKGKVQAELSQ
Cps50 (S. cerevisiae) ....
Swd1 (K. lactis) ....
Secondary Structure (Swd1)
Secondary Structure (RbBP5)
● ●
RbBP5 (H. sapiens) 527 PLTAGGAISELL
RbBP5 (M. musculus) 527 SLAAGGAISELL
Cps50 (S. cerevisiae) ....
Swd1 (K. lactis) ....
Secondary Structure (Swd1)
b
I-loop
RbBP5 (H. sapiens) 181 DSQDLVASFRVTTGTSNTTA..IKSIEFARKGSCFLI
RbBP5 (M. musculus) 181 DSQDLVASFRVTTGTSNTTA..IKSIEFARKGSCFLI
RbBP5 (D. melanogaster) 180 ETFEVVASFRIIVGTSSATA..VKSIEFARRGDAFLI
RbBP5 (C. elegans) 185 ETLKCVAWCK....QNTVQQ..IRQIIVPMKSRFIIT
Swd1 (S. cerevisiae) 199 HSLYQTE....CIHSLKITSSNIKHLIVSQNGERLAI
Swd1 (K. lactis) 197 DLESGFEDKIRCCYEEKIANANIKQIIISPSGTRIAI
A-loop
RbBP5 (H. sapiens) 216 NTADRIIRVYDGREILTCGR.....DGEPEPMQKLQD
RbBP5 (M. musculus) 216 NTADRIIRVYDGREILTCGR.....DGEPEPMQKLQD
RbBP5 (D. melanogaster) 215 NTSDRVIRVYDSKEIITLGK.....DGEPEPIQKLQD
RbBP5 (C. elegans) 216 NTQDRVIRTYELEDLLH.QR.....GQMVEAKYKVLD
Swd1 (S. cerevisiae) 232 NCSDRTIRQYEISIDDENS......AVELTLEHKYQD
Swd1 (K. lactis) 234 NGSDRTIRQYQLIVEDNESEGGSSHSVSIELEHKYQD

Figure S2.4. Sequence homology of RbBP5 and motifs among higher eukaryotes. **a**, The primary sequences of RbBP5 in *H. sapiens* and *M. musculus* (mammalian cells) as well as yeast homologues Swd1/Cps50 in *K. lactis* and *S. cerevisiae*, respectively, were used for multiple sequence alignment. The secondary structures of mouse RbBP5 and yeast Swd1 based on determined crystal structures were indicated on the top and bottom of the alignment, respectively. The **I**- and A-loops as well as the unique helix in mammalian RbBP5 were highlighted in blue, cyan, and orange boxes, respectively. Quad-R residues were shown as blue stars. Human and mouse RbBP5 had sequence divergence at five residues at the C-terminus, which were indicated by black dots. The structural part of mouse RbBP5 WD40 repeats, which cover residues 1-380, were identical between human and mouse (100% sequence identity). RbBP5 C-terminus is not included in the crystal structure (gray box). **b**, Multiple sequence alignment of RbBP5 **I**- and A-loops in eukaryotes. **I**- (blue box) and A- (red box) loops were highly conserved from *D. melanogaster* to mammals. Sequence homology analysis by USC and SHP

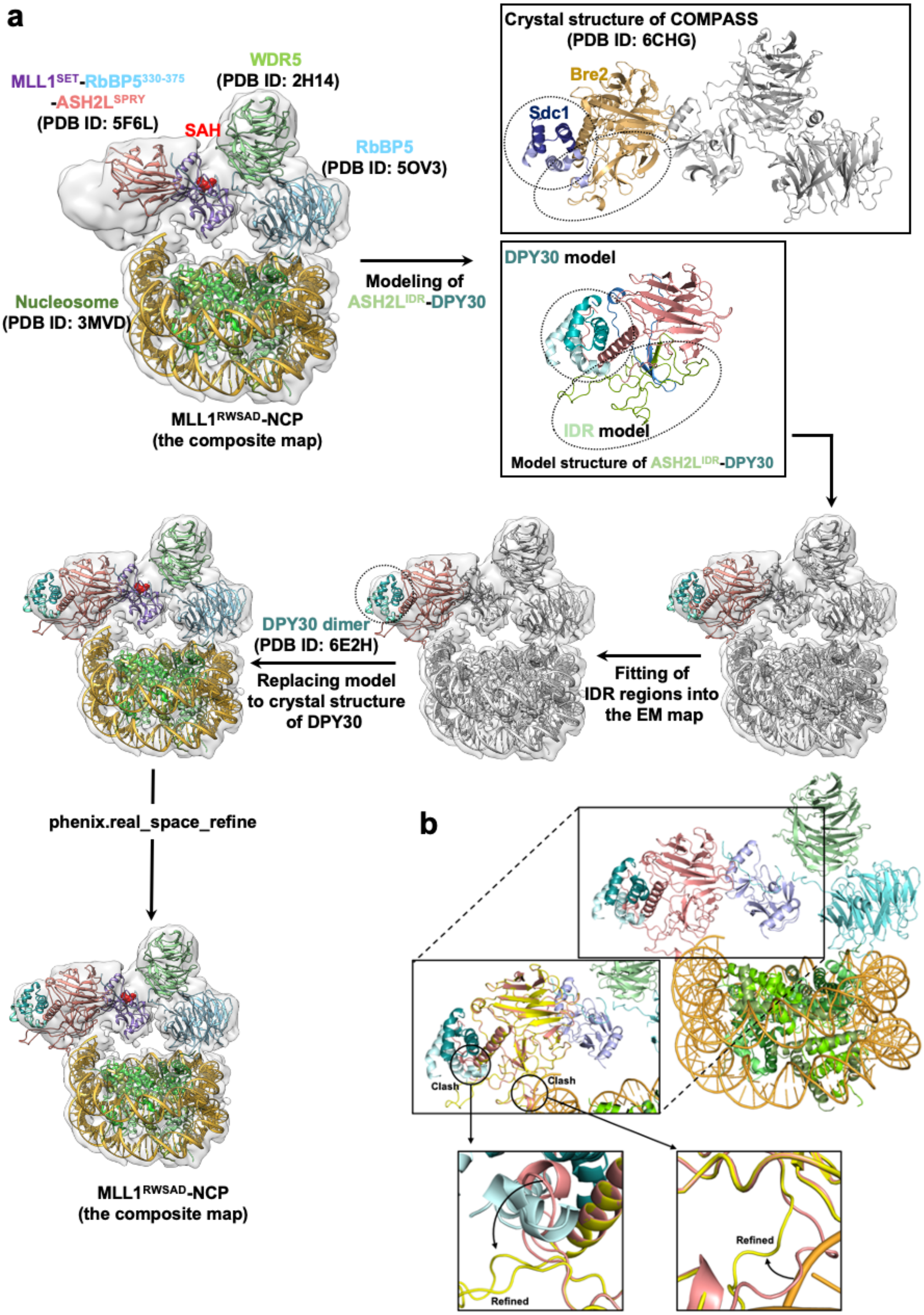

a
MLL1^{SET}-RbBP5$^{330-375}$
-ASH2L^{SPRY}
(PDB ID: 5F6L)
WDR5
(PDB ID: 2H14)
SAH
RbBP5
(PDB ID: 5OV3)
Nucleosome
(PDB ID: 3MVD)
MLL1^{RWSAD}-NCP
(the composite map)
Modeling of
ASH2L^{IDR}-DPY30
Crystal structure of COMPASS
(PDB ID: 6CHG)
Bre2
Sdc1
DPY30 model
IDR model
Model structure of ASH2L^{IDR}-DPY30
Fitting of
IDR regions into
the EM map
DPY30 dimer
(PDB ID: 6E2H)
Replacing model
to crystal structure
of DPY30
phenix.real_space_refine
MLL1^{RWSAD}-NCP
(the composite map)
b
Clash
Clash
Refined
Refined

***Figure S2.5*.** Molecular modeling of the ASH2L-NCP interaction. **a**, Flow chart of the molecular modeling of ASH2L-IDR (see Materials and Methods). The crystal structure of Bre2-Sdc1 from the yeast SET1 complex (PDB ID: 6CHG) [3] was used as a template. After modeling, DPY30 dimer was replaced by crystal structure of DPY30 dimer (PDB ID: 6E2H) [5]. **b**, Fragment-guided molecular dynamic refinement to remove two clashes (black circle: pink-before, yellow-after) between ASH2L-IDR and DPY30 or DNA using the software FG-MD [6]. Molecular modeling by WZ

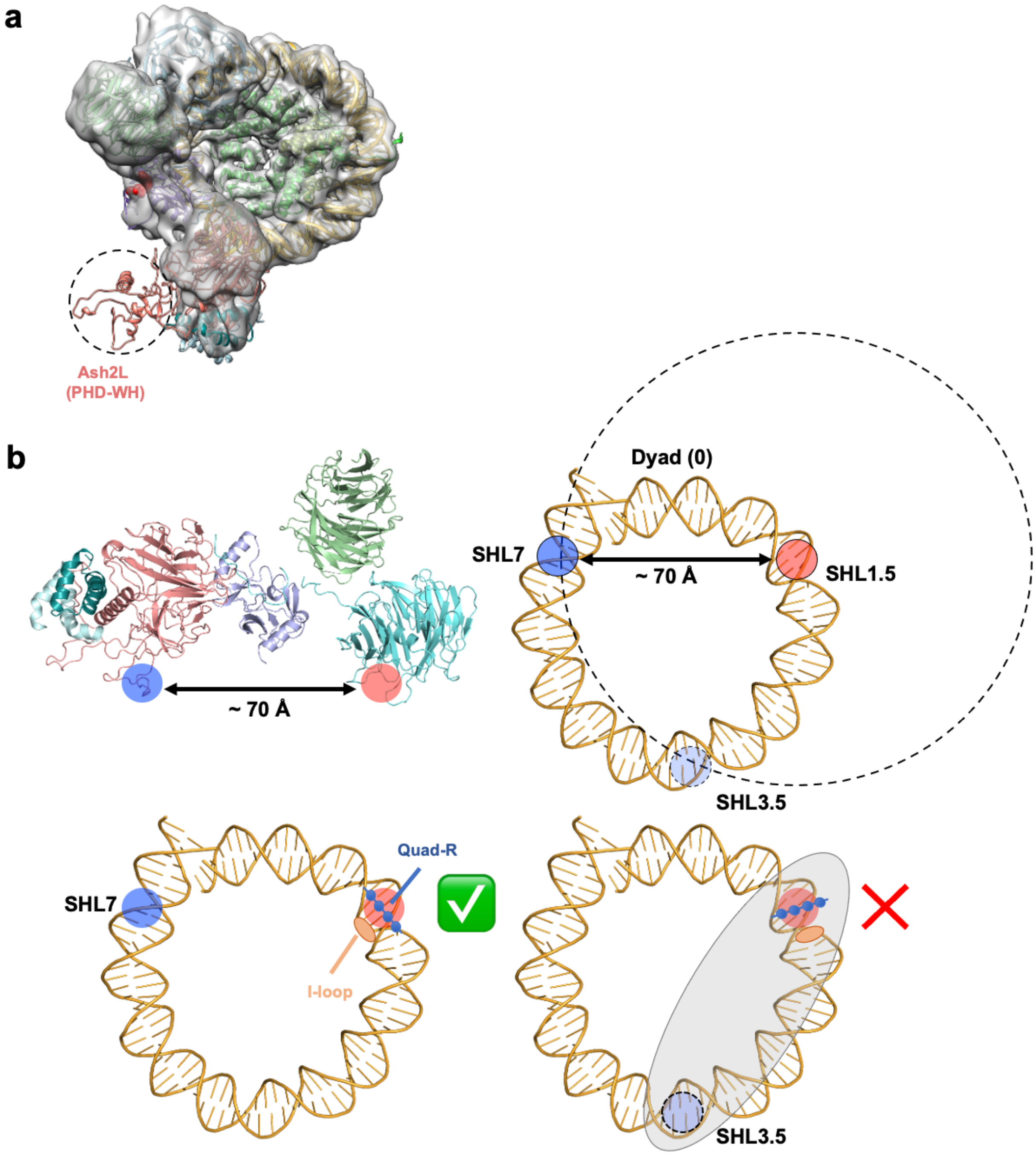


Figure S2.6. ASH2L-NCP interaction in the MLL1^{RWSAD}-NCP complex. **a**, The ASH2L PHD-WH domain (model structure; black dashed circle) was not visible in the cryo-EM map of MLL1^{RWSAD}-NCP. One potential position of the PHD-WH domain based on the structure prediction was shown **b**, The model for coordinated binding to the NCP by RbBP5 and ASH2L. The distance between the I-loop of RbBP5 (red circle) and the $_{205}$KRK$_{207}$ basic patch of ASH2L Linker-IDR (blue circle) is ~70 Å. Specific anchoring of RbBP5 on the NCP confers both orientation and distance constraint for ASH2L-NCP binding. The binding of ASH2L at DNA SHL3.5 may not be allowed due to unfavorable interactions of the Quad-R/DNA and I-loop/H4 tail in the RbBP5-NCP interaction (bottom right). PHD-WH model created by USC

