

**The function of the ASH1L histone methyltransferase in cancer:
A chemical biology approach**

by

David Rogawski

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular and Cellular Pathology)
in the University of Michigan
2016

Doctoral committee:

Assistant Professor Jolanta E. Grembecka, Chair
Assistant Professor Tomasz Cierpicki
Associate Professor Yali Dou
Associate Professor Ivan P. Maillard
Associate Professor Ray C. Trievel

When I heard the learn'd astronomer,
When the proofs, the figures, were ranged in columns before me,
When I was shown the charts and diagrams, to add, divide,
and measure them,
When I sitting heard the astronomer where he lectured with
much applause in the lecture-room,
How soon unaccountable I became tired and sick,
Till rising and gliding out I wander'd off by myself,
In the mystical moist night-air, and from time to time,
Look'd up in perfect silence at the stars.

—Walt Whitman
Leaves of Grass

For Maria

Acknowledgements

I would like to begin by thanking my high school and college teachers, who inspired in me a love for science, language, and art. A few teachers deserve special recognition: from Hammond High School (Columbia, MD), Bob Jenkins (World History), Hope Vasholz (English), and Stan Arnold (Calculus); and from Williams College, Tom Garrity (Math), Daniel Lynch (Biology), and the late Bill DeWitt (Biology). My honors thesis mentor Lois Banta at Williams College and my Masters thesis mentor Sigurd Wilbanks at the University of Otago (N.Z.) were essential role models for me as I embarked on a scientific career.

I've been fortunate to work with many great collaborators in the Michigan Pathology Department. Ivan Maillard and Jennifer Chase provided the Δ SET mice, which were originally acquired from Gang Huang, Cincinnati Children's Hospital, and Kenichi Nishioka, Saga University. Lili Chen and Andy Muntean provided the CALM-AF10 mouse bone marrow cells with doxycycline-inducible shRNA targeting Ash11. Bo Zhou and Yali Dou helped with recombinant nucleosome reconstitution.

Next I would like to thank Jola Grembecka and Tomek Cierpicki for creating a highly interdisciplinary lab environment where I could learn chemistry, biochemistry, and biology in one PhD thesis. Their lab is truly one-of-a-kind, and I am very fortunate to have had the opportunity to work there. Juliano Ndoj, who began as my undergraduate assistant, then became lab technician, and is now an important source of expertise in the lab, was instrumental to the success of this project. I thank fellow graduate students George Lund, Felicia Gray, and Jon Pollock, with whom I shared the ups and downs of lab research. I gained much from Hyoje

Cho's expertise in crystallography, Trupta Purohit patiently taught me cell culture, and Hongzhi Miao introduced me to flow cytometry. I was fortunate to work with several highly skilled synthetic chemists, without whom this project would not be possible, including Dmitry Borkin, Yu-Hsuan Kuo, Deanna Montgomery, John Zhuang, and Szymon Klossowski. Patricia Ernst at the University of Colorado Denver was a superb sabbatical mentor during the summer of 2015. Thank you to my thesis committee for their suggestions and advice over the years.

I would like to thank my family and friends who have and continue to support me during my MD/PhD journey. Thanks to my friends in grad school, the MSTP, and the med school, with whom I've shared fun times and exciting adventures. I owe a colossal thanks to my mom Carol, step-dad Jim, dad Michael, and step-mom Julie for their love and support and all the opportunities they have given me. My sister Liz and brother Alex are simply the best siblings ever.

Finally, thank you to my love and partner, Mary. You mean more to me than is appropriate to put on a PhD thesis acknowledgements page. Suffice to say that my PhD years will always feel special for having met you. Thank you for your support and all that you've done for me.

Table of Contents

Dedication	ii
Acknowledgements	iii
List of Figures	viii
List of Tables	ix
Abstract	x
Chapter 1. H3K36 methyltransferases as cancer drug targets: rationale and perspectives for inhibitor development	1
Abstract	1
Introduction	2
H3K36 methylation regulates diverse processes implicated in cancer	4
Transcriptional regulation	4
Alternative splicing	5
DNA repair	6
The reciprocal relationship between H3K36 and H3K27 methylation	7
H3K36 KMTases play important and varying roles in carcinogenesis	8
NSD1: Activator of <i>HOX</i> genes in leukemia	10
NSD2 (MMSET/WHSC1): An oncogenic driver in multiple myeloma	11
NSD3: Frequently overexpressed in diverse tumor types	13
ASH1L: <i>HOX</i> gene activator with emerging role in cancer	14
SETD2: An important tumor suppressor	15
Other H3K36 methyltransferases: SMYD2, SETMAR, and SETD3	16
H3K36-specific SET domains: structural considerations for inhibitor development	17
Structural studies of NSD, ASH1L, and SETD2 SET domains demonstrate feasibility of inhibitor design	18
SMYD2 structures: U-shaped substrate binding	21
Inhibitors of H3K36 methyltransferases	22
Challenges and opportunities in developing specific inhibitors of H3K36 KMTases	24
Targeting the catalytic SET domain	24
Alternative druggable regions of H3K36 KMTase proteins	26
Conclusions and future perspectives	29
Chapter 2. Two loops undergoing concerted dynamics regulate activity of the ASH1L histone methyltransferase	31
Abstract	31

Introduction.....	32
Results.....	34
Crystal structure of the ASH1L SET domain shows increased dynamics of the autoinhibitory and SET-I loops.....	34
NMR studies reveal that the ASH1L active site is surrounded by two loops experiencing conformational dynamics	36
ASH1L requires chromatin-interacting domains for robust enzymatic activity	38
ASH1L SET-BAH requires native nucleosome substrate for optimal enzymatic activity.....	41
Mutations of non-conserved residues in autoinhibitory loop have significant effects on ASH1L SET domain activity	41
NMR studies correlate the degree of structural perturbation to enzyme activity	43
Mutations affect structure and mobility of the ASH1L autoinhibitory loop	46
Structural insight into SET-I loop conformational dynamics	49
Conformational exchange in the SET-I loop contributes to ASH1L activity.....	50
Discussion.....	53
Chapter 3. Development of small molecule inhibitors for the ASH1L histone methyltransferase	56
Abstract.....	56
Introduction.....	57
Results.....	58
FBDD identifies a novel ligand of ASH1L SET domain	58
Medicinal chemistry optimization of fragment screening hit 1	60
Crystal structure of 6 in complex with ASH1L SET shows details of protein-ligand interaction and stabilization of autoinhibitory loop region.....	65
Structure-based design of SK-1	69
Optimizing substituents at position 6 of indole and combining with substituents at position 1.....	73
Sulfur-oxygen chalcogen bonding in ASH1L-inhibitor complexes	74
Discussion.....	76
Chapter 4. Genetic and pharmacologic approaches identify novel functions of ASH1L in breast cancer and leukemia.....	78
Abstract.....	78
Introduction.....	79
Results.....	80
ASH1L regulates proliferation of breast cancer cells	80
ASH1L regulates HOX gene expression in human leukemia cells	81
ASH1L regulates proliferation of CALM-AF10-transformed mouse bone marrow cells	83
The ASH1L SET domain is required for leukemic transformation.....	84
SK-1 inhibits growth of MA9 but not HM2 mouse bone marrow cells	87
SK-1 induces differentiation of MA9 cells.....	90
SK-1 blocks proliferation of MLL-rearranged human leukemic cell lines.....	91
Discussion.....	92

Chapter 5. Conclusions and perspectives	95
The first inhibitors of ASH1L KMTase activity represent a new paradigm in the KMTase inhibitor field	96
Advantages and disadvantages of the FBDD approach	98
Improving potency, physicochemical, and pharmacokinetic properties of ASH1L inhibitors	100
The role of ASH1L in leukemia.....	101
Future directions	104
Chapter 6. Materials and methods	106
Appendix. Selected ASH1L inhibitors synthesized to date	113
References	131

List of Figures

Figure 1.1. The functions of H3K36 methylation in biology.....	8
Figure 1.2. Domain organization of human H3K36 methyltransferases	9
Figure 1.3. Structures and structural probes of H3K36 KMTases.....	20
Figure 1.4. Sequence alignment of core SET and post-SET regions of the related NSD, ASH1L, and SETD2 KMTases.....	21
Figure 1.5. Peptide substrate-competitive inhibitors of the SMYD2 KMTase.....	24
Figure 1.6. Targeting PPI domains of H3K36 KMTases as an alternative to the SET domain....	28
Figure 2.1. Analysis of ASH1L SET domain crystal structure.....	36
Figure 2.2. NMR studies of the ASH1L SET domain	38
Figure 2.3. Enzymatic activity of ASH1L variants.....	40
Figure 2.4. Effects of autoinhibitory loop mutations in ASH1L	45
Figure 2.5. Conformational dynamics observed in the crystal structure of Q2265A	50
Figure 2.6. Structural perturbation in SET-I loop induced by the H2193F mutation.....	52
Figure 3.1. Identification of a fragment-like ligand of the ASH1L SET domain	60
Figure 3.2. Biochemical and structural characterization of ASH1L inhibitors	64
Figure 3.3. Compound 6 stabilizes autoinhibited conformation of ASH1L SET domain.....	68
Figure 3.4. Development of SK-1	72
Figure 3.5. Combination of optimized substituents at 1 and 6 positions	74
Figure 3.6. Chalcogen bonds induced by binding of SK-1 to ASH1L SET	75
Figure 4.1. Knockdown of ASH1L slows proliferation of MDA-MB-468 cells.....	81
Figure 4.2. Relative expression of <i>ASH1L</i> and <i>HOX</i> genes in THP1 and KG1 cells	82
Figure 4.3. Knockdown of <i>Ash1l</i> inhibits proliferation of CALM-AF10 BMCs.....	84
Figure 4.4. Colony formation by Δ SET and wild-type mouse BMCs.....	86
Figure 4.5. Effect of ASH1L inhibitors on cell viability of MA9 and HM2 cells.....	89
Figure 4.6. SK-1 induces differentiation of MA9 cells.....	91
Figure 4.7. Growth inhibition by ASH1L inhibitor SK-1 and negative control compound SK-nc in human leukemia cell lines	92
Figure 5.1. Regions for optimization of ASH1L inhibitors	101
Figure 5.2. Three models of the relationship between ASH1L and MLL in activating target genes	104

List of Tables

Table 2.1. Crystallographic data collection and refinement statistics for ASH1L WT and mutants	48
Table 3.1. Medicinal chemistry optimization of fragment screening hit 1	62
Table 3.2. Crystallographic data collection and refinement statistics for crystals of ASH1L Q2265A in complex with inhibitors	67
Table 3.3. Medicinal chemistry optimization of indole scaffold	71
Table A.1. Selected ASH1L inhibitors synthesized to date	113

Abstract

ASH1L (absent, small, or homeotic-like 1) is a histone lysine methyltransferase (KMTase) that is overexpressed in cancer and activates oncogenic *HOX* genes. Small molecule inhibitors of ASH1L would be invaluable tools to investigate the role of ASH1L in cancer; however, no ASH1L inhibitors have been reported to date. Previous studies of ASH1L's catalytic SET domain identified an autoinhibitory loop that blocks access of histone substrate to the enzyme active site. We used nuclear magnetic resonance (NMR) and X-ray crystallography to identify conformational heterogeneity in the ASH1L autoinhibitory loop and nearby SET-I loop, two structural features that regulate the enzymatic activity of the SET domain. These studies suggested that conformational heterogeneity in the autoinhibitory loop region of the ASH1L SET domain may create transient pockets into which small molecule ligands could bind.

We took a fragment-based drug discovery (FBDD) approach to probe the ASH1L SET domain for potential small molecule binding sites and then construct new ligands specific to ASH1L. FBDD identified a ligand that binds to the ASH1L autoinhibitory loop region. We used information from NMR and crystallographic studies to optimize the fragment-like ligand into a first-in-class potent and specific ASH1L inhibitor. Structural studies indicate that ASH1L inhibitors block enzymatic activity by stabilizing the autoinhibited conformation of the SET domain. Our current most potent compounds inhibit ASH1L activity with IC_{50} of $\sim 4 \mu M$ and bind to ASH1L with $\sim 1 \mu M$ affinity, representing greater than 1000-fold improvement over the fragment screening hit. Our work identified the first ASH1L inhibitor and represents the first example of successfully applying FBDD to KMTases.

Using our ASH1L inhibitors, we took a combined genetic and pharmacologic approach to investigate the role of ASH1L in leukemia and breast cancer. We found that ASH1L is required for proliferation of breast cancer and leukemia cells, and that ASH1L activates *HOXA* genes and *MEIS1* in leukemia. Moreover, inhibition of the ASH1L SET domain downregulated *HOX* gene expression and induced differentiation of leukemia cells transformed by MLL-AF9. Our results demonstrate cellular efficacy of ASH1L inhibitors and uncover a new role for the ASH1L SET domain in acute leukemia.

Chapter 1. H3K36 methyltransferases as cancer drug targets: rationale and perspectives for inhibitor development*

Abstract

Methylation at histone 3, lysine 36 (H3K36) is a conserved epigenetic mark regulating gene transcription, alternative splicing, and DNA repair. Genes encoding H3K36 methyltransferases (KMTases) are commonly overexpressed, mutated, or involved in chromosomal translocations in cancer. Molecular biology studies have demonstrated that H3K36 KMTases regulate oncogenic transcriptional programs. Structural studies of the catalytic SET domain of H3K36 KMTases have revealed intriguing opportunities for design of small molecule inhibitors. Nevertheless, potent inhibitors for most H3K36 KMTases have not yet been developed, underlining the challenges associated with this target class. As we now have strong evidence linking H3K36 KMTases to cancer, drug development efforts are predicted to yield novel inhibitors in the near future.

*Rogawski DS, Grembecka J, and Cierpicki T. H3K36 methyltransferases as cancer drug targets: rationale and perspectives for inhibitor development. *Future Med Chem.* Under review.

Introduction

Cellular development and differentiation are controlled by post-translational modifications on DNA and histone proteins, forming an epigenetic (“above the genes”) regulatory network. One major class of epigenetic modifications is histone lysine methylation, which alters the chromatin environment and enhances or suppresses binding of regulatory cofactors [1,2]. Methyl marks are installed by histone lysine methyltransferases (KMTases), removed by demethylases, and bound by a variety of methyl-reader proteins that initiate downstream effects. Methylation occurs at multiple different lysine residues on histone proteins, the most extensively studied of which include histone 3 lysine 4 (H3K4), H3K9, H3K27, H3K36, H3K79, and H4K20, with methylation at different sites signaling the recruitment of different cofactors [1,3]. The degree of methylation (mono-, di-, or trimethylation) at a particular site is an additional variable that affects cofactor recruitment. For example, H3K4 trimethylation is associated with the transcription start sites of active genes, where one of its multiple functions is to recruit the transcription factor TFIID complex, an essential component of the RNA polymerase II pre-initiation complex [4]. In contrast, H3K27 trimethylation is found in regions of repressed chromatin, where it plays a role in promoting chromatin compaction [5].