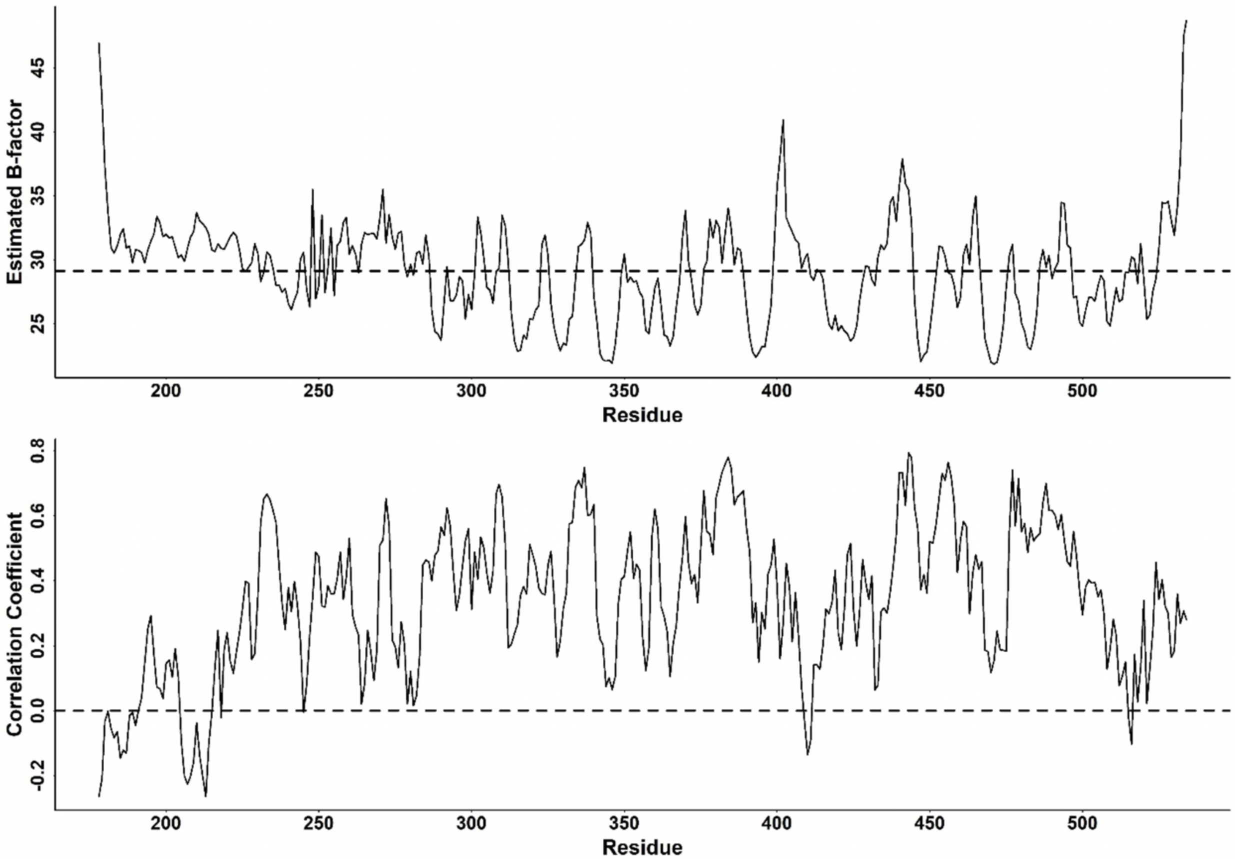


Figure S2.7**.** B-factor estimation for ASH2L. **Top**, Residue-level estimated B-factor as the function of residues of the ASH2L predicted model, where the average B-factor of the model is shown in the dashed line. **Bottom**, Correlation coefficient (CC) between the predicted model and the cryo-EM density map as the function of residues of ASH2L predicted model. Dash lines indicated the average B-factor of all residues of ASH2L (top) and CC value of zero (bottom) B-factor estimation conducted by WZ

Table 2.1. Cryo-EM data collection, refinement, and validation statistics

	MLL1^{RWSAD}-NCP (EMD-20512) (PDB:6PWV)	**MLL1^{RWS}-NCP (EMD-20513) (PDB: 6PWW)**	**RbBP5-NCP (EMD-20514) (PDB: 6PWX)**
Data Collection and Processing			
Magnification	29,000		
Voltage (kV)	300		
Electron exposure (e-/Å^2)	64		
Defocus range (μm)	-1.5 to -3.5		
Pixel size (Å)	1.01		
Symmetry imposed	C1		
Initial particle images (no.)	712,198		
Final particle images (no.)	8,433	21,114	32,563
Map resolution (Å)	6.2	4.5	4.2
FSC threshold	0.143	0.143	0.143
Refinement			
Initial model used (PDB code)	3MVD, 5OV3, 2H14, 5F6L, 6E2H	3MVD, 5OV3, 2H14, 5F6L	3MVD, 5OV3
Model resolution (Å)	6.7	4.7	4.3
FSC threshold	0.5	0.5	0.5
Map sharpening *B* factor (Å^2)	-189	-157	-100
Model composition			
Non-hydrogen atoms	21,781	18,174	14,411
Protein residues	2,005	1,550	1,074
Nucleotides	292	292	292
Ligands	2	2	-
***B* factor (Å^2)**			
Protein	192.99	109.71	107.65
Nucleotide	48.20	36.73	20.20
Ligand	781.00	826.18	-
Rmsds			
Bond lengths (Å)	0.005	0.004	0.003
Bond angles (°)	0.688	0.612	0.579
Validation			
MolProbity score	2.54	2.21	2.38
Clashscore	31.16	21.65	13.50
Poor rotamers (%)	1.42	1.22	3.58
Ramachandran plot			
Favored (%)	93.05	95.54	95.45
Allowed (%)	6.95	4.46	4.55
Disallowed (%)	0	0	0

Appendix A References

1. Kucukelbir, A., F.J. Sigworth, and H.D. Tagare, *Quantifying the local resolution of cryo-EM density maps.* Nature Methods, 2014. **11**(1): p. 63-65.
2. Afonine, P.V., et al., *New tools for the analysis and validation of cryo-EM maps and atomic models.* Acta Crystallogr D Struct Biol, 2018. **74**(Pt 9): p. 814-840.
3. Hsu, P.L., et al., *Crystal Structure of the COMPASS H3K4 Methyltransferase Catalytic Module.* Cell, 2018. **174**(5): p. 1106-1116 e9.
4. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-5.
5. Haddad, J.F., et al., *Structural Analysis of the Ash2L/Dpy-30 Complex Reveals a Heterogeneity in H3K4 Methylation.* Structure, 2018.
6. Zhang, J., Y. Liang, and Y. Zhang, *Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling.* Structure, 2011. **19**(12): p. 1784-95.

APPENDIX B.

Supplementary Information for Chapter 3

This appendix includes supplementary information and figures for Chapter 3 that were used as quality control or clarity in this study.

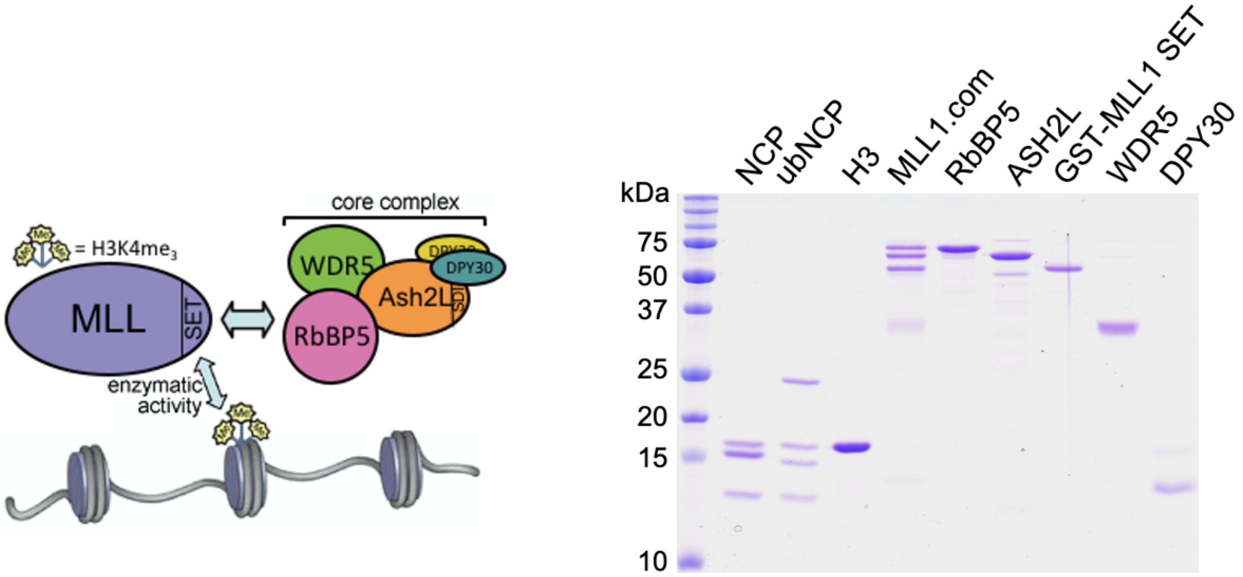


Figure S3.1. DPY30 stimulates MLL1/SET1 activity on the NCP. This figure is related to main Figure 3.1. Left, cartoon model of the subunits that comprise the MLL1 core complex. Full-length components are represented here and used in experiments with the exception of MLL1 protein, where only the C-terminal catalytic SET domain is used in the histone methyltransferase (HMT) assays. Right, Coomassie brilliant blue (CBB) staining of recombinant proteins used in this study including the NCP, the H2BK120ub-containing NCP, recombinant H3, the MLL1 core complex and individual core complex as indicated on top.

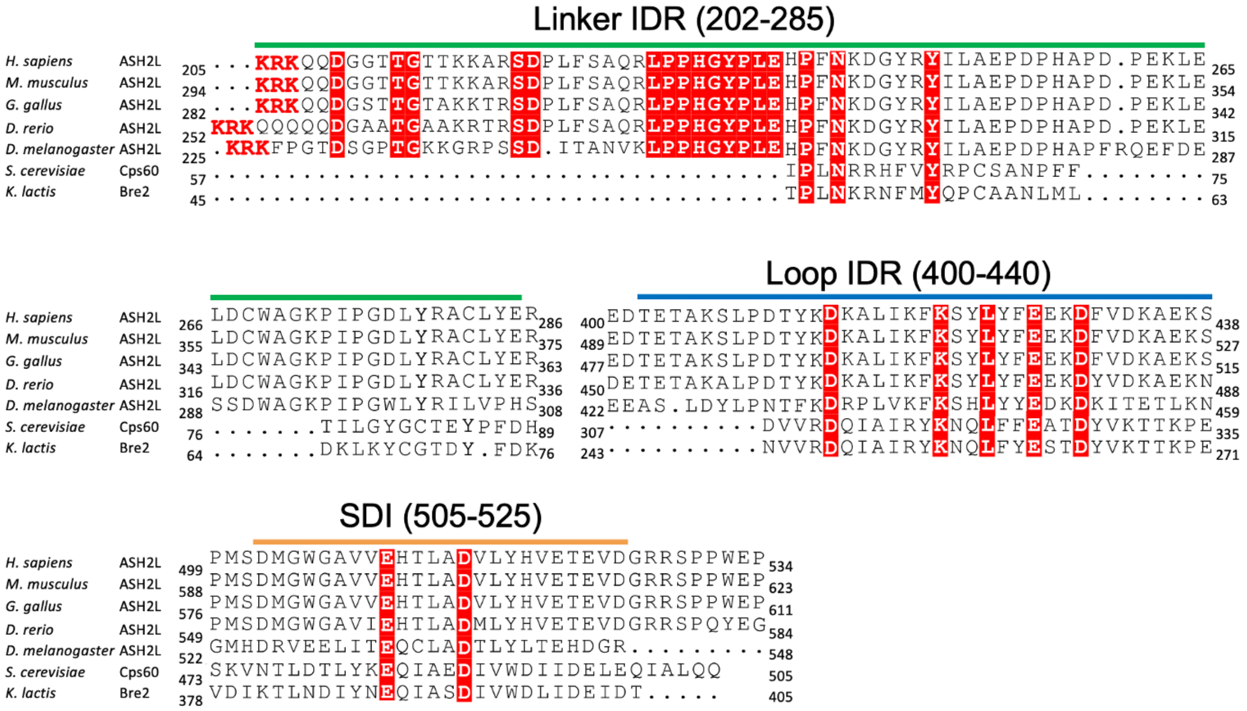

Linker IDR (202-285)
H. sapiens ASH2L 205 ...KRKQQDGGTTGTTKKARSDPLFSAQRLPPHGYPLEHPFNKDGYRYILAEPDPHAPD.PEKLE 265
M. musculus ASH2L 294 ...KRKQQDGGTTGTTKKARSDPLFSAQRLPPHGYPLEHPFNKDGYRYILAEPDPHAPD.PEKLE 354
G. gallus ASH2L 282 ...KRKQQDGSTTGTAKKTRSDPLFSAQRLPPHGYPLEHPFNKDGYRYILAEPDPHAPD.PEKLE 342
D. rerio ASH2L 252 KRKQQQQQDGAATGAAKRTRSDPLFSAQRLPPHGYPLEHPFNKDGYRYILAEPDPHAPD.PEKLE 315
D. melanogaster ASH2L 225 .KRKFPGTDSGPTGKKGRPSSD.ITANVKLPPHGYPLEHPFNKDGYRYILAEPDPHAPFRQEFDE 287
S. cerevisiae Cps60 57 ..................................IPLNRRHFVYRPCSANPFF........ 75
K. lactis Bre2 45 ..................................TPLNKRNFMYQPCAANLML........ 63
Loop IDR (400-440)
H. sapiens ASH2L 266 LDCWAGKPIPGDLYRACLYER 286 400 EDTETAKSLPDTYKDKALIKFKSYLYFEEKDFVDKAEKS 438
M. musculus ASH2L 355 LDCWAGKPIPGDLYRACLYER 375 489 EDTETAKSLPDTYKDKALIKFKSYLYFEEKDFVDKAEKS 527
G. gallus ASH2L 343 LDCWAGKPIPGDLYRACLYER 363 477 EDTETAKSLPDTYKDKALIKFKSYLYFEEKDFVDKAEKS 515
D. rerio ASH2L 316 LDCWAGKPIPGDLYRACLYER 336 450 DETETAKALPDTYKDKALIKFKSYLYFEEKDYVDKAEKN 488
D. melanogaster ASH2L 288 SSDWAGKPIPGWLYRILVPHS 308 422 EEAS.LDYLPNTFKDRPLVKFKSHLYYEDKDKITETLKN 459
S. cerevisiae Cps60 76 .......TILGYGCTEYPFDH 89 307 ..........DVVRDQIAIRYKNQLFFEATDYVKTTKPE 335
K. lactis Bre2 64 ....DKLKYCGTDY.FDK 76 243 ..........NVVRDQIAIRYKNQLFYESTDYVKTTKPE 271
SDI (505-525)
H. sapiens ASH2L 499 PMSDMGWGAVVEHTLADVLYHVETEVDGRRSPPWEP 534
M. musculus ASH2L 588 PMSDMGWGAVVEHTLADVLYHVETEVDGRRSPPWEP 623
G. gallus ASH2L 576 PMSDMGWGAVVEHTLADVLYHVETEVDGRRSPPWEP 611
D. rerio ASH2L 549 PMSDMGWGAVIEHTLADMLYHVETEVDGRRSPQYEG 584
D. melanogaster ASH2L 522 GMHDRVEELITEQCLADTLYLTEHDGR......... 548
S. cerevisiae Cps60 473 SKVNTLDTLYKEQIAEDIVWDIIDELEQIALQQ 505
K. lactis Bre2 378 VDIKTLNDIYNEQIASDIVWDLIDEIDT..... 405

Figure S3.2. IDRs in RbBP5 and MLL1^{SET} are not essential for DPY30-dependent regulation. This figure is related to main Figure 3.3-3.6. Sequence alignment for eukaryotic ASH2L and ASH2L homologs. Highly conserved sequences are highlighted in red. IDR regions of interest (i.e., Linker, Loop, and SDI) are annotated.