The human genome encodes approximately 50 KMTase proteins. With the exception of DOT1L, these KMTases contain a catalytic Suppressor of variegation 3–9 [Su(var)3–9], Enhancer of zeste [E(z)], and Trithorax (Trx) (SET) domain responsible for KMTase activity [2]. The core SET domain is approximately 130 amino acids in length and catalyzes transfer of a methyl group from S-adenosyl methionine (SAM) to the ϵ -amino group of a histone lysine. Interestingly, an increasing number of SET domain-containing proteins have been found to also methylate non-histone proteins [2]. Most SET domain-containing proteins contain additional

domains that recruit the SET domain to its target or have SET domain-independent functions [6,7]. Furthermore, some SET domain-containing proteins function in protein complexes that regulate their catalytic activity [2].

In cancer, epigenetic pathways are commonly disrupted, causing aberrations in gene expression and genome integrity (reviewed in [8]). In particular, KMTases play major roles in driving proliferation and halting differentiation by modifying transcription and other DNA-templated processes [9–11]. Over the past decade, KMTases have emerged as important drug targets for both industrial and academic research groups. Some progress has been made, most notably by selective inhibitors for the EZH2 and DOT1L KMTases that have reached clinical trials for the treatment of non-Hodgkin lymphoma and acute leukemia, respectively [12,13]. The KMTase inhibitor field, however, has yet to experience the widespread success of kinase inhibitors [14], despite similarities in the reactions catalyzed by these two classes of enzymes. Indeed, selective inhibitors for many KMTases remain unavailable.

Methylation at H3K36 represents a particularly important chromatin mark implicated in diverse forms of cancer. KMTases with specificity toward H3K36 are overexpressed in cancer cells and have been characterized as regulators of cell growth, differentiation, stemness, and DNA repair pathways [15–17]. However, very few selective and cell-active small molecule inhibitors of H3K36-specific KMTases have been reported to date. In this chapter, we provide an overview of cellular pathways involving H3K36 methylation and discuss the diverse functions carried out by the eight different human H3K36-specific KMTases. Then we analyze structural characteristics of the catalytic SET domain of H3K36 KMTases and evaluate prospects for their inhibition by small molecules. We conclude with a discussion of the challenges and opportunities for targeting these proteins.

H3K36 methylation regulates diverse processes implicated in cancer

H3K36 methylation participates in a wide variety of nuclear pathways. Many of these processes, including transcriptional regulation, alternative splicing, and DNA repair, are dysregulated in cancer.

Transcriptional regulation

On a genome scale, H3K36 methylation is enriched within the transcribed region of active genes rather than regulatory sites [18,19], with the highest levels at the 3' end of genes [20]. H3K36 methylation plays a key role within gene bodies by preventing cryptic intragenic transcription, which may generate toxic proteins or waste cell resources [21]. In this way, H3K36 methylation is associated with active genes but actually has a repressive effect on chromatin structure within these genes. In yeast, the sole H3K36 methyltransferase is Set2, which is recruited to actively transcribed genes by binding to the hyperphosphorylated C-Terminal Domain of RNA Polymerase II (Pol II) [22]. Methylation of H3K36 by Set2 recruits the Rpd3S histone deacetylase (**Figure 1.1A**), which produces chromatin compaction and prevents aberrant transcription from cryptic intragenic promoters in the wake of Pol II [23]. H3K36 methylation further suppresses cryptic transcription by preventing the incorporation of new, acetylated histones over open reading frames (ORFs). This role depends on methylated H3K36 recruiting chromatin remodelers Isw1b and Chd1 and preventing H3 binding to the histone chaperone Asf1 [24,25].

H3K36 methylation similarly prevents cryptic intragenic transcription in mammals. The human H3K36 tri-methyltransferase and Set2 homolog SETD2 is required for preventing intragenic transcription initiation in 11% of genes [26]. This function does not depend on histone deacetylases as in yeast, but on recruitment of histone demethylases and chaperones. In

embryonic stem cells, the H3K36me3 reader protein MRG15 recruits the JARID1B histone demethylase to intragenic regions, where it removes H3K4me3 to prevent cryptic intragenic transcription and facilitate transcriptional elongation of functional full-length mRNAs [27] (**Figure 1.1A**). Interestingly, human LSD2, a second H3K4 demethylase that maintains a repressive chromatin environment in gene bodies to facilitate optimal transcriptional elongation, forms a complex with the H3K36 KMTase NSD3 [28]. Additionally, H3K36me3-modified chromatin recruits the FACilitates Chromatin Transcription (FACT) histone chaperone complex, which helps restore chromatin structure following transcriptional elongation [26].

Alternative splicing

Splicing pathways are deregulated in cancer, resulting in expression of gene isoforms that contribute to oncogenic processes [29]. H3K36me3-modified nucleosomes are enriched within exons [30], and accumulating evidence indicates that H3K36 methylation signals to the splicing machinery. As an example, the human fibroblast growth factor receptor 2 (*FGFR2*) gene undergoes tissue-specific alternative splicing that is controlled by pyrimidine tract binding protein (PTB). Knockdown or overexpression of the H3K36 tri-methyltransferase SETD2 causes inclusion or exclusion, respectively, of the PTB-regulated exon in *FGFR2*. Further experiments showed that PTB is specifically recruited to exons labeled with H3K36me3 by interacting with MRG15, an H3K36 methyl-reader protein that is present in multiple protein complexes [31] (**Figure 1.1B**). In addition to the H3K36me3/MRG15/PTB alternative splicing pathway, multiple PWWP domain-containing proteins that recognize H3K36me3 have been implicated in alternative splicing [32,33]. For example, the PWWP domain-containing protein ZMYND11 localizes primarily to gene bodies due to specific recognition of K36me3 on histone variant H3.3. ZMYND11 promotes intron retention at hundreds of sites by interacting with and

antagonizing the function of the splicing regulator EFTUD2 [33]. On the other hand, alternative splicing has been shown to regulate H3K36 methylation. Deletion of splicing sites or pharmacological inhibition of splicing leads to redistribution of H3K36 tri-methylation, reduces overall levels of H3K36 tri-methylation, and blocks SETD2 recruitment to chromatin [34,35]. These findings suggest regulatory cross-talk between H3K36 methylation and alternative splicing pathways.

DNA repair

Genome instability is a key characteristic of cancer, and H3K36 methylation has a conserved role in DNA repair. In yeast, H3K36 methylation mediated by Set2 promotes the Non-Homologous End Joining (NHEJ) pathway of DNA repair during the G1 phase of the cell cycle, whereas H3K36 acetylation mediated by Gcn5 promotes Homologous Recombination (HR) repair during the S and G2 phases [36,37]. The mammalian Set2 homolog SETD2 also regulates DNA repair, but in the opposite direction from Set2, as SETD2 promotes HR rather than NHEJ. Histones tri-methylated at H3K36 by SETD2 provide a constitutive binding site for the LEDGF PWWP domain. Once a DNA double-strand break occurs, LEDGF recruits CtIP, which performs 5' end resection to produce a single stranded 3' overhang (**Figure 1.1C**). Recruitment of major HR repair factors downstream of CtIP, such as RPA and RAD51, also depends on SETD2-mediated H3K36 tri-methylation [38]. In cells with depleted SETD2, double-strand break repair proceeds via the error-prone microhomology-mediated end-joining pathway, leading to deletions between sequences of microhomology [38]. SETD2 is also responsible for regulating DNA mismatch repair through its interaction with MutS α [39]. These studies provide mechanistic basis for the genome instability observed in SETD2-deficient cancers (see below).

Another H3K36 KMTase involved in DNA repair is SETMAR, a gene that evolved from fusion of a SET domain to a transposase domain. Upon radiation damage, H3K36me₂ levels are increased in a SETMAR-dependent manner. Depleting SETMAR or expression of H3 with mutated K36 decreases Ku70 and NBS1 recruitment, thereby decreasing efficiency of the NHEJ repair pathway [40] (**Figure 1.1C**). The SET domain of SETMAR is also required for the DNA damage response at stalled replication forks and efficient restarting of stalled replication forks [41]. Interestingly, SETMAR does not methylate recombinant nucleosomes *in vitro* [42], suggesting that SETMAR may indirectly produce H3K36me₂ after DNA damage.

The reciprocal relationship between H3K36 and H3K27 methylation

How different epigenetic marks interact with each other to produce signaling outputs has important implications for epigenetic inhibitor development. H3K36 methylation interacts in an antagonistic fashion with H3K27 tri-methylation, a repressive mark mediated by Polycomb Repressive Complex 2 (PRC2). For example, in *Drosophila* the H3K36-specific KMTase Ash1 activates *Hox* genes during development by functioning as an anti-repressor and antagonizing repressive H3K27 methylation installed by PRC2 [43]. In human HeLa cells, H3K36 methylation and H3K27 methylation are rarely found together on histone H3. In fact, H3K36 pre-methylation inhibits PRC2 H3K27-KMTase activity in *in vitro* assays [44]. Conversely, H3K36 KMTases are inhibited by ubiquitinated H2A, a mark made by Polycomb Repressive Complex 1 [45]. As development proceeds, however, the PRC2 complex must invade active, H3K36-methylated chromatin to silence certain genes. In this case, PRC2 targeting and spreading is mediated by Polycomb-like proteins with Tudor domains that specifically recognize H3K36me₃ [46]. Disruption of the balance between the H3K36 and H3K27 methylation pattern is observed in multiple cancers (discussed below) (**Figure 1.1D**). In particular, cancers with

chromosomal fusions involving H3K36 KMTases have disruptions in H3K27 methylation that drive oncogenesis, in addition to aberrant H3K36 methylation [16,47].

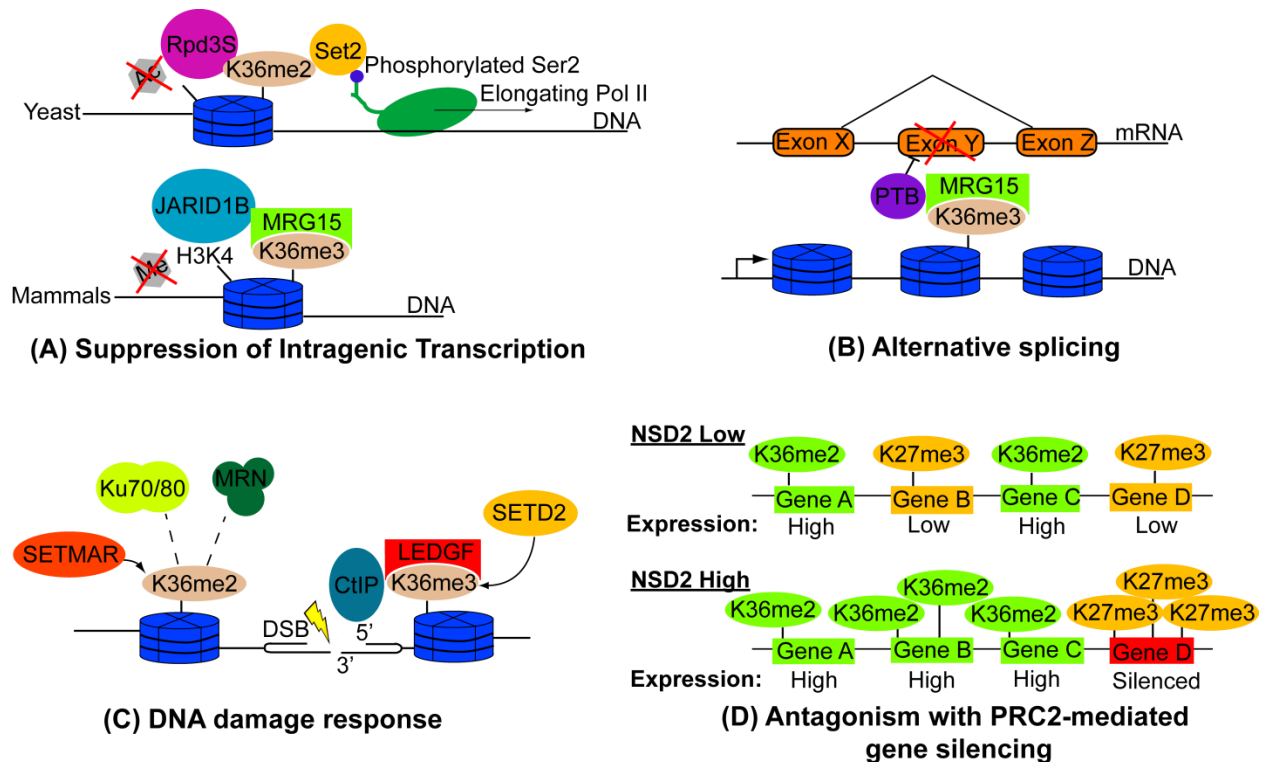


Figure 1.1. The functions of H3K36 methylation in biology. **(A)** H3K36 methylation plays a conserved role in suppressing intragenic transcription. Methylated H3K36 recruits histone deacetylases (yeast) or histone demethylases (humans) to maintain a suppressive transcriptional environment within gene bodies. **(B)** Splicing regulators including PTB are recruited to loci by H3K36 methylation. **(C)** Methyl marks installed by the SETMAR and SETD2 KMTases are important for recruiting mediators of the DNA damage response including Ku70/Ku80, MRE11-RAD50-NBS1 (MRN), and CtIP. **(D)** Methylated H3K36 antagonizes PRC2-mediated H3K27 methylation and gene silencing. In NSD2 Low cells, active genes are marked by H3K36 methylation, whereas lower expressed genes are marked by H3K27 methylation. In NSD2 High cells (as in t(4;14) multiple myeloma), overexpression of NSD2 increases genome-wide levels of H3K36 methylation, forcing accumulation of H3K27 methylation at a subset of silenced loci. Adapted from [47].

H3K36 KMTases play important and varying roles in carcinogenesis

Given the importance of H3K36 methylation in diverse cellular processes, it is not surprising that H3K36 KMTases have been implicated both as oncogenes and tumor suppressors in cancer. The human genome encodes at least eight H3K36 KMTases (**Figure 1.2**), each of which contains a SET domain responsible for transferring a methyl group from SAM to the

substrate lysine. Although these proteins have all been reported to methylate H3K36, they vary in the number of methyl groups transferred as well as whether they methylate additional substrates. NSD1-3, ASH1L, and SETD2 have closely related catalytic SET domains and show H3K36 methylation specificity *in vitro* and *in vivo*, while SMYD2, SETMAR, and SETD3 have more distantly related SET domains with less well characterized activities toward H3K36 (Figure 1.2). Of these KMTases, the NSD proteins are perhaps the best characterized oncoproteins, as they play important roles in multiple cancer types and are found in fusion proteins due to chromosomal translocations in acute myeloid leukemia (AML) and multiple myeloma (MM) [10,16]. The ASH1L KMTase is also overexpressed in cancer and regulates stem cell potential [15]. SMYD2 has oncogenic activity in multiple cancers [48–52], but whether this function depends on KMTase activity toward H3K36 or another of its many substrates is uncertain.