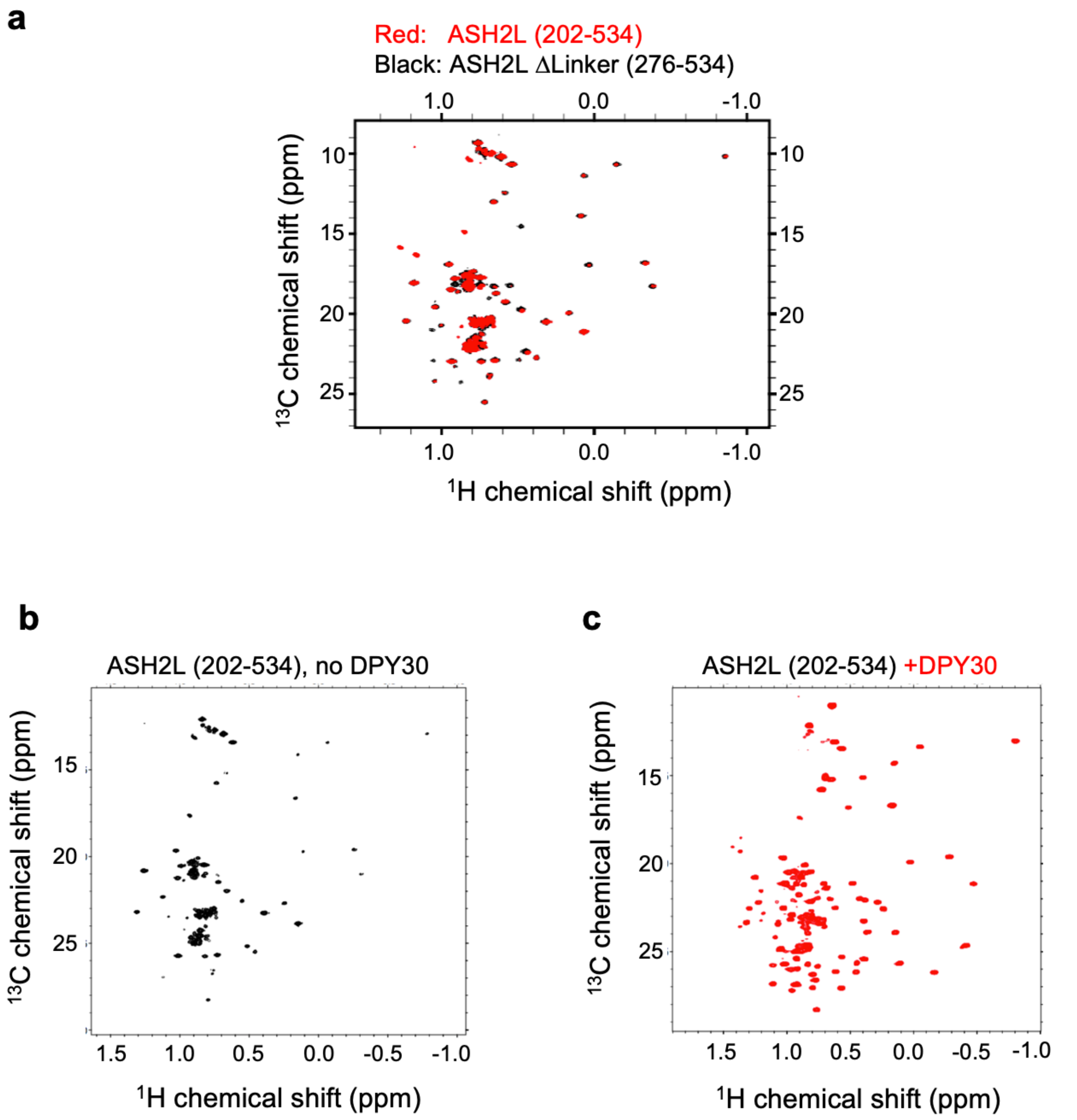


Figure S3.3. NMR spectra for ASH2L. This figure is related to main Figure 3.7 and 3.8. **a**, Linker IDR does not have unique NMR spectra in the apo-state as compared with full length ASH2L. Methyl-TROSY spectra of [^{2}H, ^{13}CH$_3$-ILV] ASH2L$^{202-534}$ (red) and [^{2}H, ^{13}CH$_3$-ILV] ASH2L$^{\Delta Linker,276-534}$ (black) were superimposed. **b**-**c**, ASH2L spectra undergo drastic changes upon addition of DPY30. **b**, Methyl-TROSY spectrum of [^{2}H, ^{13}CH$_3$-ILV] ASH2L$^{202-534}$ in the absence of DPY30. **c**, Methyl-TROSY spectrum of [^{2}H, ^{13}CH$_3$-ILV] ASH2L$^{202-534}$ in complex with DPY30. Stoichiometry between ASH2L and homodimeric DPY30 was 1:1.2. NMR experiments done by YTL

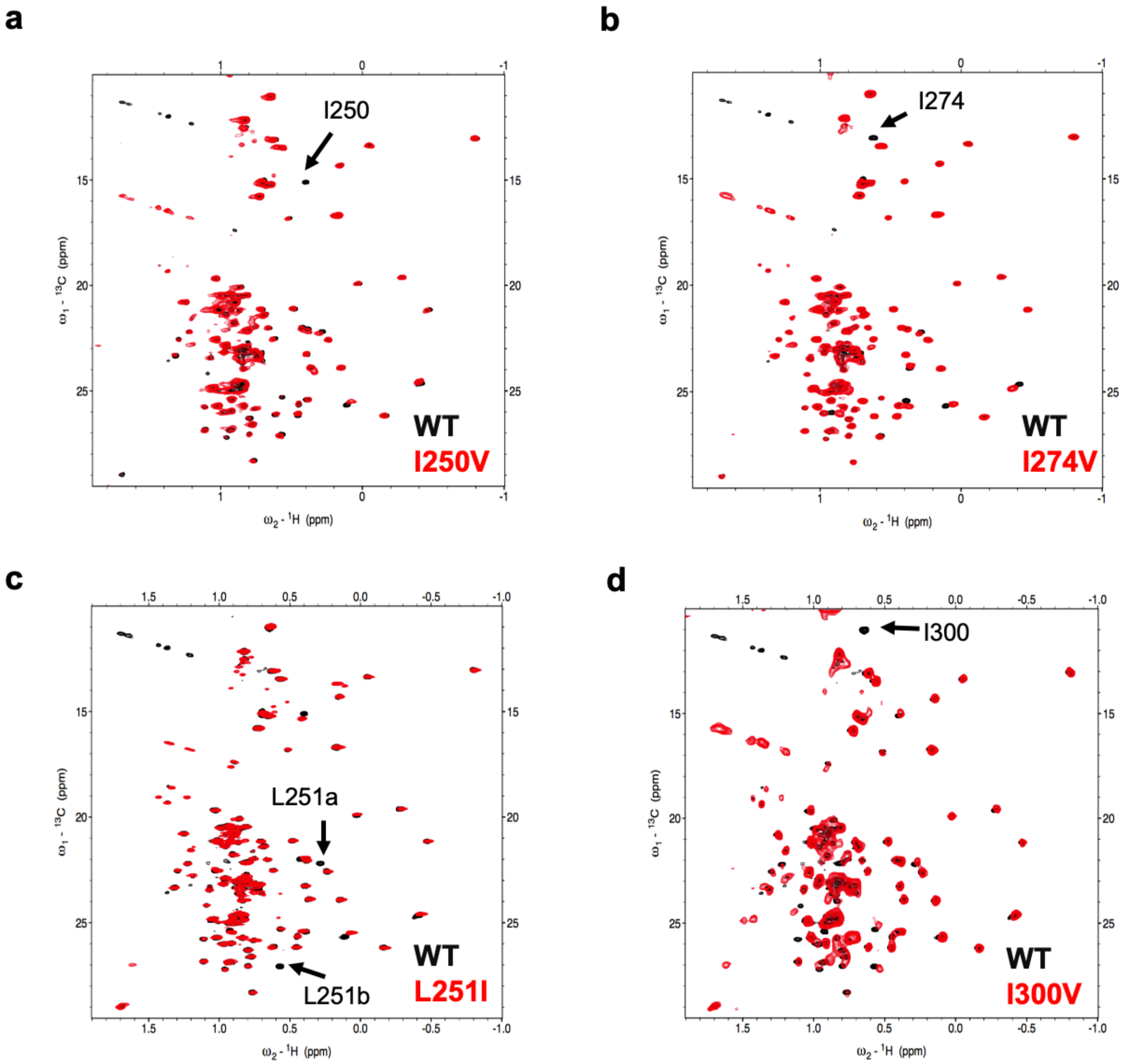


Figure S3.4. Examples of the residue-specific mutagenesis assignment by methyl-TROSY approach. This figure is related to main Figure 3.7 and 3.8. **a**, Superimposed methyl-TROSY spectra of DPY30-bound [^{2}H, ^{13}CH$_3$-ILV] wild-type ASH2L$^{202\text{-}534}$ (black) and single residue mutant ASH2L$^{202\text{-}534,\ I250V}$ constructs (red). Disappeared peak was assigned to the mutated residue. **b**, I274V, **c**, L251I and **d**, I300V were examined for residue-specific mutagenesis assignment by methyl-TROSY. Using this approach, 65% of peaks were assigned without ambiguity. See Table 3.1 in Appendix B. NMR experiments done by YTL

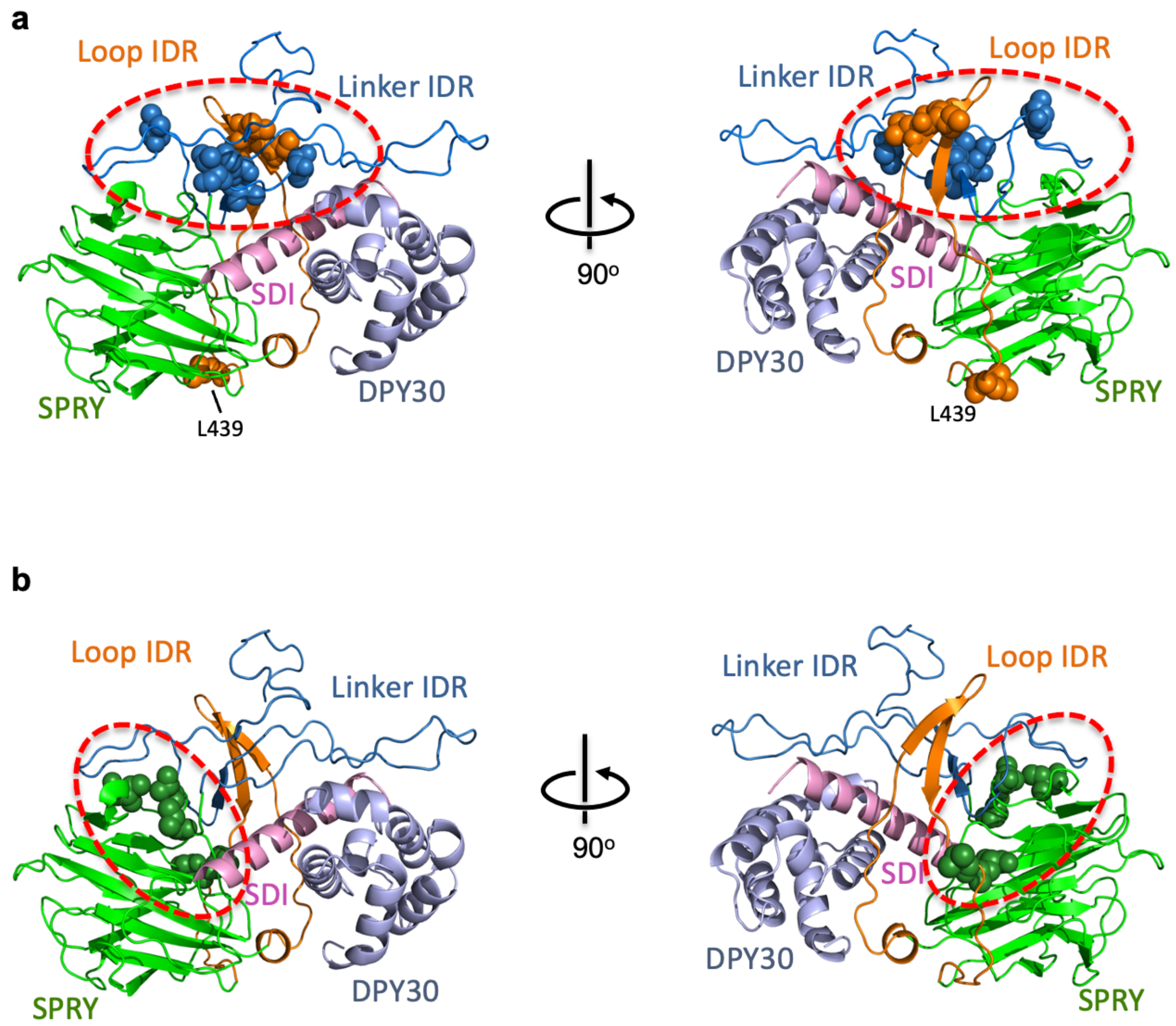


Figure S3.5. Detailed molecular models that highlight ASH2L IDRs and SPRY regions that undergo DPY30-induced changes in NMR spectra. This figure is related to main Figure 3.7 and 3.8. **a**, Newly appeared peaks are highlight by sphere representation in the ASH2L-DPY30 structural model. SPRY, green; Linker IDRs, blue; Loop IDRs, orange; SDI, pink. Different viewpoints for **a**, IDR residues and **b**, SPRY residues are shown. Computational modeling by WZ