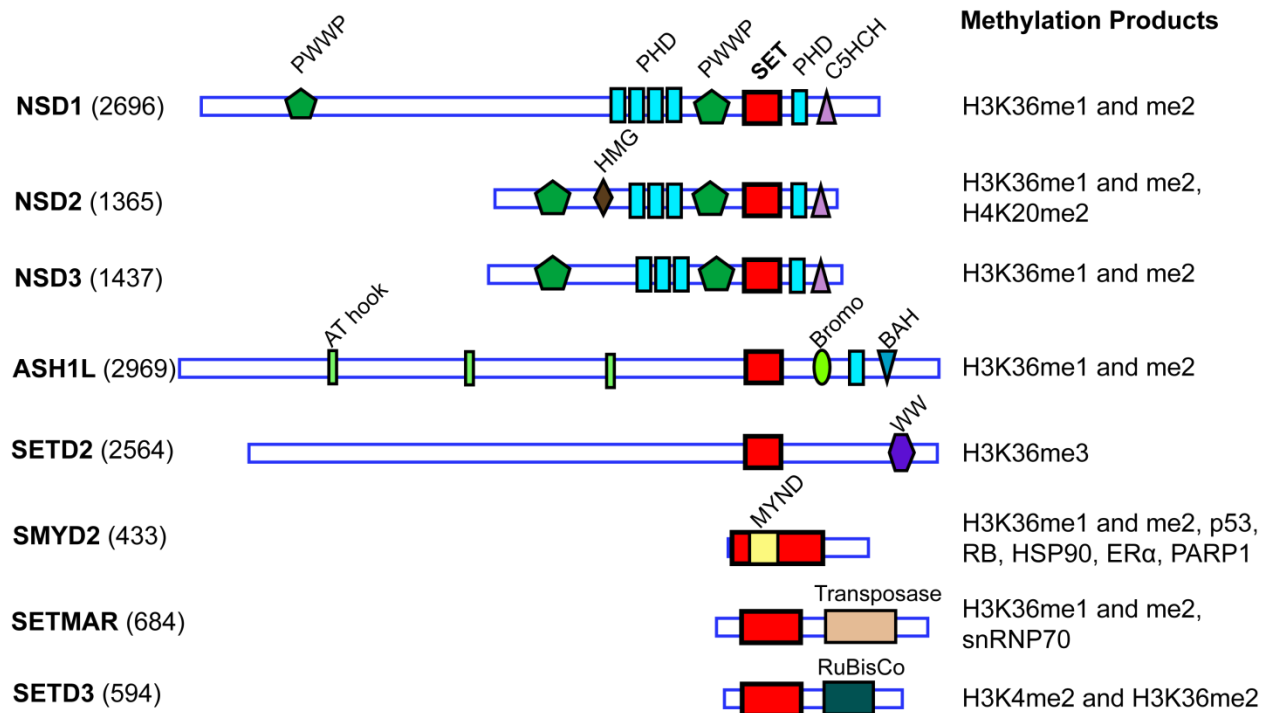


Figure 1.2. Domain organization of human H3K36 KMTases. SET domain is shown in red.

NSD1: Activator of HOX genes in leukemia

NSD1 is a mono- and di- H3K36 KMTase with functions in development and cancer. Mutations in NSD1 cause Sotos syndrome, a condition of childhood overgrowth and intellectual disability, with a 2.4% increased risk of childhood malignancy [53,54]. A chromosomal translocation resulting in the NUP98-NSD1 fusion protein is found in 16% of cytogenetically normal pediatric AML and in a smaller portion of adult AML [55]. More than 90% of NUP98-NSD1-positive leukemias are also positive for internal tandem duplication mutation of the FLT3 tyrosine kinase (FLT3-ITD), and the two genetic lesions exhibit potent cooperativity resulting in a 3-year survival rate of 31% [56]. NUP98-NSD1 induces AML *in vivo*, sustains self-renewal of myeloid stem cells *in vitro*, and enforces expression of the *HOXA7*, *HOXA9*, *HOXA10* and *MEIS1* proto-oncogenes [16]. Expression of the *HOXA* and *MEIS1* oncogenes appears to be responsible for the transforming activity of NUP98-NSD1, as inhibition of the DOT1L-AF10 complex in NUP98-NSD1 leukemia decreases *HOXA* gene expression and triggers differentiation and apoptosis [57].

NSD1 has also been reported to methylate non-histone proteins, including the p65 subunit of NF- κ B at Lys218 and Lys221. In response to cytokines such as IL-1 β and TNF α , NSD1-mediated methylation enhances NF- κ B's transcriptional activation and DNA binding activities [58], which are active in most cancer cells and regulate genes that control proliferation, resistance to apoptosis, angiogenesis, invasion, and metastasis [59]. Conversely, these activating marks on NF- κ B are removed by the FBXL11 demethylase, and increasing methylation at Lys218 and Lys221 by depleting FBXL11 enhances cell proliferation and colony formation of colon cancer cells [58]. In addition, mutation of Lys218 and Lys221 on NF- κ B showed that lysine methylation is required for activating the majority of NF- κ B target genes in mouse

embryonic fibroblasts (MEFs), including cancer-relevant genes like insulin-like growth factor 1 (*IGF1*) and engulfment and cell motility 1 (*ELMO1*) [58].

In other contexts, however, NSD1 has been reported to function as a tumor suppressor. In bladder cancer cells, for example, inhibition of NF- κ B by small molecule causes NSD1-dependent activation of pro-apoptotic BIM and the cell cycle regulator p21 [60]. Although it is not known if NSD1 KMTase activity is involved here, NF- κ B inhibition does increase global levels of H3K36 tri-methylation in addition to NSD1 expression [60]. Moreover, NSD1 frequently undergoes inactivating mutations causing premature termination of the protein in head and neck squamous cell carcinoma [61], and NSD1 is epigenetically silenced in neuroblastoma and glioma [62]. Finally, missense mutations in NSD1 are found in acute myeloid leukemia outside of the context of the NUP98-NSD1 fusion [63]. Although the functional consequences of these mutations are not known, these observations suggest variable roles for NSD1 in different cancer types.

NSD2 (MMSET/WHSC1): An oncogenic driver in multiple myeloma

NSD2 (also known as MMSET and WHSC1) is an important developmental regulator and oncogene. Germline deletion of NSD2 causes Wolf-Hirschhorn syndrome, a developmental disorder characterized by craniofacial defects, growth retardation, and microcephaly [64]. NSD2 has *in vivo* mono- and di- methyltransferase activity toward H3K36 [10]. Interestingly, NSD2 has also been reported to dimethylate H4K20, which was proposed to signal the recruitment of the DNA damage response regulator 53BP1 to sites of DNA damage [17,65], but other groups have not found evidence supporting this model [66,67].

NSD2 is a potentially attractive target for drug development due to its well characterized role in multiple myeloma (MM). In 15% of MM, the t(4;14)(p16.3;q32.3) translocation places

the NSD2 gene under control of the immunoglobulin heavy-chain promoter/enhancer, leading to overexpression of NSD2, which is believed to be the key transforming factor [68]. In human myeloma cells harboring the t(4;14) translocation, overexpressed NSD2 leads to aberrant patterns of H3K36 methylation, causing a shift away from normal plasma cell gene expression programs and increased expression of cancer-associated genes [10]. Expression of NSD2 is sufficient to rescue tumor growth of myeloma cells in which the t(4;14) translocation has been inactivated [10], and knockdown of NSD2 in a t(4;14) mouse xenograft model causes dramatic reduction in tumor growth [47]. Importantly, the KMTase activity of the NSD2 SET domain is required for the epigenetic changes and oncogenic effects caused by NSD2 overexpression [10,47].

Beyond multiple myeloma, expression of NSD2 is sufficient to transform MEF cells lacking the p19^{ARF} tumor suppressor, suggestive of it having a role as a general oncoprotein [10]. Indeed, NSD2 is overexpressed in neuroblastoma, lung, colon, and bladder cancer [69,70], and is required for proliferation of neuroblastoma cells and brain-derived neural stem cells [70]. An E1099K mutation in the SET domain of NSD2 was identified in a range of human cancer cell lines and tumor samples, including ALL, chronic lymphocytic leukemia (CLL), lung cancer, and stomach cancer [71,72]. The mutation appears to result in a hyperactive enzyme, as H3K36me2 levels are increased in cell lines harboring E1099K NSD2 [71,72]. Thus, the t(4;14) translocation in MM and the E1099K mutation in other cancers both result in increased H3K36me2 levels that drive oncogenesis.

NSD2's oncogenic role may partially reflect an imbalance between H3K36 and H3K27 methylation pathways. For example, in t(4;14) MM cells, a global loss of H3K27me3 is observed concurrently with a global increase in H3K36me2 [47], as one might predict due to the

antagonistic relationship between these two marks. However, the excess H3K36me2 forces local accumulation of H3K27me3 at a subset of genes, including tumor suppressors. MM cells with t(4;14) translocation are sensitive to EZH2 inhibitors, indicating that the accumulation of EZH2 and H3K27me3 at tumor suppressor loci enhances myeloma cell proliferation [47,73].

Interestingly, in other cancers besides MM, EZH2 and NSD2 are co-regulated in an EZH2/NSD2 oncogenic axis [74]. EZH2 represses expression of a set of microRNAs that repress NSD2, leading to EZH2/NSD2 co-regulation. In prostate cancer for example, the EZH2/NSD2 axis confers increased cell proliferation, migration, invasion, stem cell-like properties, and tumor growth in a mouse xenograft model [74]. Further studies in prostate cancer have shown that NSD2 acts as a co-activator with NF- κ B and androgen receptor [75,76], and regulates the epithelial to mesenchymal transition by dimethylating H3K36 at the TWIST1 locus and activating TWIST1 expression [77].

NSD3: Frequently overexpressed in diverse tumor types

NSD3, also known as WHSC1L1, is overexpressed in cancer [78] and catalyzes mono- and dimethylation at H3K36 *in vitro* [79]. In cells NSD3 functions as a transcriptional activator, but NSD3's KMTase activity may contribute only partially to its gene-activating role. For example, NSD3 regulates neural crest specification and migration and is required for activation of neural crest transcription factors *Sox10*, *Snail2*, *Sox9*, and *FoxD3*, but the only locus at which NSD3 is responsible for H3K36 dimethylation is *Sox10* [80]. Similar to NSD1, NUP98-NSD3 fusion proteins have been reported in AML and myelodysplastic syndrome (MDS) [81,82].

NSD3 undergoes copy number amplification in 21% of lung squamous cell carcinomas and 15% of breast invasive carcinomas, with a high correlation between copy number and mRNA expression, suggesting that NSD3 may function as an oncogenic driver [78]. Indeed, RNAi

knockdown of NSD3 in non-small cell lung cancer, colorectal cancer, bladder cancer, and breast cancer cell lines with NSD3 overexpression causes reduced cell proliferation due to increased apoptosis or cell cycle arrest in the various cancer cell lines [78,83,84]. The breast epithelial cell line MCF10A is normally highly dependent on growth factors and forms small acinar-like structures in 3D Matrigel culture, but these cells can be transformed by expression of NSD3, which confers growth factor-independent proliferation, colony-forming ability in soft agar, and expanded and disorganized growth in 3D culture [84]. In the transformed cells, NSD3 overexpression causes activation of *IRX3*, a homeobox gene family member, and downregulation of *TGFBI*, a putative tumor suppressor [84]. These findings, combined with frequent overexpression of NSD3 in breast cancer, make NSD3 an enticing therapeutic target.

ASH1L: HOX gene activator with emerging role in cancer

Multiple studies have found that ASH1L is an important transcriptional regulator in normal development, but the protein's role in cancer remains unclear. ASH1L is homologous to *Drosophila* Ash1, a trithorax group protein that activates genes involved in development and differentiation [85]. In mammals, ASH1L deficiency causes a major reduction in long-term hematopoietic stem cells (HSC) in bone marrow, but surprisingly has very modest effects on peripheral blood counts due to increased proliferation of progenitors downstream of HSCs [15]. ASH1L-deficient HSCs are also unable to reconstitute bone marrow output when transplanted into lethally irradiated mice [15]. These findings indicate that ASH1L maintains quiescence and self-renewal potential of long-term HSCs, but whether ASH1L regulates the stemness properties of leukemic stem cells is unknown.

ASH1L has been linked to liver fibrosis, facioscapulohumeral muscular dystrophy, and cancer through its role in gene activation [86,87]. ASH1L activates *HOXA* genes and *MEIS1* in

mouse HSCs [15] and activates *HOXB* and *HOXC* genes in human erythroleukemic K562 cells [88]. These findings are highly relevant because *HOX* genes are oncogenic drivers in many different blood and solid tumors [89]. For example, overexpression of *HOXA9* is highly associated with a poor prognosis in AML [90], and *HOXA9* and its collaborator *MEIS1* are required for survival of MLL-rearranged leukemia cells [91,92]. ASH1L's KMTase activity is required for at least some of its gene activating function, as deletion of the ASH1L SET domain in differentiating mouse embryonic stem cells leads to loss of expression of 152 genes, including members of the *Hox* and *Wnt* families [93].

ASH1L is overexpressed in a variety of solid tumors, including thyroid and breast cancer [94,95]. In thyroid cancer, ASH1L is overexpressed in tumor-specific truncated forms, although what parts of the protein are truncated or what processing events produce them has not been explored. Expression of the microRNA miR-142-3p suppresses ASH1L protein expression by binding to the ASH1L 3'UTR, an effect correlated with inhibition of colony formation and slowing of thyroid cancer cell growth [94]. Interestingly, in mouse embryonic stem cells ASH1L is regulated by a different microRNA, miR-290, which downregulates ASH1L expression and prevents aberrant overexpression of *Hoxa*, *Hoxb*, *Hoxc*, and *Hoxd* genes [96].

SETD2: An important tumor suppressor

SETD2 stands out among the human H3K36 KMTases as the sole mediator of H3K36 trimethylation [97] and as a tumor-suppressor in most contexts. In human development, germline mutations in SETD2 cause an overgrowth condition with features similar to Sotos syndrome [98]. Underexpression and mutation of SETD2 are associated with poor prognosis in breast and renal cancer, GI stromal tumors, and acute leukemia [99–103]. SETD2 mutations are more common in leukemias with MLL-rearrangements (22%) than leukemias without such

rearrangements (5%), and SETD2 loss cooperates with MLL fusions like *MLL-AF9* and *MLL-NRIP* to create a more aggressive leukemia with increased self-renewal of leukemia stem cells [104].

Numerous studies have examined the tumor suppressor function of SETD2 in kidney tumors, in which SETD2 is mutated in 3-12% of clear cell renal cell carcinoma (ccRCC) [101,105] and a similar proportion of papillary renal cell carcinoma [106]. Multiple different loss-of-function mutations in SETD2 occur within spatially distinct regions of a single renal tumor, suggesting strong selective pressure for SETD2 inactivation [107]. Furthermore, micro RNA-mediated repression of SETD2 expression occurs broadly in ccRCC tumor samples and cell lines, causing reductions in SETD2 function where overt mutations may not be present [108]. The molecular mechanisms underlying SETD2's suppression of renal cancer reflect the physiological functions of H3K36 methylation in transcript processing and genome integrity, as outlined above. Briefly, loss of function of SETD2 in ccRCC results in defects in DNA repair, nucleosome dynamics, RNA processing, and DNA methylation [109–112].

Other H3K36 methyltransferases: SMYD2, SETMAR, and SETD3

As a member of the SMYD family of proteins, SMYD2 contains a bipartite SET domain split by a zinc-finger MYND domain. SMYD2 promotes cancer cell proliferation in head and neck squamous cell carcinoma and esophageal squamous cell carcinoma [49,50], and overexpression of SMYD2 is a poor prognostic factor in pediatric acute lymphoblastic leukemia (ALL), gastric cancer, and bladder cancer [48,51,52]. SMYD2 methylates H3K4 and H3K36 *in vitro* [113], and SMYD2-mediated H3K36me₂ was reported to repress transcription of pro-inflammatory cytokines IL-6 and TNF- α in macrophages [114]. However, inhibition or knockdown of SMYD2 does not change global H3K36 or H3K4 mono-, di-, or trimethylation,

and most of SMYD2 is found in the cytoplasm, suggesting SMYD2's activity on chromatin may be minimal [115]. Indeed, SMYD2 methylates many non-histone substrates including p53, RB, HSP90, estrogen receptor α , and PARP1 [116–120], conveying wide-ranging effects on transcriptional regulation, protein homeostasis, apoptosis, and the DNA damage response.

The *SETMAR* gene evolved from fusion of a SET domain to a transposase domain [121]. As described above, the SETMAR protein has been linked to cancer via its role in DNA repair, which requires SETMAR's KMTase activity [40]. In colon cancer SETMAR was further found to activate genes involved in repair, synthesis, and methylation of DNA, as well as stemness markers [122], but the contribution of KMTase activity to this function was not determined. Interestingly, purified SETMAR protein is not active on nucleosomes *in vitro*, raising the possibility that SETMAR may function by methylating non-histone substrates *in vivo* [42]. In addition, SETMAR undergoes automethylation near the active site of the transposase domain [123], but how automethylation regulates its DNA repair function is unknown.

SETD3 is a poorly characterized putative tumor suppressor that methylates H3K4 and H3K36 *in vitro* and in SETD3-transfected cells [124]. SETD3 expression is lower in renal cell tumors than normal renal tissues, and low expression of SETD3 is associated with shorter survival in renal cell carcinoma patients [125]. SETD3 has been implicated in DNA replication and repair due to its interaction with proliferating cell nuclear antigen (PCNA), a conserved factor involved in DNA synthesis [126].

H3K36-specific SET domains: structural considerations for inhibitor development

Structural details of target proteins are extremely valuable in the development of potent and specific inhibitors. For the H3K36-specific KMTases, structural studies of the catalytic SET domain have revealed valuable information to jumpstart inhibitor development. An important

feature shared among the NSD, ASH1L, and SETD2 SET domains is an autoinhibitory loop blocking access of histone substrate to the active site [127–129]. This loop must undergo a conformational change to accommodate substrate binding. In contrast, structural studies of the more distantly related SMYD2 SET domain showed that substrate peptides bind this SET domain with minimal conformational changes [130,131].

Structural studies of NSD, ASH1L, and SETD2 SET domains demonstrate feasibility of inhibitor design

The SET domains of NSD1-3, ASH1L, and SETD2 consist of four subdomains: associated with SET (AWS), core SET, SET-insertion (SET-I), and post-SET (**Figure 1.3A**). AWS is a less structured region at the N-terminus that does not participate directly in catalysis, but the latter three subdomains contribute to the active site. The beta-sheet-rich core SET region is split by SET-I, a sequence-variable subdomain that has been proposed to play a role in substrate specificity [132]. Post-SET is a cysteine-rich subdomain that contains the autoinhibitory loop (**Figure 1.3A**). The methyl donor SAM binds in a pocket at the junction of core SET, SET-I, and post-SET, which make extensive contacts to SAM including hydrogen bonds and van der Waals contacts. Other important interactions involve tyrosine and carbonyl oxygens that form CH \cdots O hydrogen bonds with the SAM methyl group [133].

The autoinhibitory loop must undergo a significant conformational change to accommodate nucleosome substrate, and interestingly this loop undergoes dynamics even without nucleosome present. Molecular dynamics simulations, crystallographic, and NMR studies have shown that the NSD1 and ASH1L autoinhibitory loops experience conformational heterogeneity in the absence of nucleosome substrate [134,135]. In none of the observed conformations does the autoinhibitory loop completely open to allow nucleosome binding, suggesting that interaction with nucleosome is required to induce loop opening. Nevertheless,

conformational dynamics of the loop may be important for facilitating the larger conformational change that must occur upon nucleosome binding [134,135]. Interestingly, the autoinhibitory loop in the crystal structure of the related NSD3 SET domain is disordered [136], further suggesting that autoinhibitory loop dynamics are characteristic of the NSD and related SET domains. In the case of SETD2, both auto-inhibited and open conformations of the SET domain were observed by crystallography [129] (**Figure 1.3B**). When bound to S-adenosyl homocysteine (SAH), SETD2 adopts an autoinhibited conformation similar to NSD1 and ASH1L, with the sidechain of Arg1670 in the autoinhibitory loop extending into the putative substrate lysine binding channel (**Figure 1.3B**). In contrast, N-propyl sinefungin (Pr-SNF) was identified as a structural probe that replaces SAH and stabilizes an open conformation of SETD2 (**Figures 1.3B and C**). Pr-SNF partially occupies the lysine binding channel and causes a dramatic conformational change in the autoinhibitory loop, as this loop moves away from core SET and Arg1670 flips 15 Å away from the lysine binding channel [129].

The autoinhibitory loop region of the NSD and related KMTases provides an intriguing opportunity for H3K36 KMTase inhibitor development. The conformational heterogeneity observed in this region may create transient pockets not directly observed in crystal structures into which small molecules could bind. These ligands could function as allosteric inhibitors by preferentially interacting with and stabilizing the autoinhibited conformation of the SET domain. Importantly, the autoinhibitory loop is poorly conserved in primary sequence and in structure among NSD1, ASH1L, and SETD2 (**Figures 1.3A and 1.4**), suggesting that targeting this region could result in specific inhibitors for the different KMTases. In NSD1, the autoinhibitory loop is stabilized by a β -turn within the loop as well as hydrophobic contacts with SET-I [128]. In contrast, no β -turn exists in the ASH1L autoinhibitory loop, and there are longer range contacts

to SET-I [127,134]. In the SETD2 autoinhibitory loop, there is also no β -turn to stabilize the loop, but the interaction of Arg1670 at the beginning of the loop with the substrate lysine binding channel pulls the loop into contact with SET-I [129] (**Figure 1.3B**). Thus, the autoinhibitory loop region could be exploited for designing specific inhibitors of H3K36 KMTases.

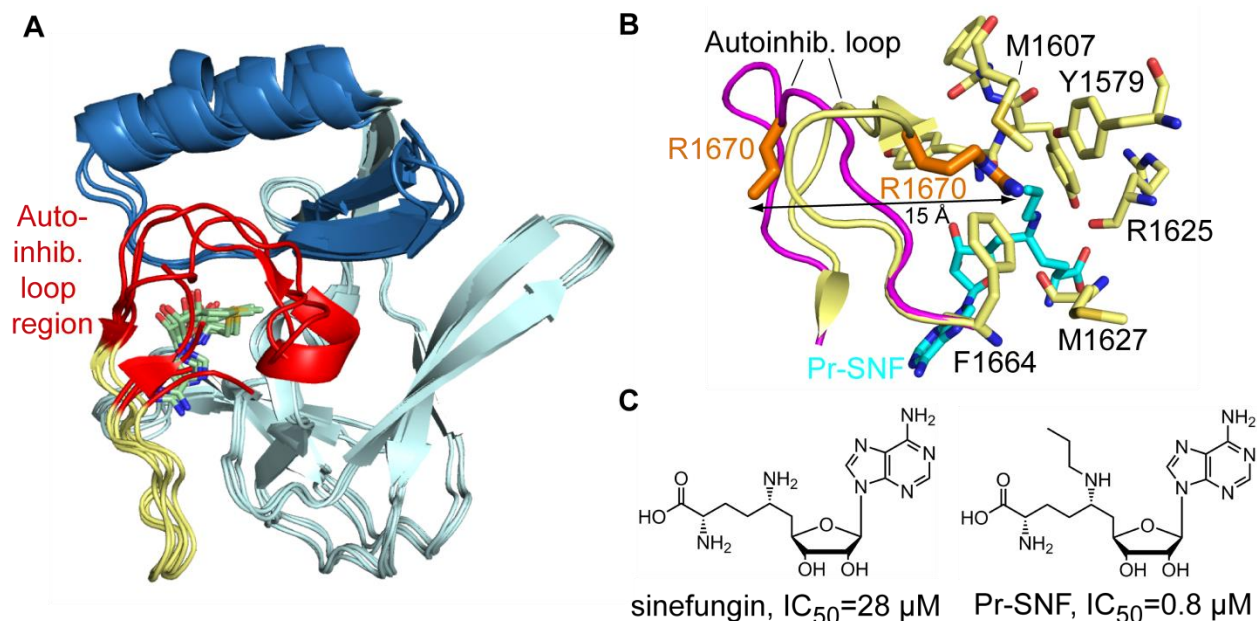


Figure 1.3. Structures and structural probes of H3K36 KMTases. **(A)** Overlay of the core SET (pale cyan), SET-I (sky blue), and post-SET (pale yellow) subdomains of NSD1 (PDB code 3OOI), NSD3 (4YZ8), ASH1L (4YNM), and SETD2 (4H12). The structurally variable autoinhibitory loop region is colored in red. **(B)** Binding of the Pr-SNF inhibitor to SETD2 causes opening of the autoinhibitory loop. Autoinhibitory loop conformation is shown with SAH bound (pale yellow, 4H12) and with Pr-SNF bound (magenta, 4FMU). Steric clash between the propyl moiety of Pr-SNF (cyan) and the Arg1670 sidechain (orange) causes Arg1670 to flip out a distance of 15 Å. Residues stabilizing Arg1670 in the putative substrate lysine binding channel in the SAH-bound form of SETD2 are shown in pale yellow sticks. **(C)** Chemical structures of SETD2 inhibitors sinefungin and Pr-SNF. *In vitro* IC_{50} values for SETD2 are listed.

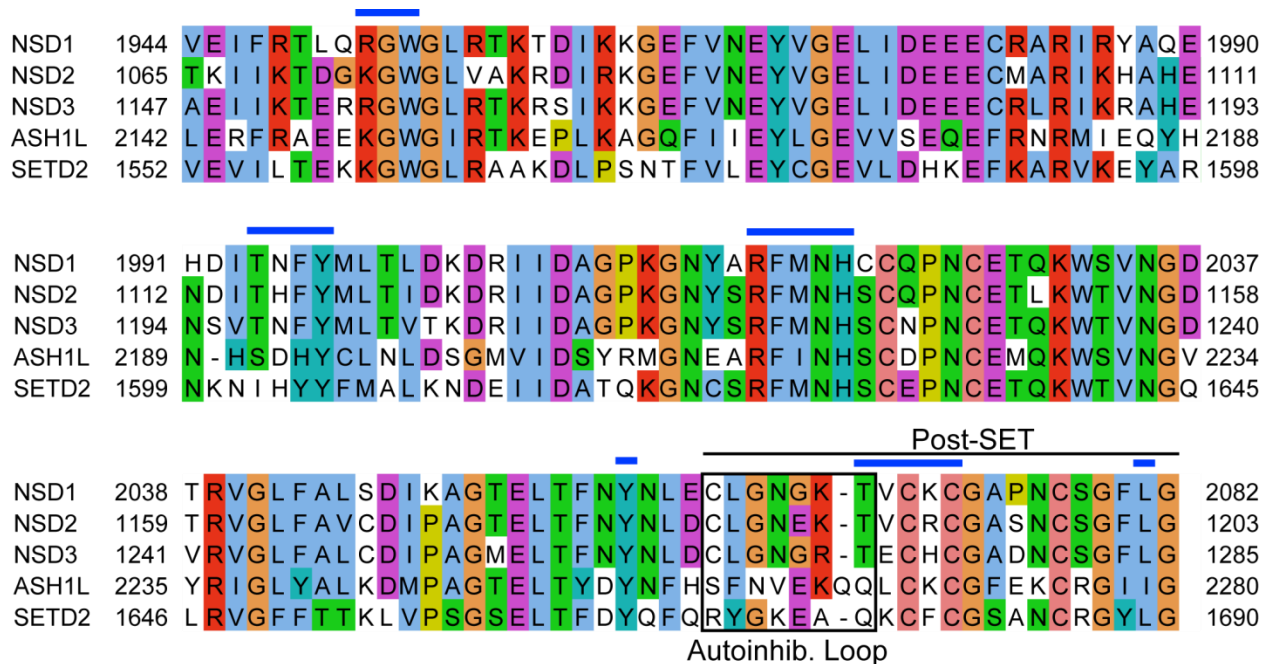


Figure 1.4. Sequence alignment of core SET and post-SET regions of the related NSD, ASH1L, and SETD2 KMTases. Blue lines indicate conserved contacts to SAM cofactor. Black box indicates the autoinhibitory loop, black line indicates post-SET subdomain.

SMYD2 structures: U-shaped substrate binding

Crystal structures of full-length SMYD2 show an overall structure of two lobes or domains separated by a deep groove (**Figure 1.5A**). The N-terminal lobe contains the core SET region surrounded by a MYND zinc finger, the SET-I subdomain, and the cysteine-rich post-SET subdomain. The C-terminal domain (CTD) is made up of seven anti-parallel α -helices in a structure reminiscent of the tetratricopeptide repeat (TPR), a motif involved in protein-protein interactions. The CTD regulates KMTase activity of the SET domain, possibly via interactions the CTD forms with the substrate protein [130]. The CTD also controls activity and substrate specificity via an interaction with the SMYD2 cofactor Hsp90 [113,137]. The SAM cofactor binds in the crevasse between the N- and C-terminal lobes and makes direct contacts with core SET, SET-I, and post-SET (**Figure 1.5A**). Regulatory cross-talk between the CTD and SET domain active site was suggested by the observation that binding of the SAM analog sinefungin

induces a long range conformational change in the first two α -helices of the CTD, resulting in a more open conformation of SMYD2 [137,138].