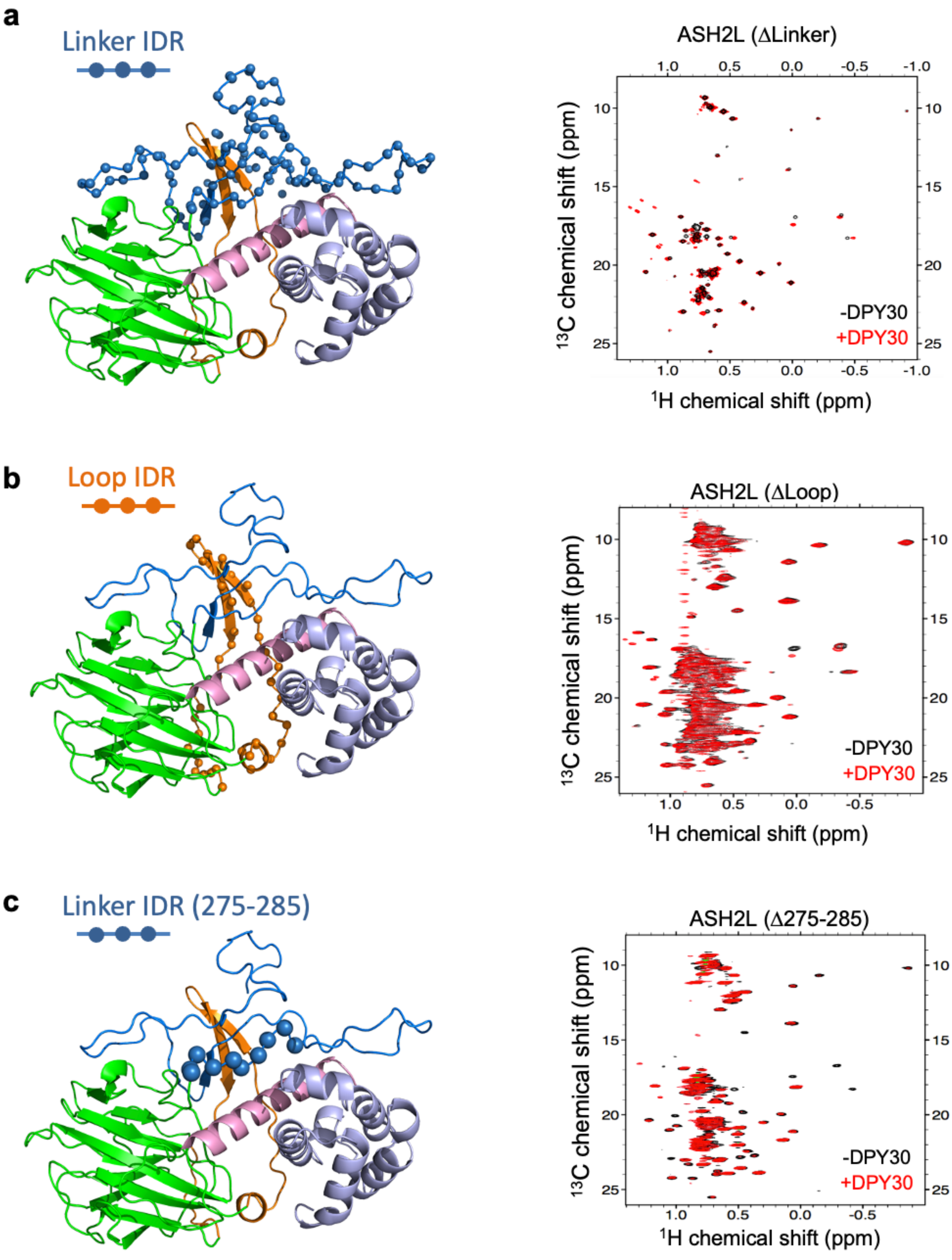

a
Linker IDR
ASH2L (ΔLinker)
1.0 0.5 0.0 -0.5 -1.0
10 15 20 25
^{13}C chemical shift (ppm)
^{1}H chemical shift (ppm)
-DPY30
+DPY30
b
Loop IDR
ASH2L (ΔLoop)
1.0 0.5 0.0 -0.5
10 15 20 25
^{13}C chemical shift (ppm)
^{1}H chemical shift (ppm)
-DPY30
+DPY30
c
Linker IDR (275-285)
ASH2L (Δ275-285)
1.0 0.5 0.0 -0.5
10 15 20 25
^{13}C chemical shift (ppm)
^{1}H chemical shift (ppm)
-DPY30
+DPY30

Figure S3.6. Deletion of ASH2L IDR segments abolished DPY30-induced changes in NMR spectra. This figure is related to main Figure 3.7 and 3.8. **a-c**, Left, computational models that highlight specific ASH2L IDRs in a spherical representation. **a.** Link IDR; **b.** Loop IDR; **c.** residues 275-285 in Linker IDR. Right, superimposed Methyl-TROSY NMR spectra of [^{2}H, ^{13}CH$_3$-ILV] ASH2L$^{202-534}$ with a designated deletion (indicated on top) in absence (black) or presence of homodimeric DPY30 (red). Compared to wild-type ASH2L$^{202-534}$, most DPY30-induced changes were abolished in the ASH2L mutants. SPRY, green; Linker IDRs, blue; Loop IDRs, orange; SDI, pink. ∆Linker NMR by YTL; computational models by WZ

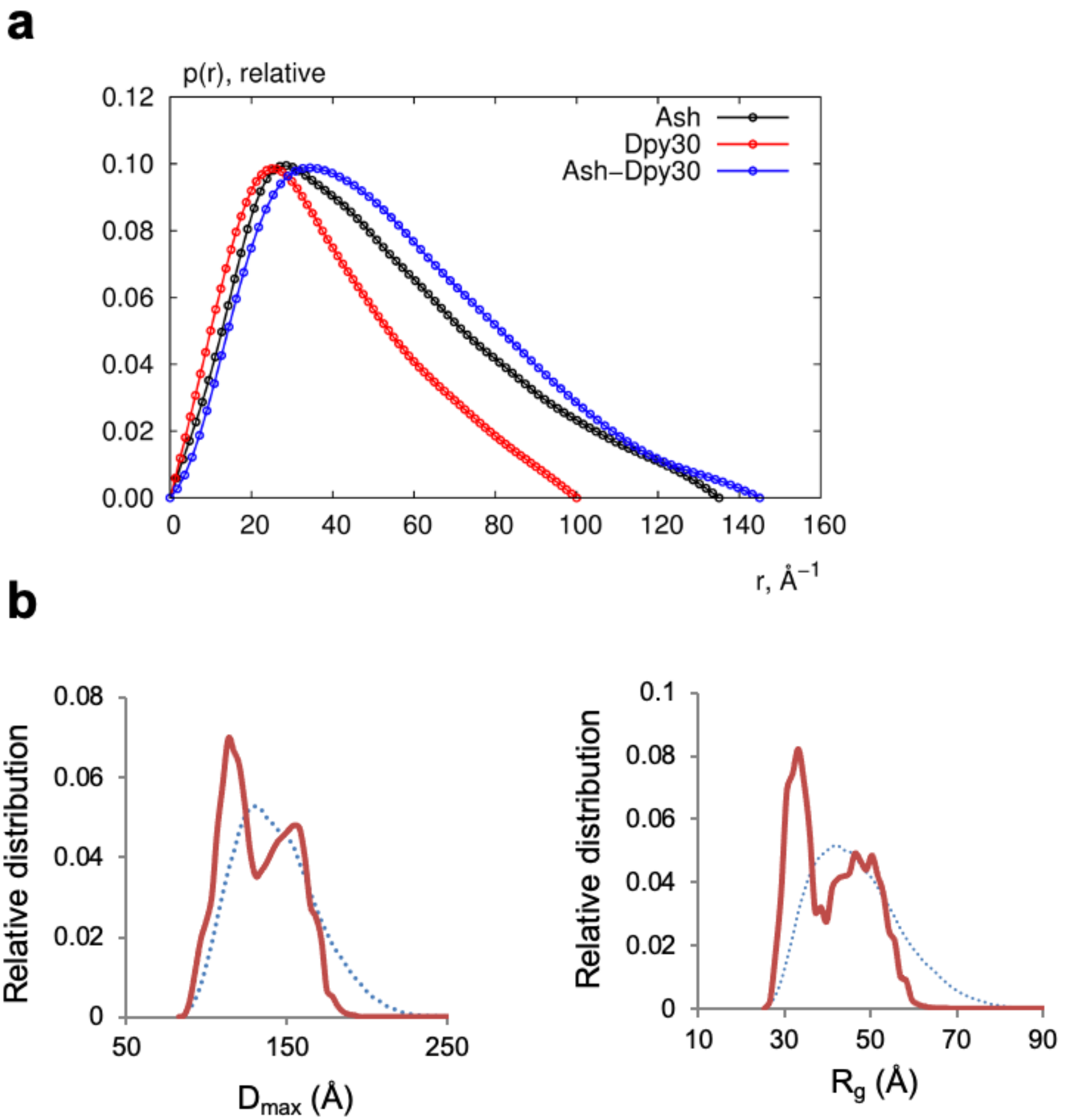


Figure S3.7. SAXS studies on free ASH2L, DPY30 and ASH2L/DPY30. This figure is related to main Figure 3.7 and 3.8. **a**, Pair distance distribution *P(r)* functions of ASH2L, DPY30 and the ASH2L-DPY30 complex. **b**, Ensemble Optimized Method (EOM) analyses for free ASH2L. Distribution of a pool of 10,000 structures (blue) and optimally fit ensemble (red) are plotted against *D$_{max}$* (left) and *R$_g$* (right). SAXS by YTL

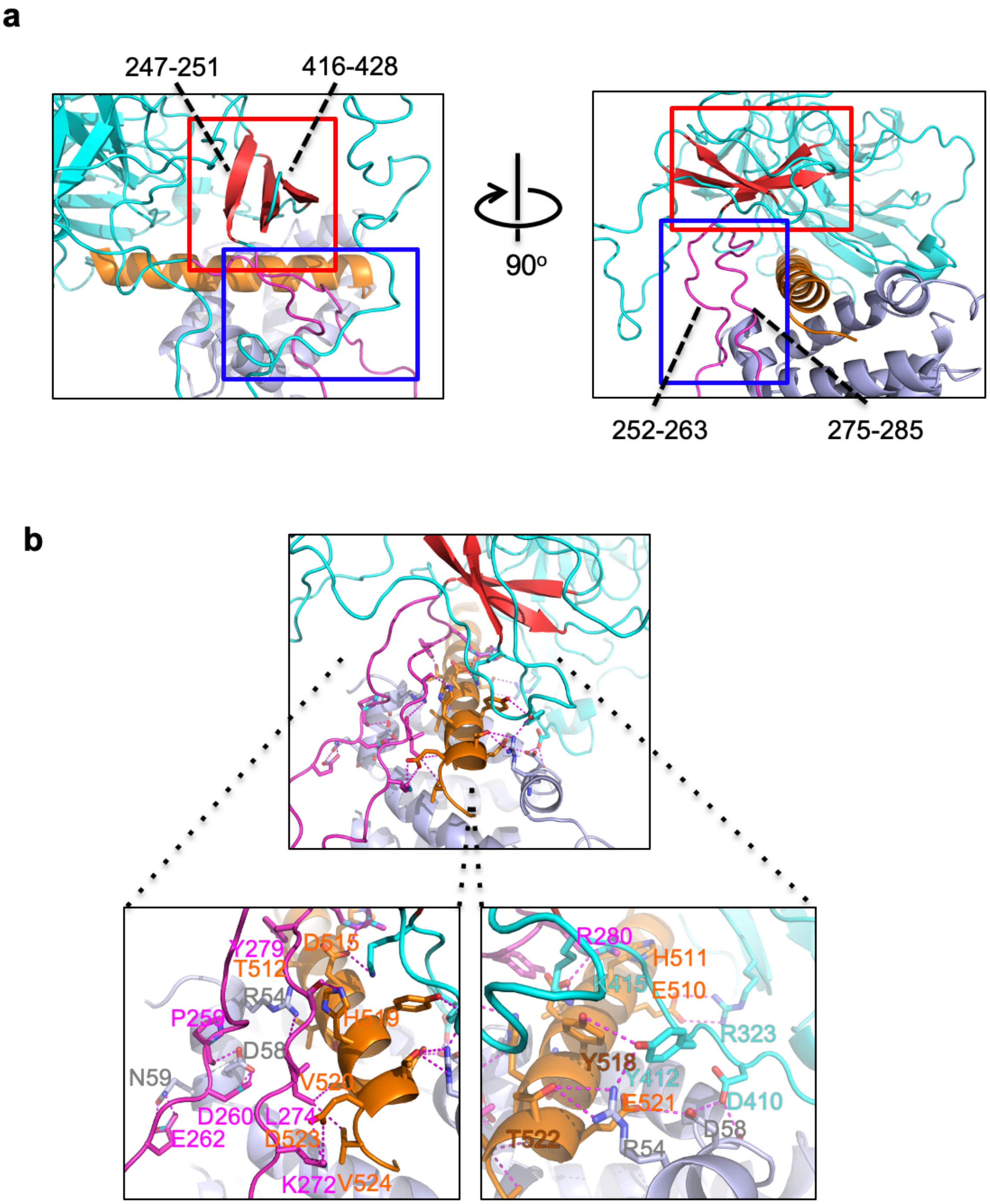


Figure S3.8. The computational model for ASH2L IDR residues that are important for *in vitro* HMTs activities. This figure is related to main Figures 3.1-3.8 **a**, The computational model shows that ASH2L Linker and Loop IDRs likely form a three-strand β-sheet (left) and a β-like structures upon DPY30 interaction. The β-sheet, β-like structure, SPRY domain, SDI of ASH2L as well as DPY30 were labeled in red, pink, cyan, orange, and purple, respectively. **b**, The computational model shows likely interactions among residues in ASH2L IDRs and the α-helical ASH2L SDI. The enlarged interaction interface was shown on bottom. Characteristic secondary structures of ASH2L (β-sheet and β-like, α-helical ASH2L SDI) were shown in red, pink, and orange, respectively. Residues that make potential direct contacts in the model were highlighted. Notably, some of the highlighted ASH2L residues were tested for MLL1 methyltransferase activity on the NCP in main Figures 3.1-3.5. Computational modeling by WZ