Ternary complexes of SMYD2 have been solved with SAM and p53 peptide, SAH and methylated p53 peptide, and SAM and ER α peptide (**Figure 1.5A**). The p53 and ER α peptides bind in a U-shaped conformation at the hinge region between the N- and C-terminal lobes, with the target lysine at the base of the “U” [130,131,139]. The target lysine projects deep into a hydrophobic cavity, with the lysine ϵ -amino group oriented by the hydroxyl groups of two conserved tyrosine residues and several main chain carbonyl groups. Although the CTD does not directly interact with SAM cofactor, residues from the CTD are involved in binding to substrate peptides. Remarkably, given the narrow groove between the N and C-terminal lobes and the previously described flexibility in the CTD, only minor structural changes in SMYD2 occur upon binding to substrate peptides [130,131].

Inhibitors of H3K36 methyltransferases

Inhibitors for two members of the H3K36 methyltransferases, SETD2 and SMYD2, have been reported. Zheng et al. developed N-propyl sinefungin (Pr-SNF) and N-benzyl sinefungin, analogs of the broad-spectrum KMTase inhibitor sinefungin, as specific inhibitors of SETD2. Pr-SNF and N-benzyl sinefungin inhibit SETD2 KMTase activity *in vitro* with IC₅₀ values of 0.8 μ M and 0.5 μ M, respectively (**Figure 1.3C**). These values represent activities more than 10-fold better than unmodified sinefungin, and with at least 2-fold selectivity over a panel of related KMTases. A co-crystal structure of SETD2 in complex with Pr-SNF revealed that the N-propyl moiety partially occupies the lysine binding channel and causes a dramatic conformational change in the autoinhibitory loop [129]. No activity for Pr-SNF has been reported in cells, and sinefungin analogs in general have poor cell membrane permeability. Nevertheless, combined

occupancy of the SAM and substrate lysine binding pockets by small molecules is an intriguing strategy that warrants further study in other KMTases.

The first specific SMYD2 inhibitor, AZ505, was discovered using an AlphaLISA-based high throughput screen of 1.23 million compounds from the Astra Zeneca library [131] (**Figure 1.5B**). The compound binds with K_d of 0.5 μ M to the peptide binding groove of SMYD2 and is greater than 100-fold selective over related KMTases including SMYD3. The AZ505 binding site partially overlaps with that of p53 peptide (**Figure 1.5C**). Interestingly, the benzoxazinone moiety of AZ505 occupies the substrate lysine binding channel, indicating that this channel may accommodate much larger and more rigid groups than a lysine sidechain. The cyclohexyl group of AZ505 is positioned in a hydrophobic pocket occupied by Leu369 of the p53 peptide. However, the dichlorophenethyl group of AZ505 extends into a second hydrophobic pocket that is not occupied by the p53 substrate peptide.

AZ505 was optimized to improve potency while retaining cellular permeability. Nguyen et al. developed the SMYD2 inhibitor LLY-507 by making substantial changes to AZ505, using a pyrrolidine group rather than benzoxazinone to occupy the substrate lysine binding channel (**Figures 1.5B and C**). LLY-507 inhibits SMYD2 methyltransferase activity in an *in vitro* assay with $IC_{50} < 15$ nM but maintains an excellent selectivity profile over other KMTase and non-KMTase enzymes. LLY-507 inhibits p53 methylation in HEK293 and U2OS cells overexpressing SMYD2 and/or p53, and the compound inhibits proliferation of esophageal squamous cell carcinoma (ESCC), hepatocellular carcinoma (HCC), and breast cancer cell lines [115].

A second SMYD2 inhibitor, A-893, was also inspired by the AstraZeneca HTS hit but makes fewer changes to the AZ505 scaffold (**Figures 1.5B and C**). A-893 inhibits SMYD2

KMTase activity *in vitro* with an IC₅₀ of 2.8 nM and blocks p53 methylation in the A549 lung carcinoma cell line. This large increase in activity over AZ505 was achieved by moving a hydroxyl group from the benzooxazinone to the diethylamine linker. The increase in potency is likely mediated by a hydrogen bond that the relocated hydroxyl group forms with the backbone carbonyl of Tyr240 [140].

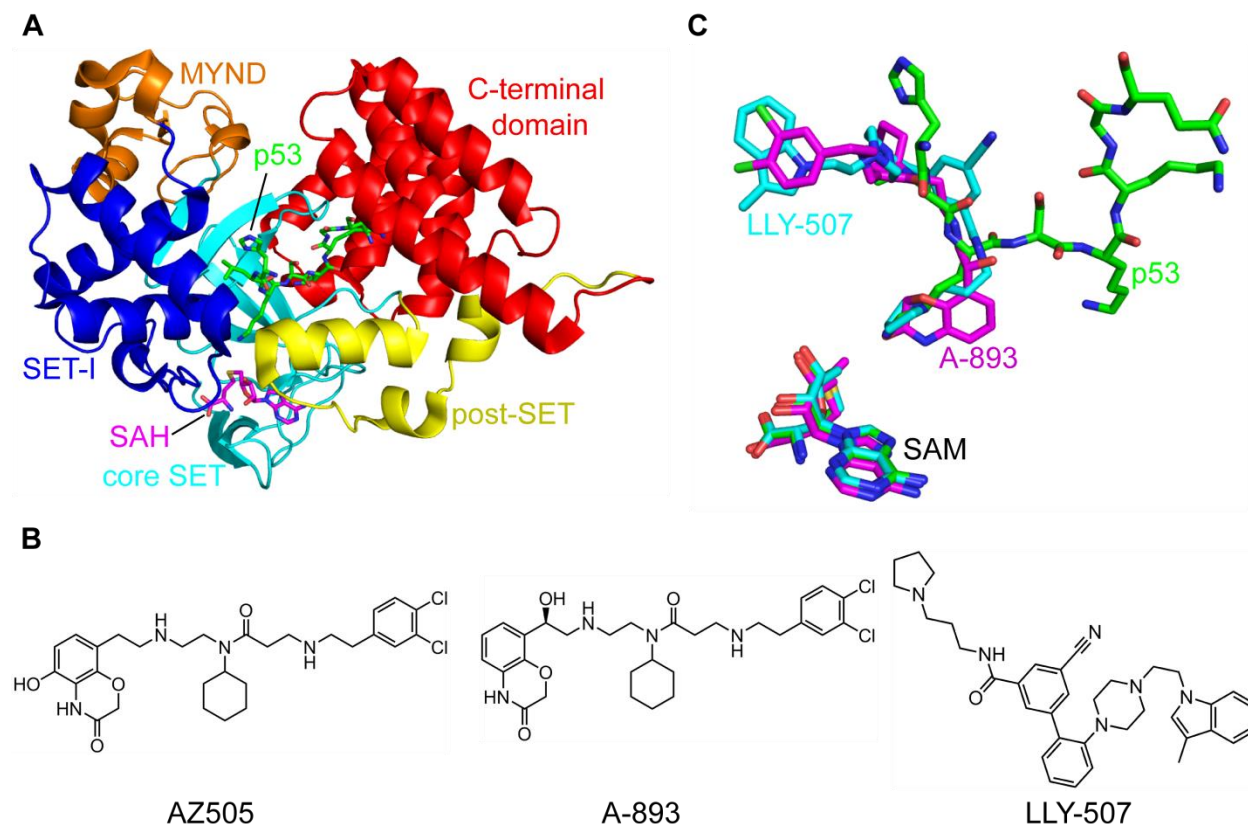


Figure 1.5. Peptide substrate-competitive inhibitors of the SMYD2 KMTase. **(A)** Crystal structure of SMYD2 bound to p53 peptide (PDB code 3TG5). **(B)** Chemical structures of SMYD2 inhibitors. **(C)** Overlay of binding modes of SMYD2 inhibitors LLY-507 (4WUY) and A-893 (4YND) and p53 peptide. The weaker compound AZ505 binds to the same site but has been omitted for clarity. SAM position included for reference.

Challenges and opportunities in developing specific inhibitors of H3K36 KMTases

Targeting the catalytic SET domain

The H3K36 KMTases are highly relevant to biology and medicine, and development of potent and specific inhibitors for this class of proteins is urgently needed. However, very few

specific inhibitors of H3K36 KMTases have been reported, highlighting difficulties associated with this target class. First, developing *in vitro* methylation assays for H3K36 KMTases can be challenging because the SET domain may require chromatin-interacting domains and/or co-factor proteins for optimal activity on nucleosome substrate. For example, the isolated ASH1L SET domain has weak KMTase activity on nucleosomes, but constructs including the PHD, Bromo, and BAH domains have increased activity [134]. An additional obstacle is that inhibitors of KMTases appear to be rare in current high-throughput screening (HTS) libraries, as three different pharmaceutical companies published SAM-competitive inhibitors of the EZH2 KMTase with similar pyridone-containing scaffolds [11]. KMTase-privileged screening libraries, which could be based on known KMTase inhibitors like sinefungin and chaetocin [129,141], may help increase the number of screening hits.

In contrast to screening through large compound libraries, a more focused approach to developing KMTase inhibitors could proceed by modifying the SAM cofactor. Structural chemical analysis of ten different human KMTase SET domains suggested that the SAM binding pocket represents a druggable region because a low proportion of the SAM molecular surface is accessible to solvent when bound to the SET domain. Moreover, although the conformation of bound SAM cofactor is very similar across different SET domains due to a precise hydrogen bond network, SAM binding pockets contain enough structural diversity surrounding this network to allow development of specific inhibitors for different KMTases [142]. Indeed, chemical modification of SAM or closely related compounds like sinefungin has led to potent and specific inhibitors for several KMTases [129,143,144]. An additional consideration for designing SAM-derivative inhibitors is that the SAM binding pocket is immediately adjacent to the narrow substrate lysine binding channel. Therefore, weak or non-specific SAM derivatives

may be improved to target a specific KMTase by adding hydrophobic groups that interact with the lysine binding channel. Proof of principle for this hypothesis was demonstrated by sinefungin analogs that specifically inhibit SETD2 [129]. The potency of SAM-derivative inhibitors could be further enhanced by including moieties that mimic the transition state for the methyl transfer reaction, as enzyme theory predicts that KMTases should bind the transition state more tightly than the individual SAM and histone substrates [145]. Kinetic isotope effects were recently used to generate a model of the NSD2 transition state, providing an opportunity for development of transition state mimics [146]. Thus far, poor membrane permeability of the available SAM-derivative inhibitors has proven a challenge in applying these compounds to cellular systems [147].

Alternative druggable regions of H3K36 KMTase proteins

Most H3K36-specific KMTases are large epigenetic regulators that contain multiple protein-protein interacting (PPI) domains in addition to the catalytic SET domain. In a growing number of cases, PPI domains of KMTase proteins have been found essential for recruiting the SET domain to its methylation targets. PPI domains may also have SET domain-independent oncogenic roles. Structural studies have revealed that PPI domains could be highly druggable targets [148,149]. Therefore, to overcome challenges associated with inhibitor development for H3K36 SET domains, an alternative approach is to target the PPI domains.

For example, PPI domains of NSD2 are essential for the protein's oncogenic activity. The N-terminal PWWP domain of NSD2 preferentially binds H3K36me₂ and is important for recruiting NSD2 to its target genes [7]. Point mutations in the PWWP domain that prevent recognition of methyllysine inhibit NSD2's ability to increase H3K36 methylation and cancer cell proliferation. The fact that the NSD2 PWWP domain recognizes the same mark produced by

the NSD2 SET domain suggests a role for the PWWP domain in spreading H3K36 dimethylation to neighboring genomic regions and to daughter cells during DNA replication [7]. NSD2 also contains four PHD domains, a family of epigenetic reader modules that bind methylated lysines on H3 [150]. The NSD2 PHD domains are required for epigenetic changes and oncogenic activity mediated by NSD2 due to their role in recruiting NSD2 to chromatin [47,151]. Structural studies have shown that PWWP and PHD domains interact with histone via an aromatic cage that surrounds methyllysine, as well as a hydrogen bonding network with nearby histone residues (**Figure 1.6A and B**) [148,149]. Although no well validated inhibitors of PWWP or PHD domains have yet been reported, the histone-binding surface of these domains provides opportunities for development of small molecules that compete with histone [150,152].

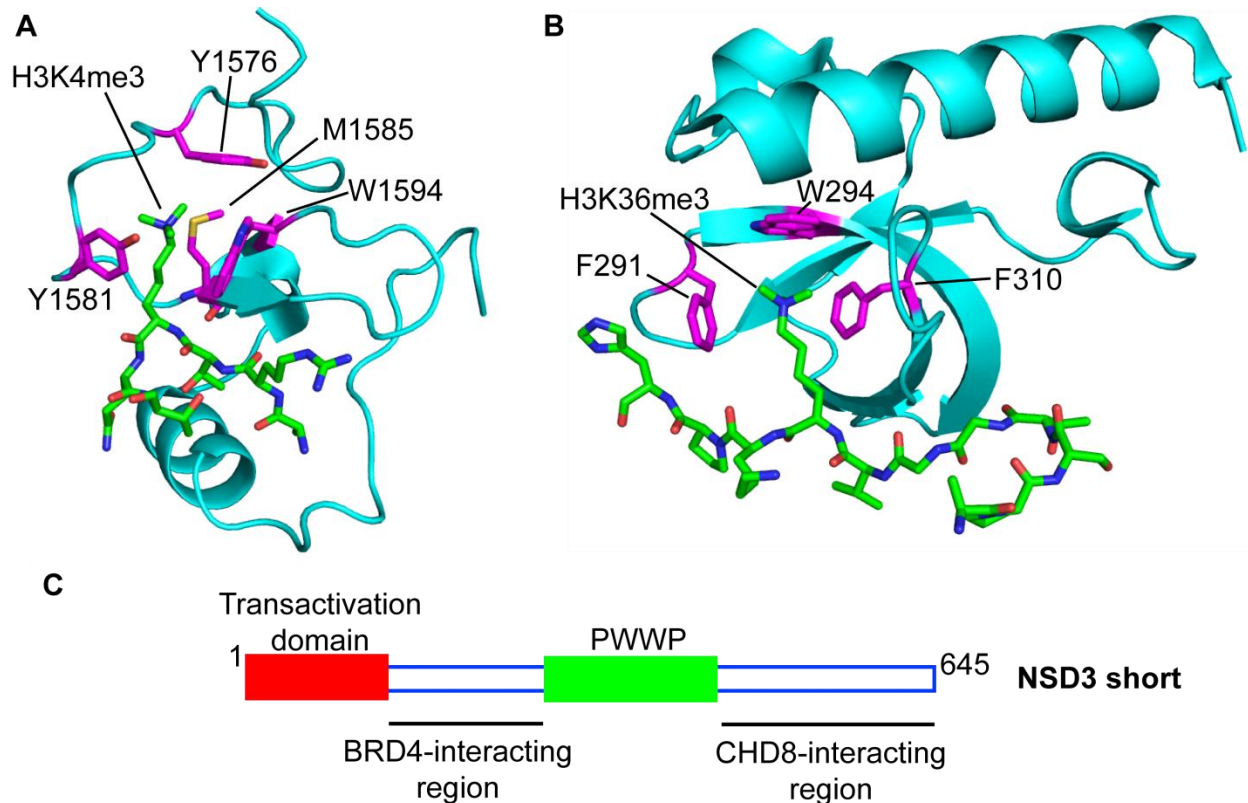


Figure 1.6. Targeting PPI domains of H3K36 KMTases as an alternative to the SET domain. Examples of PHD and PWWP domains: (A) Third PHD domain of MLL1 (cyan) bound to H3K4me3 peptide (green) (PDB code 3LQJ) and (B) PWWP domain of ZMYND11 (cyan) bound to H3.3K36me3 peptide (green) (4N4I). Residues on the PPI domain that form the aromatic cage are colored magenta. (C) PPI regions in the N-terminus of NSD3 that are required for leukemogenesis, adapted with permission from [6].