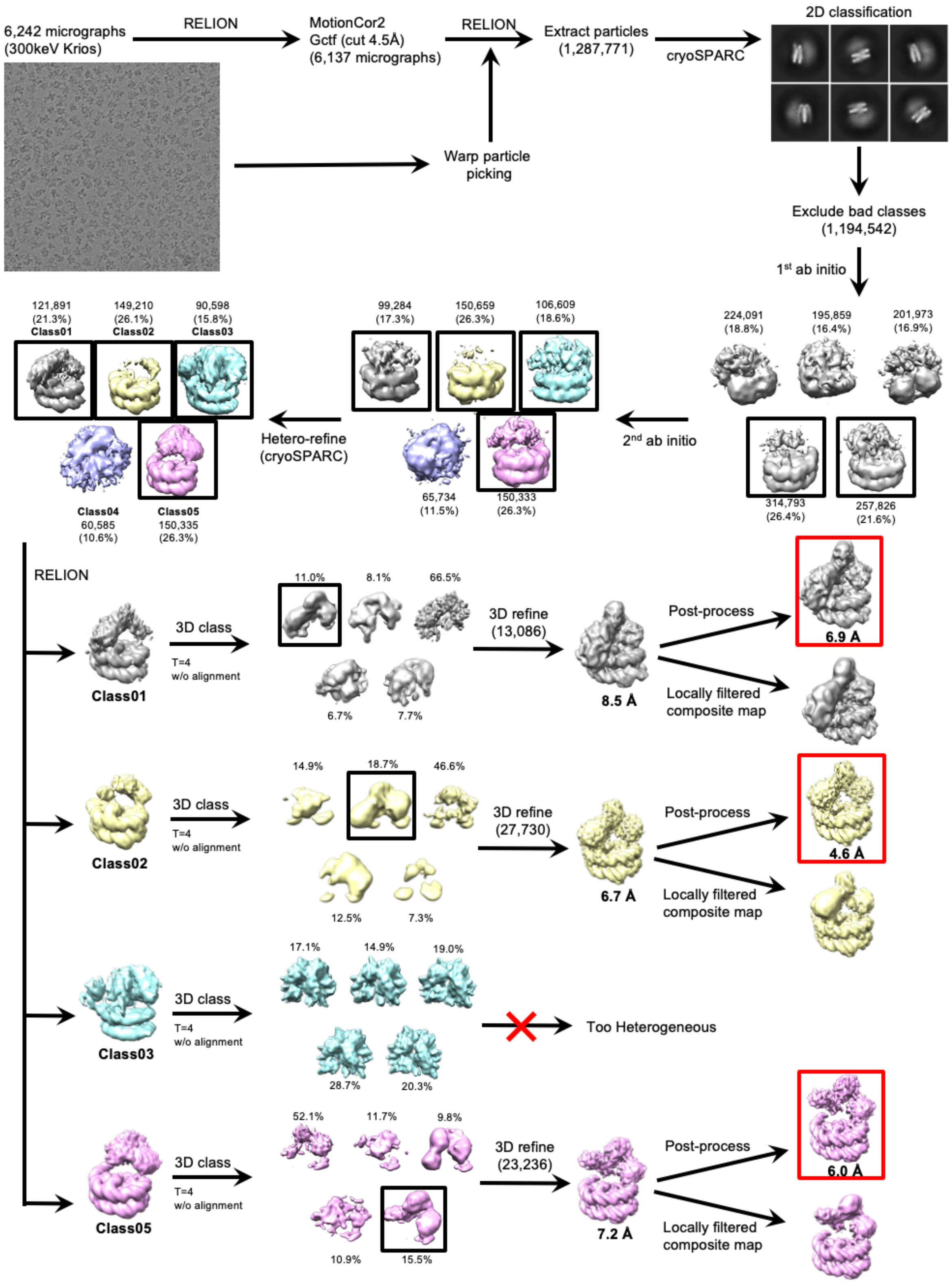


Figure S3.9. Cryo-EM data processing for 4-MLL1-NCP complex. This figure is related to main Figure 3.9. Representative micrograph image (Titan Krios 300 keV) and 2D classifications of the 4-MLL1-NCP complex. The number of particles for each classification and an estimated resolution for overall and selected subcomplexes (red box) were provided. Additional data processing information can be found in the Material and Methods in Chapter 3. Cryo-EM data collection and processing by USC and SHP

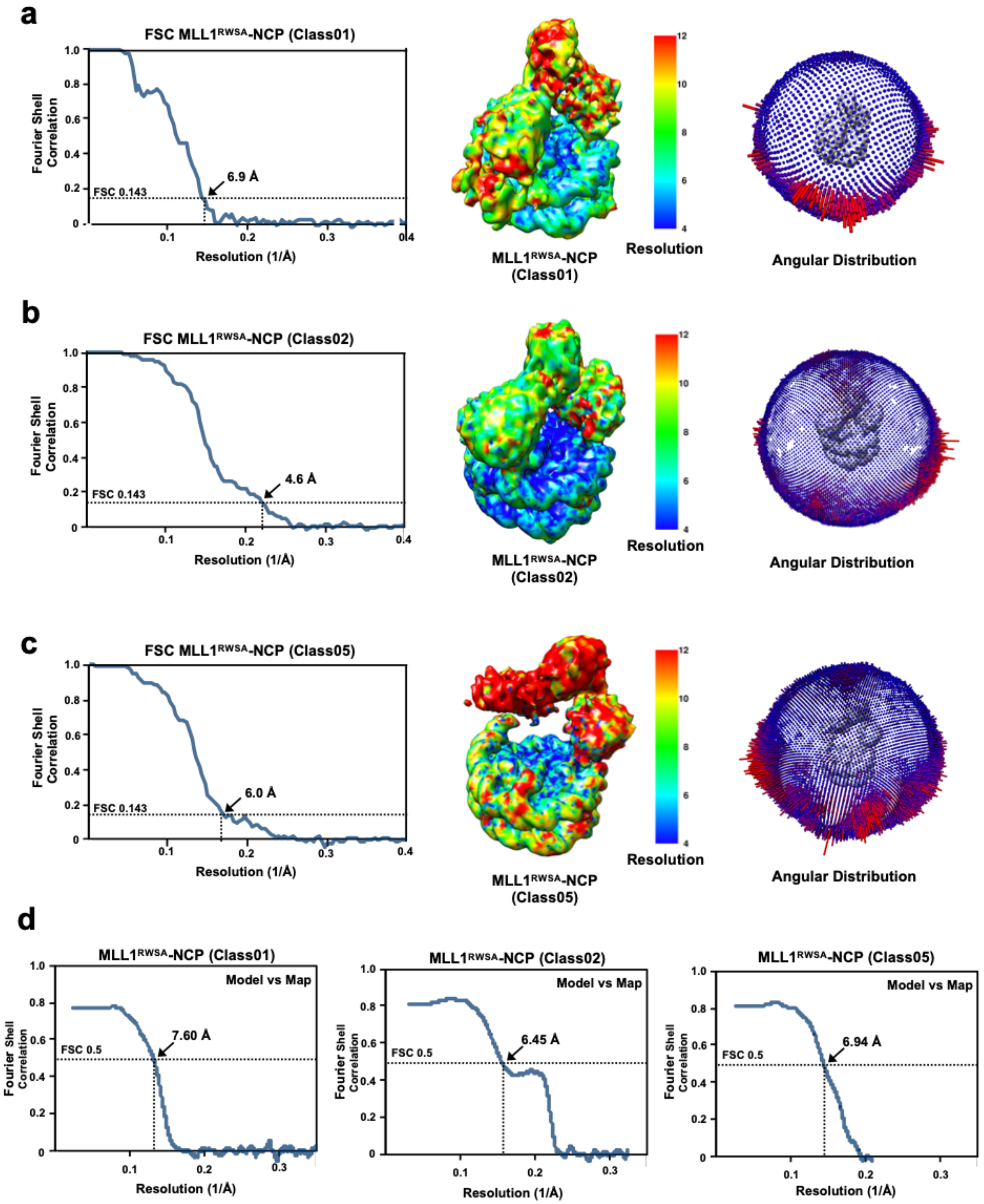


Figure S3.10. Cryo-EM map validation of 4-MLL1-NCP classifications. This figure is related to main Figure 3.9. Fourier Shell Correlation (FSC) curves (left), the corresponding local resolution assessment by RESMAP (middle) [1], and angular distribution plots (right) for the 4-MLL1-NCP **a** (Class01), **b** (Class02), and **c** (Class05) particles. The final resolution was determined using FSC = 0.143 criterion, represented by an arrow on each FSC curve. **d**, Model-map FSC curves for 4-MLL1-NCP Class01, 02 and 05 were calculated using phenix.mtriage [2]. The resolution was found using FSC = 0.5 criterion as indicated by an arrow on each FSC curve. Cryo-EM data collection and processing by USC and SHP

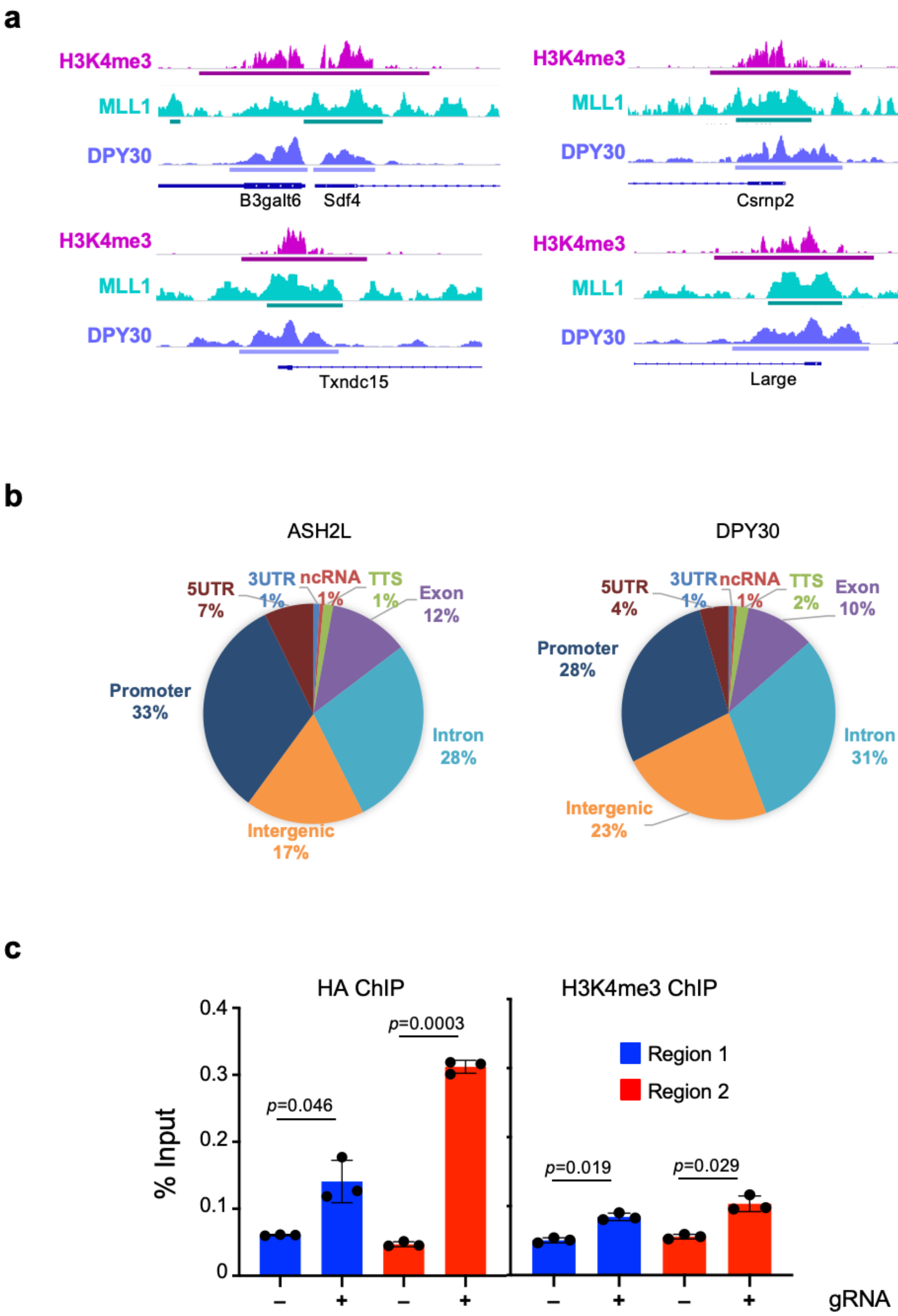


Figure S3.11. *In vivo* analysis of DPY30 and ASH2L binding and H3K4me3. This figure is related to main Figure 3.11. **a**, UCSC browser views of H3K4me3, MLL1 and DPY30 tracks at two randomly selected genomic loci. Peaks called by MACS2 were highlighted on bottom. **b**, Pie charts for distribution of ASH2L (left) and DPY30 (right) binding sites relative to annotated gene structures. **c**, The ChIP assay for HA (left) or H3K4me3 (right) in HA-dCas9 cells transfected with or without the pooled gRNAs. This experiment serves as the control for Figure 3.11b. ChIP signals were normalized against input and presented as %input. Means and standard deviations (error bars) from at least three independent experiments were presented. Two-sided student *t* test was performed to calculate *p*-value. Bioinformatic analysis by FM, ChIP-seq and CUT&RUN by LS, dCas9 experiments by JX