In some cases, KMTase proteins may engage in PPIs that drive oncogenesis independently of the SET domain. For example, recent studies have shown that NSD3 engages in a rich repertoire of PPIs with other chromatin readers and remodelers. In a subset of NUT midline carcinoma (NMC) cases, an NSD3-NUT fusion protein, which lacks the NSD3 SET domain, blocks differentiation [153]. The NSD3 N-terminus in the fusion protein interacts with the bromodomain-containing protein BRD4 to drive oncogenesis, and NSD3-NUT NMC cells are sensitive to the bromodomain inhibitor JQ1. In addition, wild-type *NSD3* is required for proliferation of NMC cells harboring a BRD4-NUT fusion protein, demonstrating that the NSD3/BRD4 interaction is key to the pathogenesis of NMC in general [153]. The interaction

between NSD3 and BRD4 was further characterized in AML, in which a short isoform of NSD3 lacking the SET domain drives leukemogenesis by serving as a bridge between BRD4 and the chromatin remodeler CHD8. An N-terminal transactivation domain and PWWP domain are also necessary for NSD3 to activate transcription and maintain leukemia (**Figure 1.6C**) [6]. Thus, in some contexts PPIs rather than KMTase activity are key to NSD3's oncogenic role.

These studies reveal that H3K36 KMTases harbor a variety of potentially druggable domains and regions in addition to the SET domain. Precise targeting of individual domains with small molecule inhibitors will ideally leave other protein functions unaffected and may reveal highly specific cancer dependencies. On the other hand, in some cases the entire KMTase protein may need to be eliminated to achieve anti-cancer activity. To overcome this challenge, one could use small molecule ligands capable of targeting proteins for proteolytic degradation [154,155]. For example, Winter et al. have recently found that coupling a ligand of a target protein to phthalimide results in engagement of the cellular ubiquitin ligase machinery and degradation of the target protein [154]. This method could be applied to KMTases, such that phthalimide conjugates of KMTase ligands could be used to degrade an entire KMTase protein.

Conclusions and future perspectives

The H3K36-specific KMTases are prominent drug targets due to numerous studies linking them to oncogenesis. However, large knowledge gaps remain surrounding the mechanisms by which H3K36-specific KMTases drive cancer growth, whether it is by controlling transcription, splicing, DNA repair, or other cellular processes. Small molecule inhibitors for H3K36-specific KMTases will be invaluable chemical tools to better understand the role of these proteins in cancer. Developing potent and specific inhibitors for this challenging target class will require creative and non-traditional approaches. Derivatives of SAM and

sinefungin have shown promise as specific inhibitors of KMTases and this compound class could be further expanded to cover H3K36-specific KMTases. In addition, the autoinhibitory loop is a unique feature of H3K36-specific SET domains that could be exploited for inhibitor development. Furthermore, some PPI domains of KMTase proteins are required for oncogenic functions and provide excellent opportunities for targeting by small molecules. Major progress on the structure and function of H3K36 KMTases in recent years leads us to predict that the KMTase inhibitor field lies on the precipice of many innovative and exciting discoveries.

Chapter 2. Two loops undergoing concerted dynamics regulate activity of the ASH1L histone methyltransferase*

Abstract

ASH1L (absent, small, or homeotic-like 1) is a histone methyltransferase (KMTase) involved in gene activation that is overexpressed in multiple forms of cancer. Previous studies of ASH1L's catalytic SET domain identified an autoinhibitory loop that blocks access of histone substrate to the enzyme active site. Here, we used both NMR and X-ray crystallography to identify conformational dynamics in the ASH1L autoinhibitory loop. Using site-directed mutagenesis we found that point mutations in the autoinhibitory loop that perturb the structure of the SET domain result in decreased enzyme activity, indicating that the autoinhibitory loop is not a simple gate to the active site but is rather a key feature critical to ASH1L function. We also identified a second loop in the SET-I subdomain of ASH1L that experiences conformational dynamics, and we trapped two different conformations of this loop using crystallographic studies. Mutation of the SET-I loop led to a large decrease in ASH1L enzymatic activity in addition to a significant conformational change in the SET-I loop, demonstrating the importance of the structure and dynamics of the SET-I loop to ASH1L function. Furthermore, we found that three C-terminal chromatin-interacting domains greatly enhance ASH1L enzymatic activity and that ASH1L requires native nucleosome substrate for robust activity. Our study illuminates the role of concerted conformational dynamics in ASH1L function and identifies structural features important for ASH1L enzymatic activity.

* Rogawski DS, Ndoj J, Cho HJ, Maillard I, Grembecka J, Cierpicki T. Two Loops Undergoing Concerted Dynamics Regulate the Activity of the ASH1L Histone Methyltransferase. *Biochemistry* 54(35), 5401-13 (2015).

Introduction

ASH1L (absent, small, or homeotic-like 1) is a mammalian homolog of Ash1, a member of the trithorax group of proteins essential for epigenetic mechanisms of gene activation [156]. ASH1L is a SET domain-containing histone methyltransferase (KMTase) with controversial substrate specificity. ASH1L was shown to methylate H3K4 [157,158] and H3K36 [44,93,127,159,160] in mammals. Although the specificity of the SET domain and function of ASH1L *in vivo* are not yet clear, emerging data link ASH1L to multiple cancers. In breast cancer, 27 percent of aggressive, basal-like breast cancers have high-level copy number amplifications of the *ASH1L* gene [161]. Moreover, high levels of ASH1L mRNA are associated with shorter survival in breast cancer patients [161]. In thyroid cancer, ASH1L is overexpressed in tumor-specific truncated forms and is downregulated by a tumor suppressor microRNA [94]. Amplifications of ASH1L are found in a variety of other tumors, such as lung and uterine cancer [162,163], while mutations in ASH1L have been identified in gastric cancer [164], colorectal cancer [165], esophageal squamous cell cancer [166], and lung cancer [167]. ASH1L activates genes in the *HOX-A*, *HOX-B*, and *HOX-C* clusters [15,88], which are overexpressed in multiple cancers and correlated with metastasis and aggressive disease [89]. Notably, ASH1L activates *HOXA9* and its collaborator *MEIS1* [15], which are oncogenes in leukemia [168].

The ASH1L protein has a large and unannotated N-terminus, a SET domain responsible for KMTase activity, and three C-terminal chromatin-interacting domains: bromodomain, plant homeodomain (PHD), and bromo-associated homology (BAH) domain [93]. ASH1L's catalytic SET domain is required for its gene activating function, as deletion of the ASH1L SET domain in differentiating mouse embryonic stem cells leads to a decrease in expression of multiple genes, including members of the *Wnt* and *Hox* families [93]. Most SET domains including

ASH1L SET can be divided into four subdomains: associated with SET (AWS), core SET, post-SET, and the variable SET-I subdomain positioned in the middle of core SET, which is a putative substrate specificity cassette for KMTases [132]. In ASH1L and the closely related NSD1 and SETD2 KMTases, a region in the post-SET subdomain called the autoinhibitory loop blocks access of histone substrate to the enzyme active site [127–129]. It is unclear how the autoinhibitory loop reorients to accommodate substrate binding, as the structures for ASH1L, NSD1, and SETD2 were determined in the absence of histone substrate. Furthermore, whether the autoinhibitory loop functions simply as a gate to the active site or whether it forms important interactions with nucleosome substrates is unknown. In the case of NSD1, molecular dynamics simulations showed modest flexibility of the NSD1 autoinhibitory loop that may permit the H3K36 sidechain to access NSD1's lysine binding channel [128], but this computational study was not confirmed experimentally. In the case of SETD2, an open and substrate-accessible conformation of the autoinhibitory loop was observed by crystallographic studies upon binding of *N*-propyl sinefungin, which forces reorientation of Arg1670 that normally occupies the substrate lysine binding channel [129]. Even less is known about the function of the autoinhibitory loop of ASH1L. In the ASH1L SET domain crystal structure, high B factors for the autoinhibitory loop led the authors to conclude that this loop is highly mobile [127]. The functional significance of this mobility in solution, however, remains unclear. Interestingly, a Q2265A ASH1L mutant with increased catalytic activity had a highly disordered autoinhibitory loop by preliminary structural analysis, suggesting that destabilization of the autoinhibitory loop might be sufficient to increase ASH1L enzymatic activity [127].

Here we aimed to better understand the function of ASH1L by investigating structural features of the ASH1L SET domain and assessing how they regulate its KMTase activity. Using

X-ray crystallography and NMR we found that two loops surrounding the active site of ASH1L—the autoinhibitory loop and a loop in the SET-I subdomain—undergo concerted conformational dynamics. We designed several mutations in the SET domain to perturb the conformation of these loops and characterized the structure and activity of the mutants. We found that the autoinhibitory loop is not a simple gate blocking access to the active site. Instead, the autoinhibitory loop and the SET-I loop represent important structural features required for ASH1L SET domain activity. Moreover, our study emphasizes that concerted dynamics play a significant role in ASH1L KMTase activity. These results shed light on the mechanisms of SET domain function and may provide a foundation for development of ASH1L inhibitors.

Results

Crystal structure of the ASH1L SET domain shows increased dynamics of the autoinhibitory and SET-I loops

To gain insight into the mechanism of methyltransferase activity by ASH1L, we determined the crystal structure of the ASH1L SET domain (residues 2069-2288) (**Figure 2.1A**). The crystals diffracted to 2.2 Å resolution, representing a significant improvement over the previously published 2.9 Å structure of the ASH1L SET domain [127]. Higher resolution allowed us to more precisely model residues throughout the structure, including regions associated with the catalytic activity of ASH1L. We observed satisfactory electron density for the main chain and for the majority of sidechains in the autoinhibitory loop (residues 2258-2266) (**Figure 2.1B**). The poor electron density for some sidechains suggested that the autoinhibitory loop may experience conformational dynamics, and to further evaluate dynamics of the SET domain we used crystallographic B factors. We found that the autoinhibitory loop and the AWS region have the highest B factors, suggesting that these are more mobile regions (**Figure 2.1C**). Interestingly, we also found that the SET-I subdomain has higher B factors than the neighboring

C-terminal core SET region (SET-C) (**Figure 2.1C**). The SET-I subdomain consists of a helix-loop-strand-turn-strand motif, with the SET-I loop (residues 2187-2195) in this motif located directly beneath the autoinhibitory loop (**Figure 2.1D**). Residues 2192-2194 in the SET-I loop make both polar and hydrophobic contacts with S-adenosyl methionine (SAM) cofactor. We observed well-defined electron density for SET-I residues, with the exception of the sidechain of His2193, which is poorly defined, indicating that this residue may undergo conformational exchange (**Figure 2.1D**).

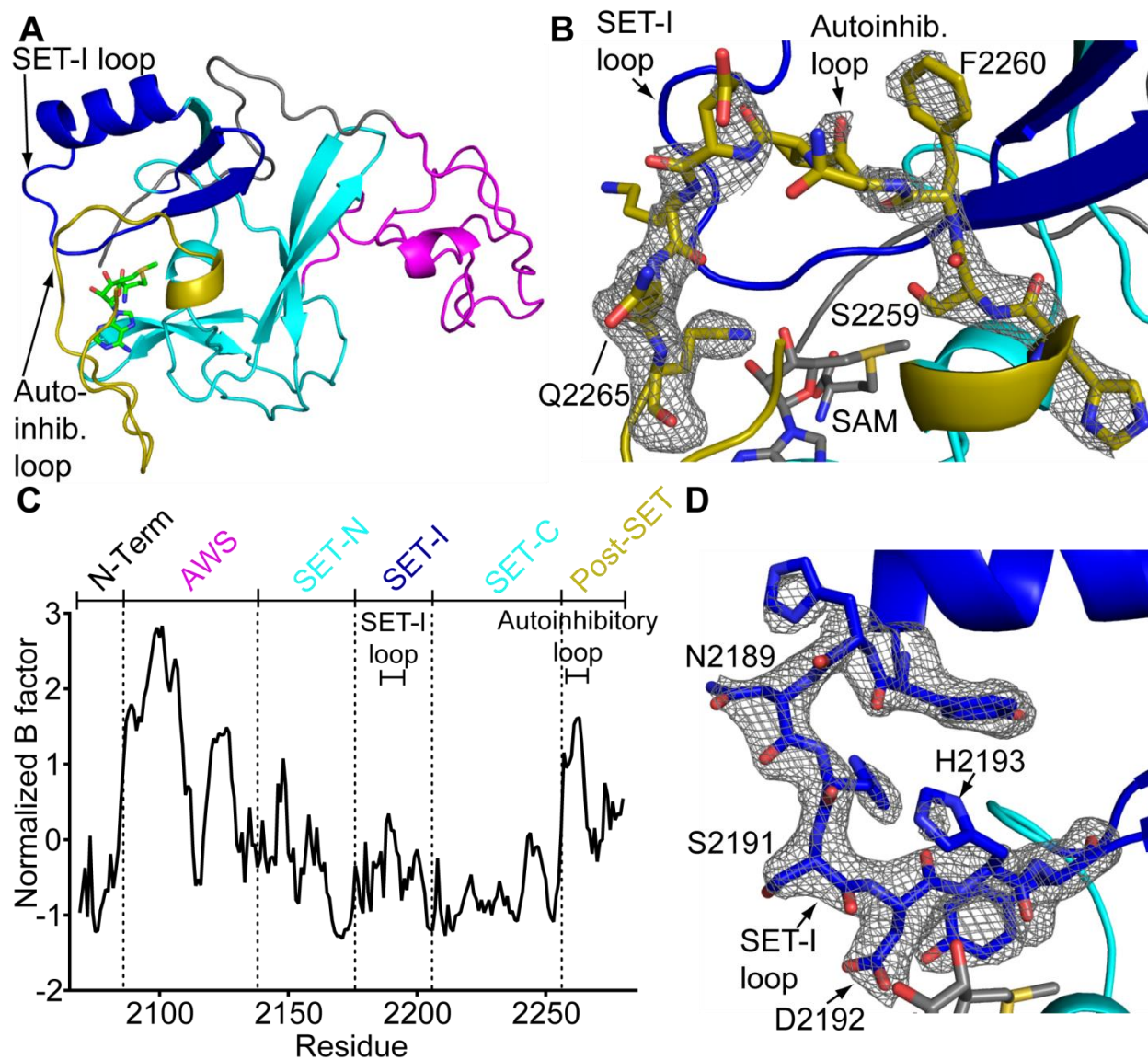


Figure 2.1. Analysis of ASH1L SET domain crystal structure. (A) Cartoon representation of the ASH1L SET domain, colored by subdomain: N-terminus, grey; AWS, magenta; core SET, cyan; SET-I, blue; post-SET, dark yellow. (B) The autoinhibitory loop of ASH1L with mFo-DFc omit map contoured at 2.5σ . Subdomains colored as in (A). (C) Normalized crystallographic B factors for the ASH1L SET domain. (D) The SET-I loop of ASH1L with mFo-DFc omit map contoured at 2.5σ . Subdomains colored as in (A).