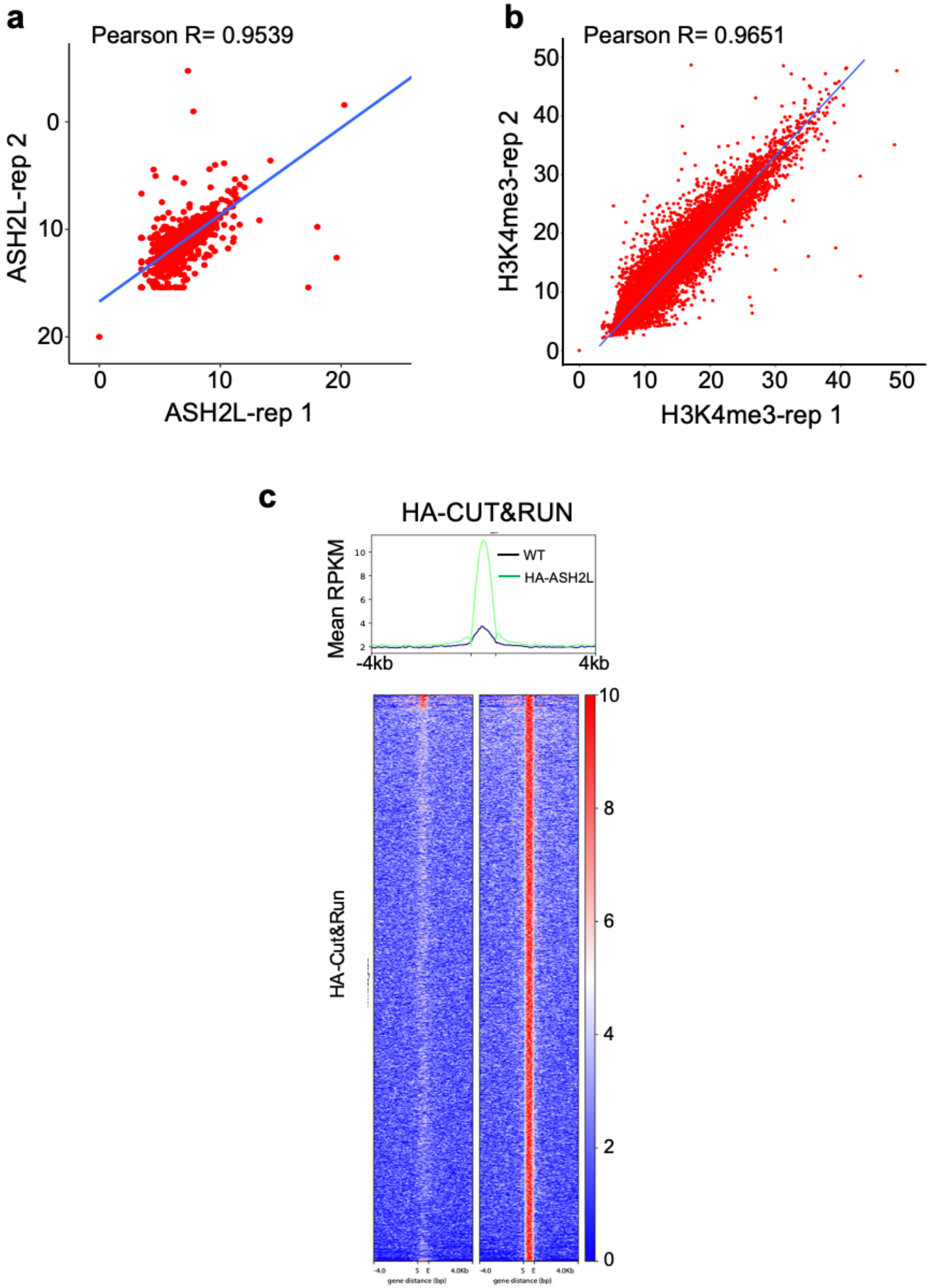


Figure S3.12. Biological duplicates for ASH2L and H3K4me3 CUT&RUN show good correlation and signal-to-noise ratio. This figure is related to main Figure 3.11. **a** and **b**, Scatter plots for peaks in two independent biological replicates of HA (**a**) or H3K4me3 (**b**) CUT&RUN. Pearson correlation coefficient for two samples were shown on top. **c**, Heatmap for HA peaks in the control E14 parental cell line and the HA-ASH2L cell line. Merged signals from biological duplicates were shown, with the heat map key at right. Bioinformatics by FM; CUT&RUN by LS

Table 3.1. Methyl chemical shift of ASH2L$^{202\text{-}534}$ bound to DPY30. This is related to Figure 3.7. Mutational analysis of residues resolved by YTL

Residue	CM1[a]	QM1[a]	CM2[a]	QM2[a]		Residue	CM1[a]	QM1[a]	CM2[a]	QM2[a]
L225	20.459	0.661	21.885	0.75		V391	18.047	1.171	n.d.	n.d.
L231	20.948	0.767	22.188	0.832		L392	19.826	0.163	n.d.	n.d.
L238	19.794	0.947	22.092	0.966		I396	10.303	-0.874		
I250	12.364	0.322				L398	23.874	0.697	n.d.	n.d.
L251	19.437	0.207	24.297	0.489		L408	n.d.	n.d.	n.d.	n.d.
L264	n.d.[b]	n.d.	n.d.	n.d.		L417	21.913	-0.492	22.938	0.03
L266	n.d.	n.d.	n.d.	n.d.		I418	12.245	0.618		
I274	10.333	0.549				L424	21.157	0.297	22.659	0.324
L278	20.459	0.636	23.211	0.845		V432	20.154	0.877	n.d.	n.d.
L283	19.257	0.351	23.385	0.539		L439	19.779	0.54	23.441	-0.238
V287	16.882	-0.355	17.179	-0.049		I447	11.562	0.077		
L288	19.461	1.142	24.123	0.855		I448	10.728	0.485		
L289	n.d.	n.d.	n.d.	n.d.		V454	17.723	0.734	n.d.	n.d.
L291	20.615	0.654	22.66	0.844		V458	18.456	0.79	18.659	0.612
L298	n.d.	n.d.	n.d.	n.d.		I463	10.616	-0.123		
I300	8.308	0.564				V467	n.d.	n.d.	n.d.	n.d.
L305	20.588	1.238	22.974	0.938		I472	12.475	0.57		
V307	n.d.	n.d.	n.d.	n.d.		L474	n.d.	n.d.	n.d.	n.d.
V308	16.942	0.951	17.776	0.905		V480	n.d.	n.d.	n.d.	n.d.
V316	17.723	0.734	21.441	1.012		I482	13.957	0.094		
V322	18.408	-0.545	n.d.	n.d.		L495	20.52	0.316	21.173	0.069
I331	13.053	0.645				V508	19.812	1.213	n.d.	n.d.
V333	n.d.	n.d.	n.d.	n.d.		V509	n.d.	n.d.	n.d.	n.d.
L344	22.929	0.373	23.052	1.032		L513	n.d.	n.d.	n.d.	n.d.
L350	19.329	0.304	23.416	0.378		V516	n.d.	n.d.	n.d.	n.d.
L353	19.284	0.581	n.d.	n.d.		L517	n.d.	n.d.	n.d.	n.d.
L357	n.d.	n.d.	n.d.	n.d.		V520	n.d.	n.d.	n.d.	n.d.
I378	9.422	0.744				V524	19.023	0.822	n.d.	n.d.

[a] Methyl resonances are arbitrarily listed without stereospecific assignment for Leu and Val. For Ile, CM1 and QM1 are equivalent to CD1 and QD1, respectively.

[b] Not determined due to strong ambiguity.

Table 3.2. Cryo-EM data collection, refinement, and validation statistics. This is related to main Figure 3.9, and Figures S3.9 and S3.10 in Appendix B.

	4-MLL1-NCP, Class01 (EMD-21542) (PDB: 6W5I)	**4-MLL1-NCP, Class02 (EMD-21543) (PDB: 6W5M)**	**4-MLL1-NCP, Class05 (EMD-21544) (PDB: 6W5N)**
Data Collection and Processing			
Magnification	29,000		
Voltage (kV)	300		
Electron exposure (e-/Å^{2})	64		
Defocus range (μm)	-1.5 to -2.5		
Pixel size (Å)	1.00		
Symmetry imposed	C1		
Initial particle images (no.)	1,287,711		
Final particle images (no.)	13,086	27,730	23,236
Map resolution (Å)	6.9	4.6	6.0
FSC threshold	0.143	0.143	0.143
Refinement			
Initial model used (PDB code)	6PWV	6PWV	6PWV
Model resolution (Å)	7.6	6.5	6.9
FSC threshold	0.5	0.5	0.5
Map sharpening *B* factor (Å^{2})	-442.70	-177.74	-199.18
Model composition			
Non-hydrogen atoms	19,672	19,574	19,667
Protein residues	1,741	1,729	1,740
Nucleotides	292	292	292
Ligands	-	-	-
***B* factor (Å^{2})**			
Protein	214.20	195.04	221.62
Nucleotide	50.98	31.39	51.78
Ligand	-	-	-
Rmsds			
Bond lengths (Å)	0.005	0.005	0.005
Bond angles (°)	0.674	0.673	0.671
Validation			
MolProbity score	2.68	2.59	2.69
Clashscore	44.52	36.36	44.38
Poor rotamers (%)	1.63	1.58	1.63
Ramachandran plot			
Favored (%)	94.21	94.06	94.09
Allowed (%)	5.79	5.94	5.91
Disallowed (%)	0	0	0

Table 3.3. Survey of IDR content in histone methyltransferases. This is related to Figure 3.3a.

Modification Site	Enzyme	Length (AAs)	IDR (%)
H3R2	CARM1	608	50
	PRMT6	375	12
H3K4	PRDM16	1276	71
	SMYD3	428	0
	MLL1	3969	84
	MLL3	4911	76
H3K9	Suv39H1	412	36
	G9a	1210	66
	SETDB1	1291	38
H3K36	SETD2	2564	90
	NSD2	1365	73
	ASH1L	2969	85
H3K79	Dot1L	1739	82

Table 3.4. gRNA and primer sequence information.

gRNAs:

gRNA1-1 sense	CACCGCCCTCTGATCTGTAGCGCAG
gRNA1-1 antisense	AAACCTGCGCTACAGATCAGAGGGC
gRNA1-2 sense	CACCGAGCTGGGTGGTGGACAATGC
gRNA1-2 antisense	AAACGCATTGTCCACCACCCAGCTC
gRNA1-3 sense	CACCGAAGTGCCCAGGGATGATTGA
gRNA1-3 antisense	AAACTCAATCATCCCTGGGCACTTC
gRNA2-1 sense	CACCGTCCTGTGAGGTCCTGCGAAA
gRNA2-1 antisense	AAACTTTCGCAGGACCTCACAGGAC
gRNA2-2 sense	CACCGTGAGGCTAAGGTAATTCAGC
gRNA2-2 antisense	AAACGCTGAATTACCTTAGCCTCAC
gRNA2-3 sense	CACCGCATCTCTGCGTATAGACCAC
gRNA2-3 antisense	AAACGTGGTCTATACGCAGAGATGC

Primers:

CHIP-region 1-F	AGGTCTAACTCAGGCTCCCG
CHIP-region 1-R	ACTGAAGTGACATGTGCGTGTG
CHIP-region 2-F	TGCTGCATTGCCTGTCTTGCT
CHIP-region 2-R	GGTTGCTTACACCTGCCTGTAAC

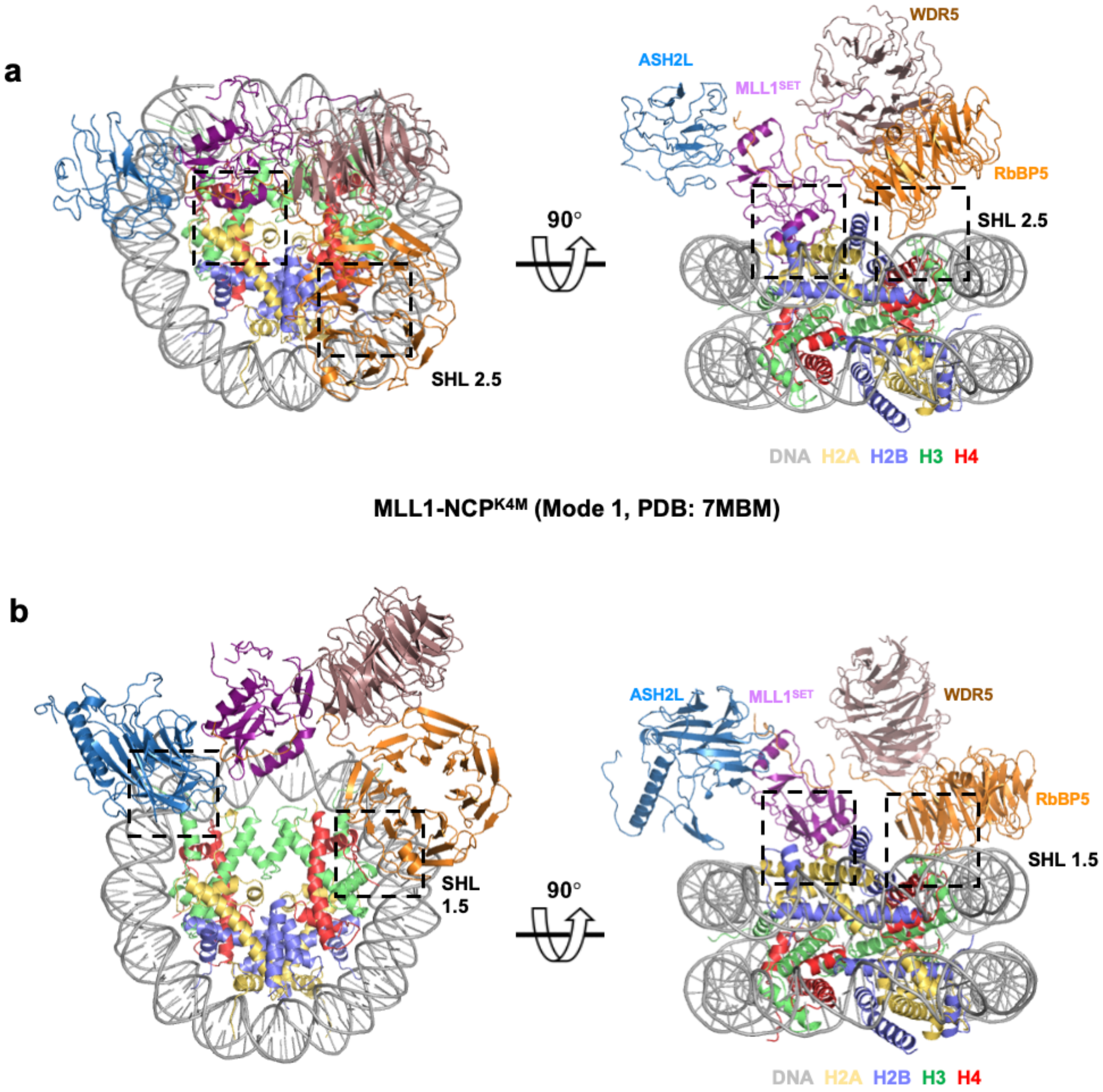
Appendix B References

1. Kucukelbir, A., F.J. Sigworth, and H.D. Tagare, *Quantifying the local resolution of cryo-EM density maps.* Nature Methods, 2014. **11**(1): p. 63-65.
2. Afonine, P.V., et al., *Real-space refinement in PHENIX for cryo-EM and crystallography.* Acta Crystallogr D Struct Biol, 2018. **74**(Pt 6): p. 531-544.