NMR studies reveal that the ASH1L active site is surrounded by two loops experiencing conformational dynamics

To further investigate the dynamics of the ASH1L SET domain we performed NMR studies in solution. We collected a ^{15}N - ^1H TROSY NMR spectrum for the ASH1L SET domain. While the ASH1L SET construct contains 213 non-proline residues, we observed only 181

backbone amide peaks in the NMR spectrum. The large number of missing peaks suggested that a significant portion of the protein is undergoing intermediate exchange dynamics on the microsecond-millisecond timescale [169,170]. To identify the regions undergoing such dynamics, we completed backbone assignment based on triple resonance experiments for $^{15}\text{N}^{13}\text{C}$ -labeled ASH1L (**Figure 2.2A**). We were able to assign 168 (93%) of the 181 peaks observed for backbone amides. Interestingly, we were not able to observe any backbone amide peaks for residues in the autoinhibitory loop and in a large portion of the SET-I subdomain including the entire SET-I loop, while we obtained nearly complete assignment for the remaining part of the SET domain (**Figure 2.2A and B**). This observation strongly suggests that the autoinhibitory and SET-I loops surrounding the active site of ASH1L experience conformational dynamics, in agreement with crystallographic data presented above. Interestingly, while the crystallographic B-factors for the AWS subdomain are very high, we observe nearly all the AWS residues in the NMR spectrum (**Figure 2.2C**), suggesting that in solution the AWS subdomain undergoes dynamics at a much faster time scale than the SET-I loop and autoinhibitory loop. Overall, we concluded that two loops surrounding the active site of ASH1L, the autoinhibitory loop and SET-I loop, are undergoing dynamics in solution on the microsecond-millisecond timescale.

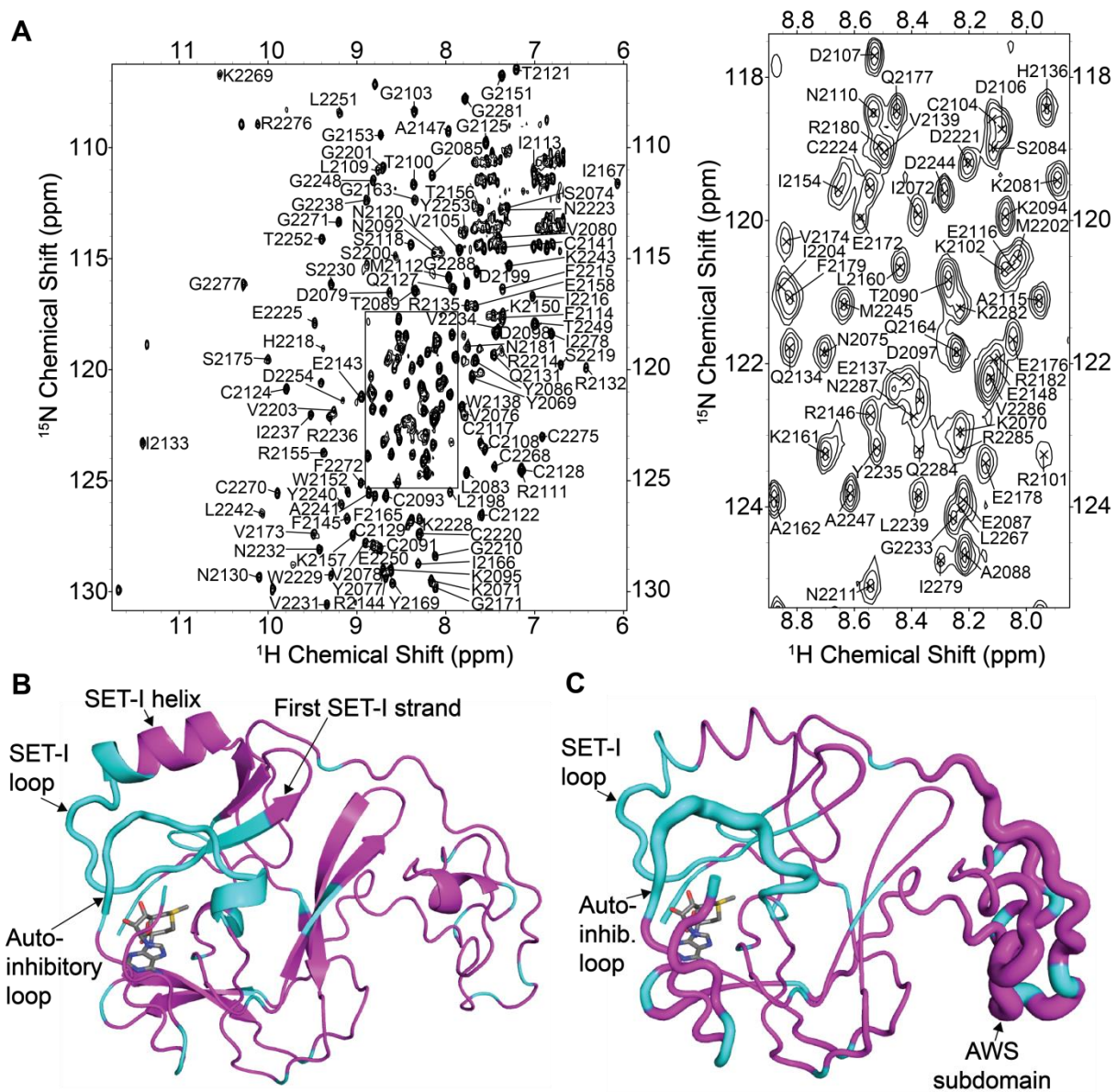


Figure 2.2. NMR studies of the ASH1L SET domain. (A) ^{15}N - ^1H TROSY spectrum of ASH1L SET domain with assignment. Right panel shows zoomed in view of the central region of the spectrum. (B) NMR assignment mapped onto ASH1L SET domain crystal structure. Residues with assigned peaks in the TROSY spectrum are magenta, while unassigned residues are cyan. (C) Cartoon putty representation of ASH1L with thickness of cartoon directly proportional to B factor. Assigned and unassigned residues colored as in (B).

ASH1L requires chromatin-interacting domains for robust enzymatic activity

We next aimed to develop a robust assay to measure the catalytic activity of ASH1L. A previous study showed weak catalytic activity for the isolated SET domain of ASH1L [93]. We tested the catalytic activity of ASH1L SET in a radiometric KMTase assay with different

substrates including chicken nucleosomes and detected weak activity using 1 μM ASH1L SET construct (**Figure 2.3A**) and 0.8 μM nucleosome (**Figure 2.3B, inset**). Because full-length ASH1L is a large multi-domain protein with three chromatin-interacting domains at its C terminus (a bromodomain [171], a PHD finger [172], and a BAH domain[173]) (**Figure 2.3A**), we wondered whether these domains would enhance ASH1L KMTase activity.

To investigate the effect of the chromatin-interacting domains on ASH1L KMTase activity, we designed three additional ASH1L constructs of different lengths systematically incorporating the chromatin reader domains at the C-terminus of the protein (**Figure 2.3A**). We tested the enzymatic activity of the different ASH1L constructs using 0.25 μM ASH1L and 0.2 μM chicken mono/di nucleosome substrate. We found that the isolated SET domain has no detectable KMTase activity on chicken nucleosome substrate under these assay conditions (**Figure 2.3B**). A larger construct incorporating the N-terminal flanking region of the SET domain (N-SET)[93] also had very low activity (**Figure 2.3B**). In contrast, the SET-PHD and SET-BAH constructs had significantly higher KMTase activity, with the longest construct SET-BAH being most active (**Figure 2.3B**). These results show that chromatin-interacting domains are necessary for robust enzymatic activity of ASH1L, likely through recruitment of nucleosome substrates. Despite significantly enhanced activity we were not able to characterize kinetic parameters for ASH1L because the signal was too weak at the low nanomolar enzyme concentrations required to accurately determine the Michaelis constants.

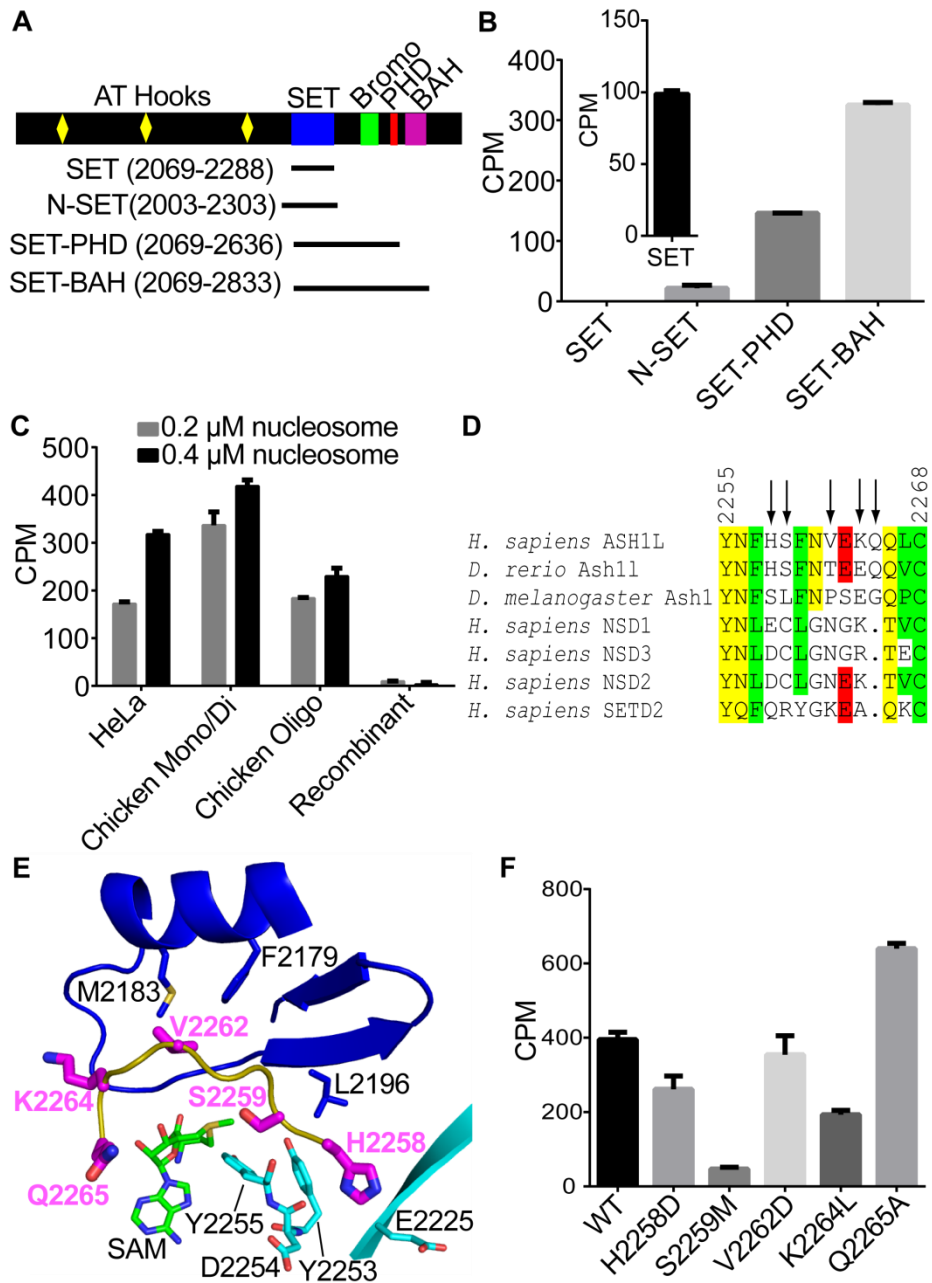


Figure 2.3. Enzymatic activity of ASH1L variants. (A) Schematic of ASH1L variants tested in the enzymatic assay. (B) Activity in counts per minute (CPM) of ASH1L constructs with chicken mono/di nucleosome substrate at 0.25 μ M ASH1L and 0.2 μ M nucleosome. Inset: activity of ASH1L SET at 1 μ M enzyme and 0.8 μ M nucleosome. (C) Activity of ASH1L SET-BAH construct with various native and recombinant nucleosome substrates at 0.2 or 0.4 μ M concentration. (D) Sequence alignment of the autoinhibitory loop of human ASH1L with related KMTases. Sites selected for mutagenesis are shown with arrows. (E) Location of residues selected for mutagenesis (magenta). (F) Activity of ASH1L SET-BAH WT and mutant proteins on chicken mono/di nucleosomes using 0.25 μ M ASH1L and 1.7 μ M chicken mono/di nucleosomes.

ASH1L SET-BAH requires native nucleosome substrate for optimal enzymatic activity

Previous studies have shown that ASH1L requires nucleosomes as substrate [93,127], and a recent study by Eram et al. showed enhanced activity of the ASH1L SET domain on native chicken nucleosomes compared to recombinant, reconstituted nucleosomes [160]. We tested whether our longer, most active ASH1L SET-BAH construct also has enhanced activity on native versus recombinant nucleosomes. We found that, indeed, SET-BAH had much higher activity on native nucleosomes (**Figure 2.3C**). Interestingly, among the native nucleosomes we tested, ASH1L SET-BAH exhibited nearly twofold higher activity on chicken mono/di-nucleosomes than HeLa nucleosomes and chicken oligonucleosomes at 0.2 μ M nucleosome concentration (**Figure 2.3C**). Taken together with the results of Eram et al., our studies indicate that ASH1L exhibits greater activity on native nucleosomes, and this may be partly due to recognition of covalent modifications on native nucleosomes through the chromatin-interacting domains of ASH1L.

Mutations of non-conserved residues in autoinhibitory loop have significant effects on ASH1L SET domain activity

A previous study suggested that the autoinhibitory loop of ASH1L regulates its KMTase activity, likely by physically blocking access to the active site [127]. On the other hand, our results indicated that the autoinhibitory and SET-I loops surrounding the active site of the SET domain experience significant dynamics in solution. In addition, sequence analysis of SET domains related to human ASH1L indicate that residues in the autoinhibitory loop are not conserved (**Figure 2.3D**). We aimed to dissect the contribution of different residues in the autoinhibitory loop to the enzymatic activity of ASH1L by making a series of point mutations based on our crystal structure of the SET domain (**Figure 2.3E**). We explored rather severe mutations with the overall goal to affect the conformation of the autoinhibitory loop while

avoiding global disruption of the SET domain.