APPENDIX C.

Supplementary Information for Chapter 4

This appendix includes supplementary information and figures for Chapter 4 that were used as quality control or clarity in this study.



MLL1-NCP^{K4M} (Mode 1, PDB: 7MBM)

MLL1- NCP^{K4M} (Mode 2, PDB: 7MBN)

Figure S4.1. MLL1-NCP^{K4M} mode 1 and 2 exploit unique interfaces to bind the NCP. **a**, Top (left) and left (right) views of MLL1^{RWSAD}-H3K4M NCP$^{mode\ 1}$ cryo-EM structure (PDB: 7MBM, EMDB: 23738). **b**, Top (left) and front (right) views of MLL1^{RWSAD}-H3K4M NCP$^{mode\ 2}$ cryo-EM structure (PDB: 7MBN, EMDB: 23739). **a-b**, Ribbon representations of each core complex component are color-coded as follows: ASH2L (light blue), MLL1SET (purple), WDR5 (tan), and RbBP5 (orange). 147 bp 601 DNA is gray and core histones H2A (yellow), H2B (lavender), H3 (green), and H4 (red) are shown. Unique interaction motifs are in dashed boxed regions. Cryo-EM data collected and processed by USC and SHP

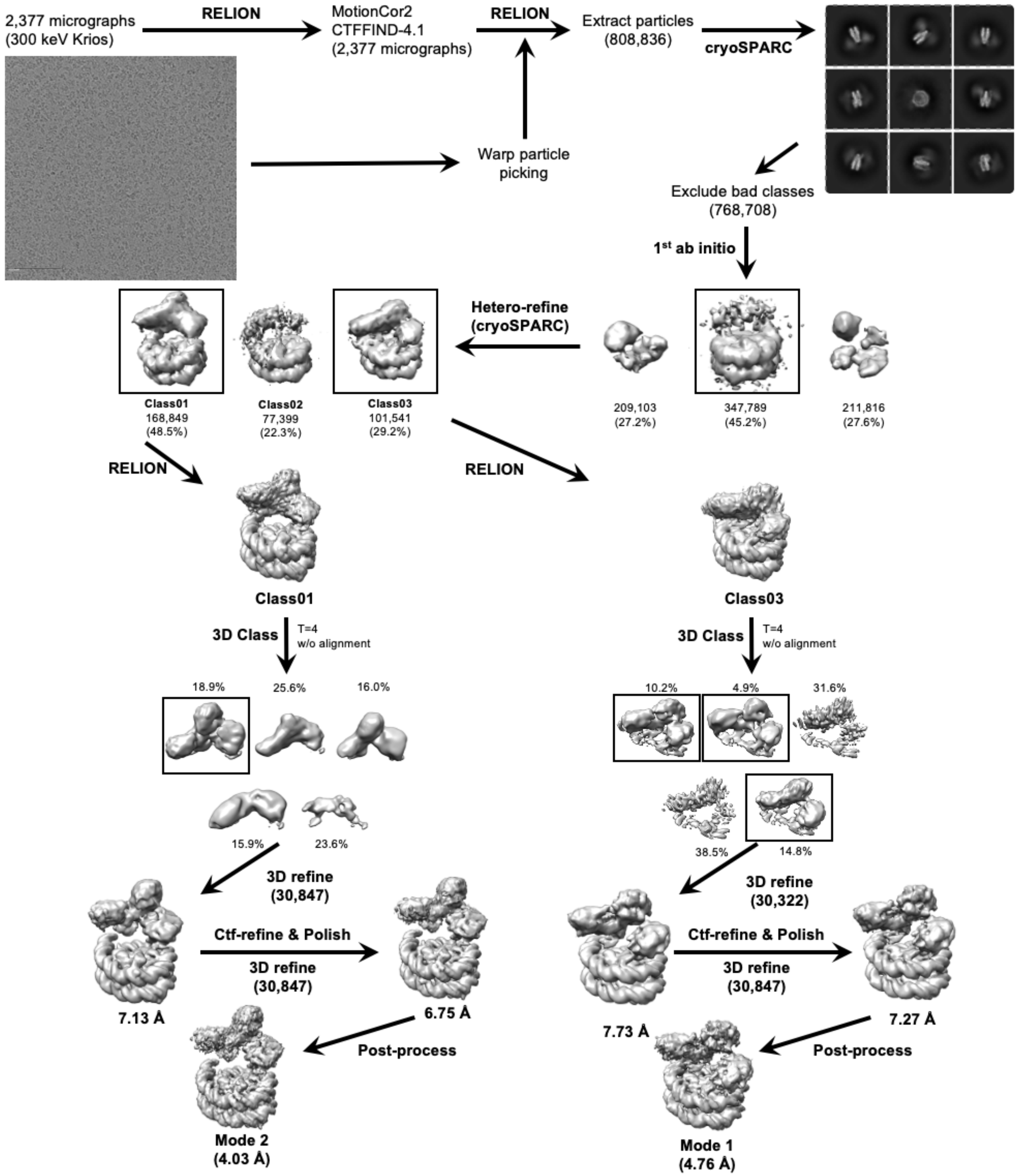


Figure S4.2. Cryo-EM data processed for the MLL1^{RWSAD}-NCP^{K4M}. Representative micrograph image (Titan Krios 300 keV) and 2D classifications of the MLL1^{RWSAD}-NCP^{K4M} complex are shown. Particle numbers for each classification step are shown. Estimated resolutions of each classification are provided. The details of data processing steps for MLL1^{RWSAD}-NCP^{K4M} and programs used therein are described in detail in the Methods section of Chapter 4. Cryo-EM data collected and processed by USC and SHP

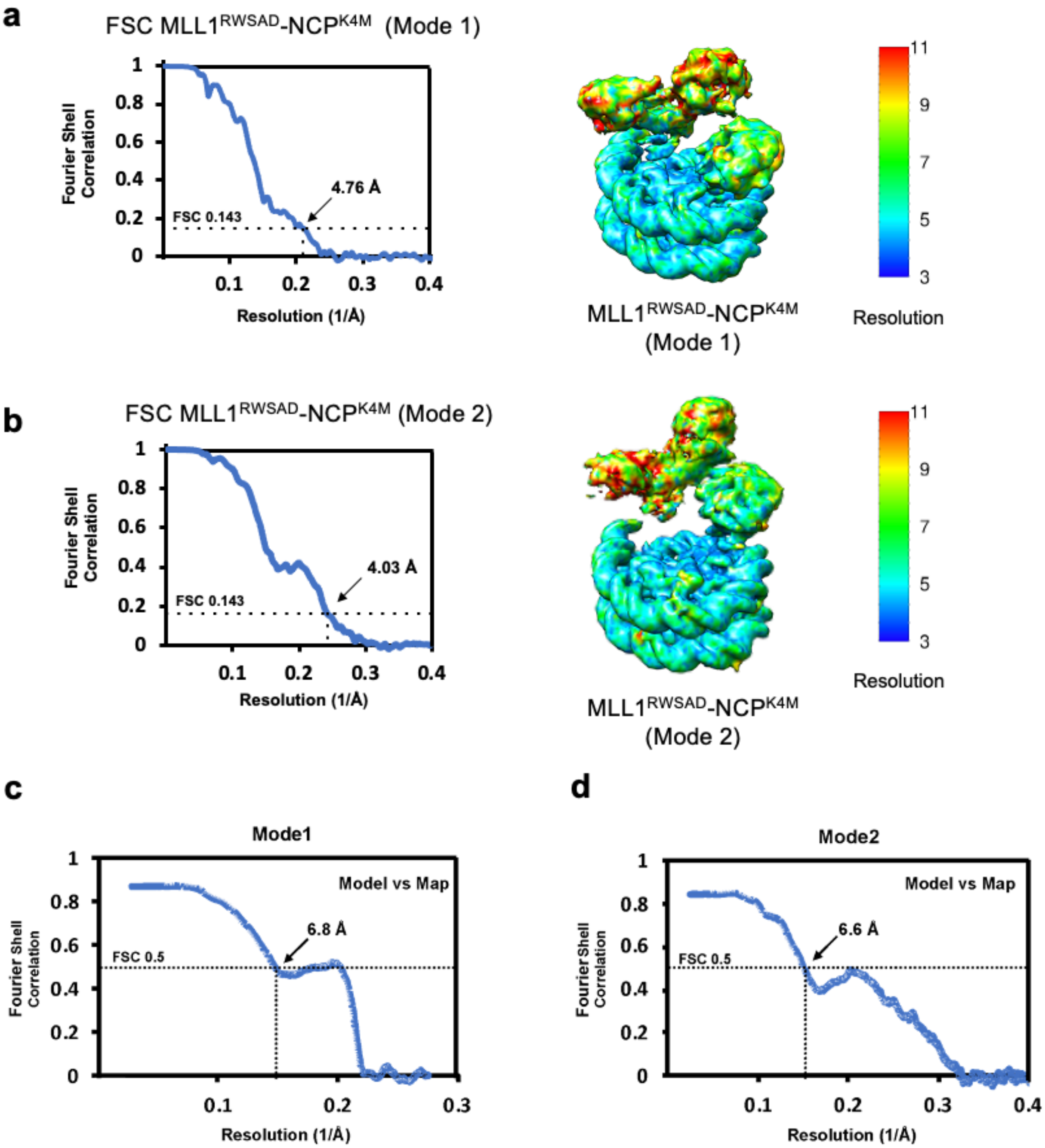


***Figure S4.3*.** Cryo-EM map validation of MLL1^{RWSAD}-NCP^{K4M}**. a** and **b**, Fourier Shell Correlation (FSC) curve for MLL1^{RWSAD}-NCP^{K4M} is shown on the left along with the corresponding local resolution as determined using RESMAP1. The final resolution was determined using FSC = 0.143 criterion, designated with an arrow on the FSC curve. **c** and **d**, Model-map FSC curve for MLL1^{RWSAD}-NCP^{K4M}. The resolution was indicated using FSC = 0.5 criterion, shown by an arrow on the FSC curve.

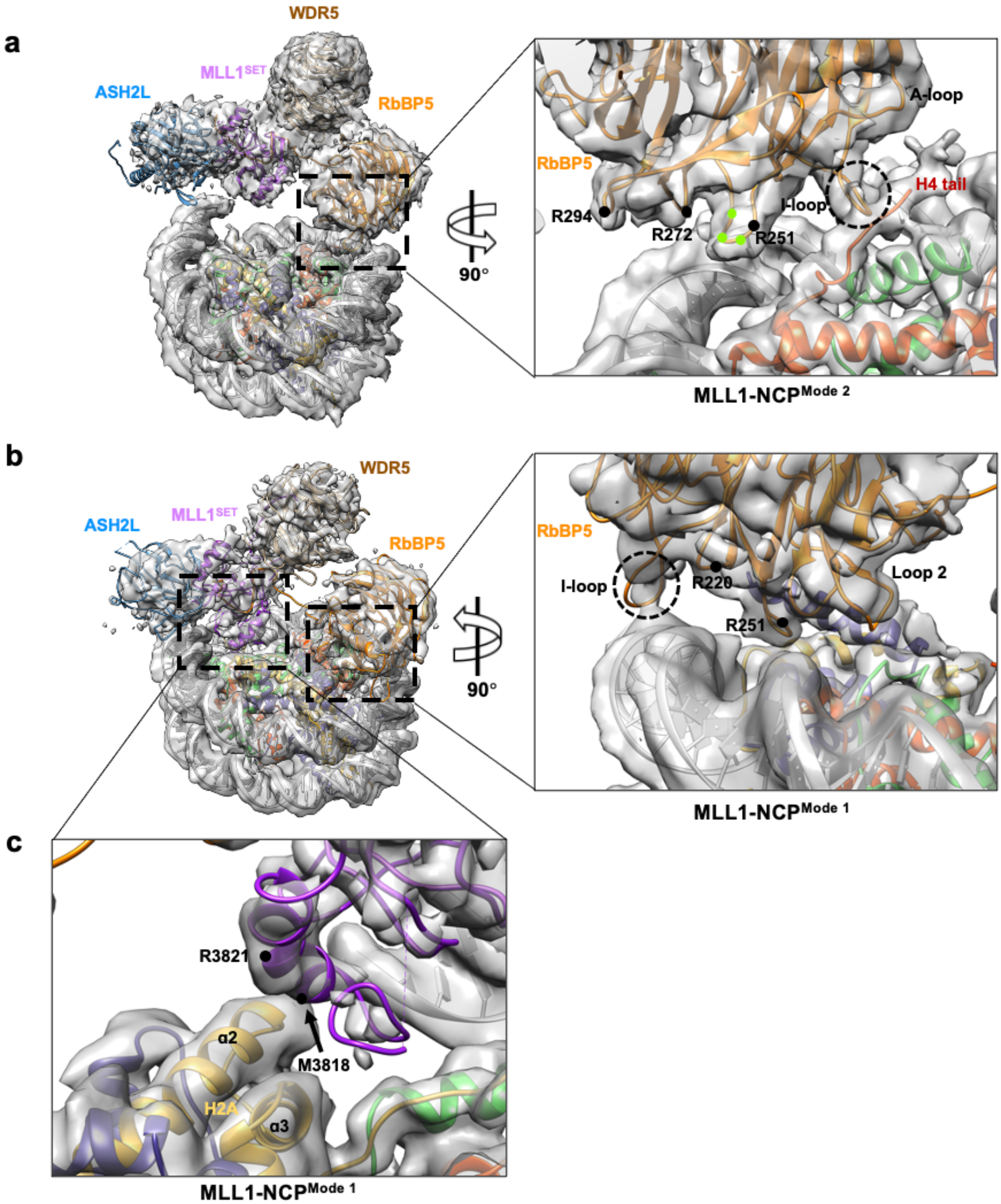
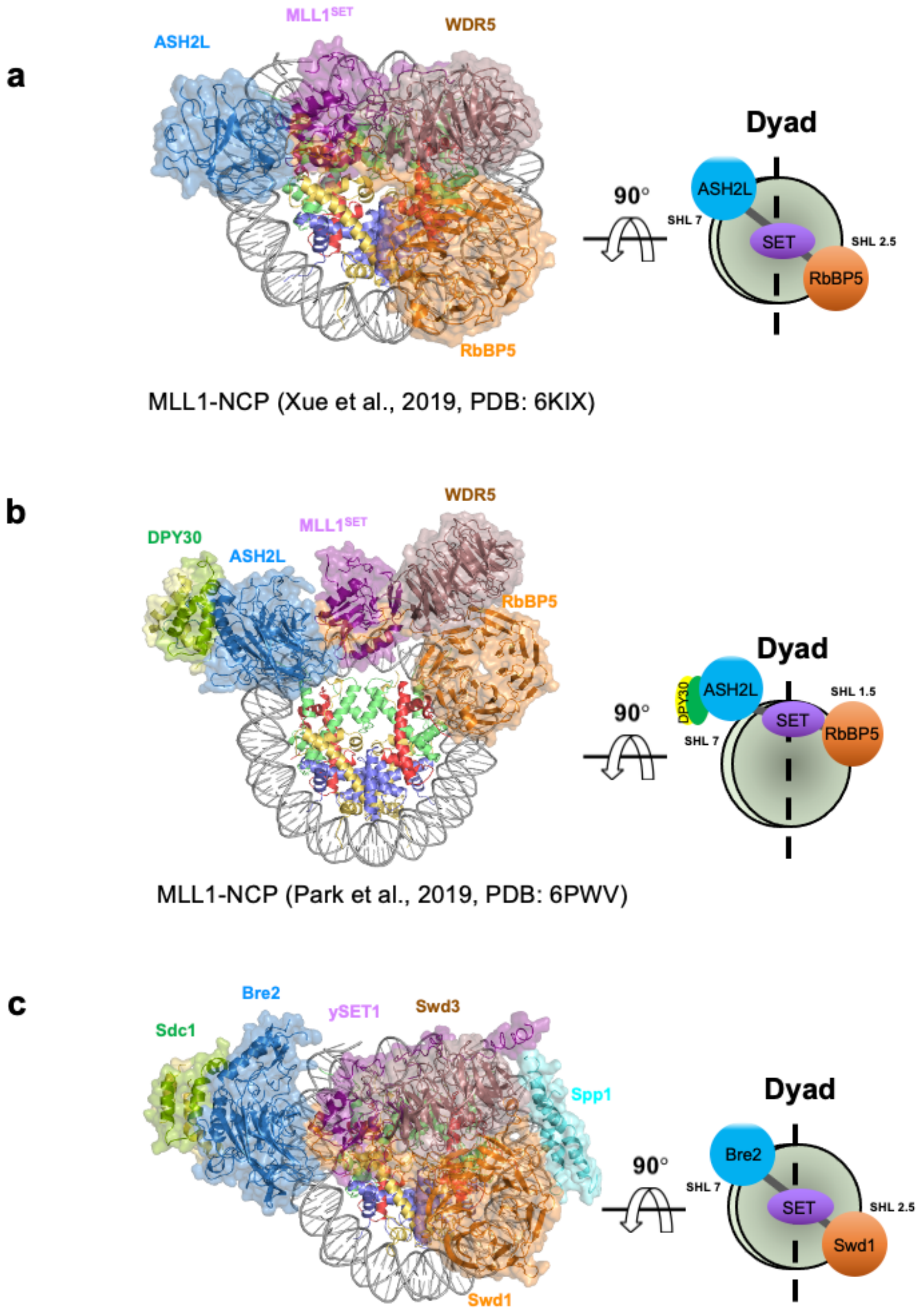


Figure S4.4. Cryo-EM density map. The crystal structure of the MLL1 complex was fitted to show potential interactions at key interfaces. **a,** MLL1-NCP$^{mode\ 2}$ highlighting R251 ($_{248}$LVN$_{250}$ of Loop 1 shown in green), R272, and R294 of RbBP5 (orange) with DNA backbone and I-loop (dashed circle) sandwiched in DNA groove and H4 tail (red); **b,** MLL1-NCP^{mode1} highlighting reorientation of R220, R251, I-loop (dashed circle) and Loop 2 ($_{294}$RGE$_{296}$) of RbBP5 (orange) relative to DNA backbone; **c**, MLL1-NCP$^{mode\ 1}$ highlighting proximity of the MLL1^{SET} (purple) hydrophobic helix, including M3818 and R3821 to H2A (yellow) α2 and α3 helices. Cryo-EM map and crystal structure docking by SHP



MLL1-NCP (Xue et al., 2019, PDB: 6KIX)

MLL1-NCP (Park et al., 2019, PDB: 6PWV)

ySET1-NCP (Hsu et al., 2019, PDB: 6UH5)

Figure S4.5. Unique orientations of MLL1 and ySET1 on NCP**. a**, Top (left) and cartoon (right) views of MLL1^{RWSAD}-NCP structure from Xue et al., 2019 (PDB: 6KIX, EMDB: EMD-0694) [1]. **b**, Top (left) and cartoon (right) views of MLL1^{RWSAD}-NCP structure from Park et al., 2019 (PDB: 6PWV, EMDB: EMD-20512) [2]. **c**, Top (left) and cartoon (right) views of ySET1-NCP structure from Hsu et al., 2019 (PDB: 6UH5, EMDB: EMD- 20767) [3]. **a-c**, The representation of each core complex component labeled and color-coded: DPY30 (Sdc1) homodimer is green/yellow, ASH2L (Bre2) in light blue, MLL1^{SET} (ySET1^{SET}) in purple, WDR5 (Swd3) in tan, RbBP5 (Swd1) in orange and ySET1-specific subunit Spp1 in cyan In **c**, H2BK120ub is colored dark orange. 147 bp 601 DNA is gray and core histones H2A (yellow), H2B (lavender), H3 (green), and H4 (red) are shown. Protein complexes are semi-transparent to reveal minor differences in structural alignment for each complex.

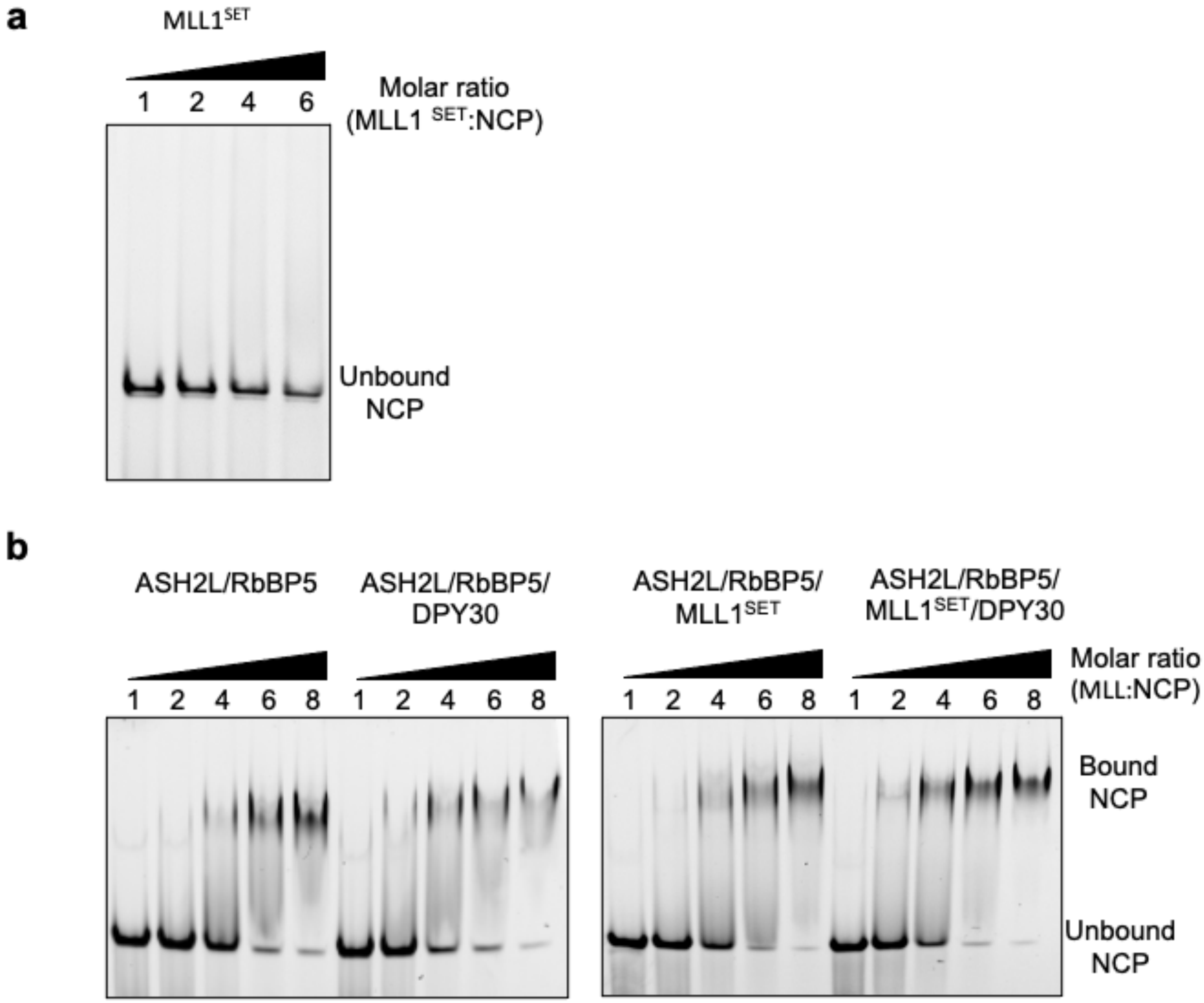


***Figure S4.6*.** MLL^{SET} does not contribute significantly to NCP binding**. a,** Electrophoretic mobility shift assay for MLL1^{SET} on the NCP. **b,** Electrophoretic mobility shift assay comparing relative affinity of various MLL1 sub-complexes as indicated above. For **a** and **b**, values represent increasing molar ratio of proteins on top relative to NCP. EMSA experiments done by YTL

Table 4.1 Cryo-EM Data Collection, Refinement, and Validation Statistics

	MLL1^{RWSAD}-K4M NCP, Mode 1 (EMD-23738, PDB: 7MBM)	MLL1^{RWSAD}-K4M NCP, Mode 2 (EMD-23739, PDB: 7MBN)
Data Collection and Processing		
Magnification	130,000	
Voltage (kV)	300	
Electron exposure (e-/Å^2)	53.4	
Defocus range (μm)	-1.0 to -2.5	
Pixel size (Å)	1.06	
Symmetry imposed	C1	
Initial particle images (no.)	808,836	
Final particle images (no.)	30,322	30,847
Map resolution (Å)	4.76	4.02
FSC threshold	0.143	0.143
Refinement		
Initial model used (PDB code)	6KIX	6PWV
Model resolution (Å)	6.8	6.6
FSC threshold	0.5	0.5
Map sharpening *B* factor (Å^2)	-135.69	-88.62
Model composition		
Non-hydrogen atoms	19,757	20,222
Protein residues	1,747	1,806
Nucleotides	290	292
Ligands	-	-
***B* factor (Å^2)**		
Protein	234.84	31.50
Nucleotide	109.84	9.31
Ligand	-	-
Rmsds		
Bond lengths (Å)	0.006	0.005
Bond angles (°)	0.835	0.669
Validation		
MolProbity score	2.01	2.60
Clashscore	9.39	37.16
Poor rotamers (%)	0	1.50
Ramachandran plot		
Favored (%)	91.26	93.81
Allowed (%)	8.74	6.19
Disallowed (%)	0	0

Appendix C References

1. Xue, H., et al., *Structural basis of nucleosome recognition and modification by MLL methyltransferases.* Nature, 2019. **573**(7774): p. 445-449.
2. Park, S.H., et al., *Cryo-EM structure of the human MLL1 core complex bound to the nucleosome.* Nature Communications, 2019. **10**(1): p. 5540.
3. Hsu, P.L., et al., *Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS.* Molecular Cell, 2019. **76**(5): p. 712-723.e4.