In an attempt to destabilize the autoinhibitory loop, we mutated His2258 to Asp to introduce electrostatic repulsion with two adjacent acidic residues, Glu2225 and Asp2254 (**Figure 2.3E**). Interestingly, we found that the H2258D mutant exhibited a modest ~30% decrease in activity (**Figure 2.3F**), suggesting that electrostatic interactions involving His2258 play a minor role in regulating enzyme activity. Next we investigated the neighboring residue Ser2259, whose sidechain points toward the *S*-methyl group of SAM cofactor. Crystallographic studies of the related SETD2 KMTase suggested that the amino acid occupying this position could play a key role in flipping the autoinhibitory loop from a closed to open conformation [129]. To investigate the regulatory potential of Ser2259 we mutated it to Met to enhance hydrophobic contacts with core SET and SET-I. We found that the S2259M mutation strongly decreased ASH1L activity by ~90% (**Figure 2.3F**), consistent with an important regulatory role for Ser2259.

In the middle of the autoinhibitory loop, Val2262 forms hydrophobic contacts with Phe2179 and Met2183 in SET-I. To explore the role of these contacts, we mutated Val2262 to Asp. We found that the V2262D mutation caused nearly no effect on enzymatic activity compared to wild-type ASH1L (**Figure 2.3F**). This result suggests that the hydrophobic contacts made by Val2262 are not critical for enzymatic activity. The result is also consistent with the dynamic nature of the autoinhibitory loop, which may contact SET-I in only a subset of its heterogeneous conformations.

At the end of the autoinhibitory loop, we tested whether we could stabilize the autoinhibited form of the SET domain by mutating Lys2264 to Leu and thereby enhance hydrophobic contacts between the autoinhibitory loop and SET-I. The K2264L mutation

decreased activity by ~50% (**Figure 2.3F**), which is consistent with a need for polar residues at this position based on sequence alignment to related methyltransferases (**Figure 2.3D**). Finally, it was previously reported that mutation of the solvent-exposed Gln2265 to Ala results in enhancement of ASH1L activity, likely through destabilization of the autoinhibitory loop [127]. To further explore this finding we introduced the Q2265A mutation into the SET-BAH construct and found that indeed its activity is increased by ~50% compared to wild type (**Figure 2.3F**). Altogether, we found that despite their low-level conservation, the mutated autoinhibitory loop residues confer significant regulatory control over ASH1L enzyme activity. Our results suggest that the ASH1L autoinhibitory loop is a precisely tuned structural feature with a more complex role than simply blocking the active site.

NMR studies correlate the degree of structural perturbation to enzyme activity

We used NMR to assess structural perturbations to the ASH1L SET domain caused by the autoinhibitory loop mutations. For these studies we selected four mutants: S2259M, V2262D, K2264L, and Q2265A. We collected ^{15}N - ^1H TROSY spectra for all four variants and mapped the chemical shift perturbations caused by the mutations onto the ASH1L crystal structure (**Figure 2.4A-D**). We found that both the S2259M and K2264L mutations caused perturbations to an area surrounding the dynamic autoinhibitory loop region of the protein (**Figure 2.4A and B**). The S2259M mutation caused chemical shift perturbations mostly in β -sheets in the SET-I region and core SET region that contact the autoinhibitory loop (**Figure 2.4A**). The K2264L mutation caused a greater number of perturbations in a complete shell surrounding the autoinhibitory loop (**Figure 2.4B**). In addition to the β -sheets affected by S2259M, the perturbed shell includes the SET-I helix and the C-terminal tail of the post-SET region, which are located above and below the autoinhibitory loop, respectively. The widespread chemical shift perturbations caused by

single amino acid substitutions in the autoinhibitory loop suggest regulatory cross-talk between the autoinhibitory loop and the rest of the SET domain.

Interestingly, we found that the Q2265A mutation, which was previously reported to destabilize the substrate-unbound conformation of the autoinhibitory loop [127], led to very few chemical shift perturbations on the NMR spectrum of ASH1L (**Figure 2.4C**). Chemical shift perturbations were limited mostly to a small local area immediately adjacent to the Q2265A mutation, including residues in the post-SET subdomain. Finally, the V2262D mutation, which had nearly no effect on enzyme activity, also had a relatively small effect on the structure of the SET domain as determined by NMR chemical shift perturbations (**Figure 2.4D**). Notably, we did not observe an increase in the total number of signals observed in the ^{15}N - ^1H TROSY spectra for any of the mutants. Therefore, none of these mutations significantly alter the intermediate exchange dynamics in the autoinhibitory and SET-I loops.

In summary, our NMR analysis showed that mutations leading to large decreases in enzyme activity (S2259M and K2264L) caused significant structural perturbations throughout the SET domain. In contrast, mutations that had no effect or caused enhanced activity (V2262D and Q2265A) caused only small perturbations to the SET domain as measured by chemical shift perturbations (**Figure 2.4A-E**). These results suggest that the proper structure and dynamic properties of the ASH1L autoinhibitory loop are required for enzyme activity.

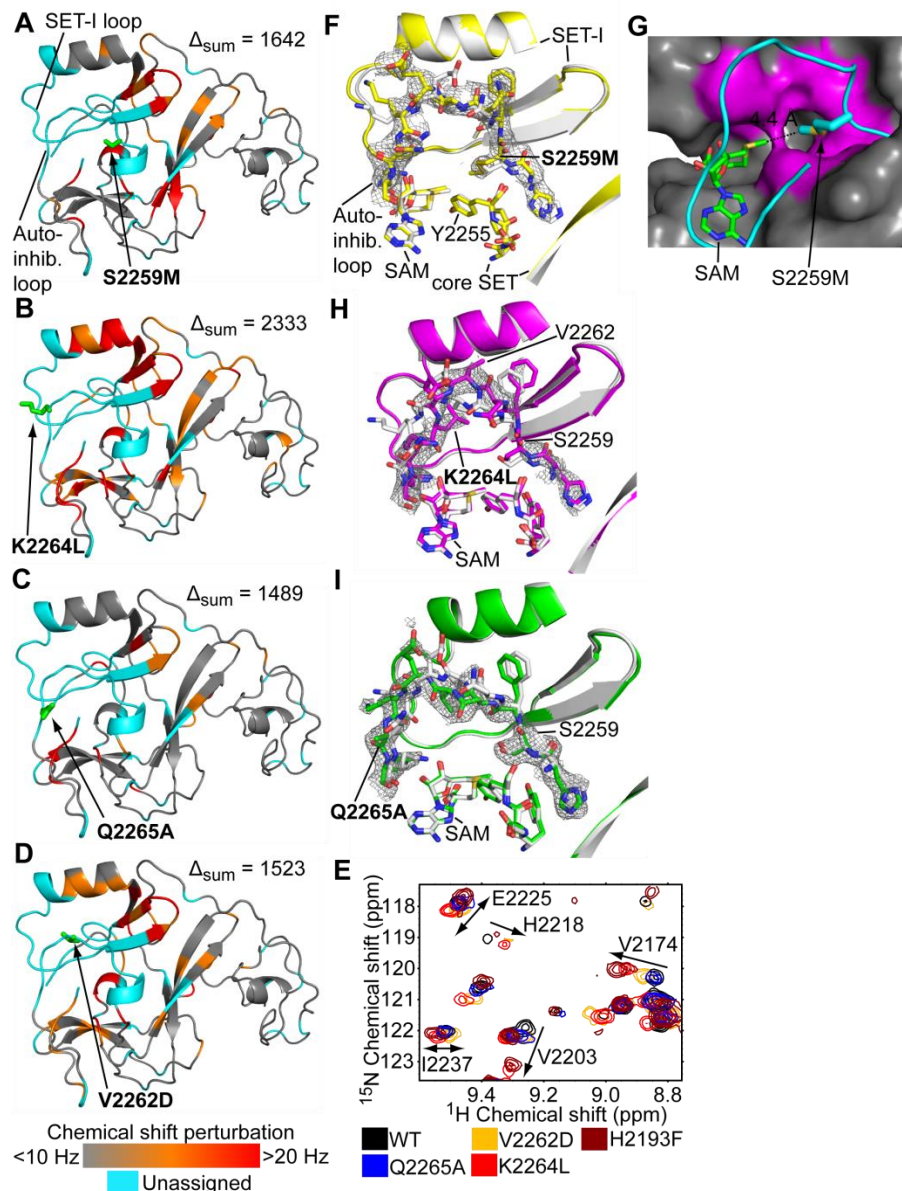


Figure 2.4. Effects of autoinhibitory loop mutations in ASH1L. (A-D) Chemical shift perturbations (Δ , see Methods) of backbone amides in TROSY spectra induced by S2259M (A), K2264L (B), Q2265A (C), and V2262D (D) mutations mapped on the crystal structure of wild-type ASH1L SET domain. Residues perturbed by $\Delta < 10$ Hz, grey; $10 \leq \Delta < 20$ Hz, orange; $\Delta \geq 20$ Hz, red; unassigned, cyan. Mutated residues are labeled and shown in green. The sum of all chemical shift perturbations (Δ_{sum}) is shown for each mutant. (E) A region of the TROSY spectra showing greater chemical shift perturbations caused by the K2264L and H2193F mutants. (F), (H), (I) Comparison of autoinhibitory loop conformation between the mutants (S2259M in yellow (F), K2264L in magenta (H), Q2265A in green (I)) and wild type ASH1L (grey), with 2mFo-DFc maps for the mutants contoured at 0.8σ . (G) Mutant Met2259 partially occupies substrate lysine binding channel. Post-SET subdomain depicted as cartoon, all other regions depicted as surface. Residues His2193, Tyr2194, Cys2195, Leu2196, Tyr2253, and Tyr2255 that form the substrate lysine binding channel are highlighted in magenta.

Mutations affect structure and mobility of the ASH1L autoinhibitory loop

Next we determined the crystal structures of the ASH1L SET domain mutants that caused the largest changes in activity: S2259M, K2264L, and Q2265A. After refinement of these structures, the R_{free} statistic remained high (**Table 2.1**), primarily due to the more flexible AWS subdomain. Compared to the WT structure, the S2259M structure showed modest changes to autoinhibitory loop sidechains (**Figure 2.4F**). However, the sidechain of Met2259 in the mutant protrudes into a channel bordered by the backbone atoms of His2193, Tyr2194, and Cys2195 and the sidechains of Leu2196, Tyr2253, and Tyr2255, which is the putative substrate lysine binding site (**Figure 2.4F, G**). The Met2259 sidechain methyl group is a close 4.4 Å distance from the *S*-methyl group of SAM. These data indicate that the profound reduction in activity of the S2259M mutant (**Figure 2.3F**) is likely caused by stabilization of the inactive conformation of the SET domain via interaction of the methionine sidechain with the substrate lysine binding channel. Therefore, our findings suggest that Ser2259 is an important gatekeeper residue that partially occupies the lysine binding channel in the autoinhibited conformation of the SET domain.

In the K2264L structure, we observed intra-loop hydrophobic contacts between Leu2264 and Val2262 (**Figure 2.4H**), rather than contacts with SET-I that we had designed to stabilize the loop. The new intra-loop interaction between Leu2264 and Val2262 distorts the loop and disrupts the interaction between the autoinhibitory loop and SET-I subdomain. These results suggest that the ~50% reduced activity of the K2264L mutant is caused by perturbations to the structure of the autoinhibitory loop. Interestingly, we observed very high B factors for the K2264L autoinhibitory loop (**Table 2.1**), indicating that despite new contacts between Leu2264 and Val2262 the loop has enhanced dynamics relative to WT ASH1L. These observations further support the conclusion that disruption of the structure and dynamic properties of the

autoinhibitory loop is detrimental to enzyme activity.

The crystal structure of the hyperactive Q2265A mutant shows only minor differences in the conformation of the autoinhibitory loop as compared to wild-type protein (**Figure 2.4I**), in agreement with the minimal effect of the Q2265A mutation on the NMR spectrum. In addition, crystallographic B factors for the Q2265A autoinhibitory loop are comparable to those of wild-type (**Table 2.1**). Although possible, we did not observe evidence that different crystal packing of the Q2265A mutant artificially constrained the loop's mobility. Therefore, our Q2265A structure suggests that enhanced activity of the Q2265A mutant may not be correlated with disordering of the autoinhibitory loop as previously suggested [127], but a co-crystal structure of ASH1L with nucleosome would be required to fully explain the effect of the Q2265A mutation. Finally, for all the mutants we did not find evidence that altered SAM binding contributed to changes in enzymatic activity, as there were no significant differences in SAM orientation or B factor relative to the structure's average B factor (**Table 2.1**).

Table 2.1. Crystallographic data collection and refinement statistics for ASH1L WT and mutants.

	ASH1L WT	ASH1L S2259M	ASH1L H2193F	ASH1L Q2265A	ASH1L K2264L
PDB code	4YNM	4YNP	4YPE	4YPA	4YPU
Data Collection					
Space group	P3 ₂ 21	P3 ₂ 21	P3 ₂ 21	P1	P3 ₂ 21
Cell dimensions					
a, b, c (Å)	59.1, 59.1, 231.0	59.3, 59.3, 233.9	58.8, 58.8, 232.2	53.7, 61.8, 73.2	59.1, 59.1, 226.0
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	91.6, 93.8, 90.5	90, 90, 120
No. protein molecules in asymmetric unit	2	2	2	4	2
Resolution (Å)	50-2.20 (2.24-2.20)	50-2.90 (2.95-2.90)	50-2.20 (2.24-2.20)	50-2.30 (2.34-2.30)	50-2.60 (2.64-2.60)
R_{merge} (%)	9.5 (59.0)	8.8 (39.3)	7.3 (38.5)	6.0 (25.5)	8.8 (48.3)
R_{meas} (%)	10.1 (62.6)	9.4 (41.7)	8.0 (41.6)	8.6 (36.2)	9.3 (51.0)
CC_{1/2} in outer shell	0.87	0.93	0.92	0.85	0.94
I/σI	33.7 (3.5)	25.2 (4.4)	24.4 (4.1)	15.5 (2.8)	42.2 (6.4)
Completeness (%)	99.9 (100.0)	96.1 (96.0)	86.1 (79.5)	96.0 (97.3)	99.4 (98.2)
Redundancy	9.1 (9.1)	7.7 (7.4)	5.2 (5.7)	2.0 (2.0)	9.1 (9.0)
Refinement					
Resolution (Å)	46.82-2.20	47.00-2.90	42.56-2.20	73.04-2.30	75.35-2.60
No. of reflections	23,762	10,352	20,191	38,087	14,072
No. of atoms	3497	3364	3572	7254	3373
Protein	3333	3304	3405	6920	3277
SAM	54	54	54	108	54
Zn ²⁺	6	6	6	12	6
Water	104	0	107	214	36
R_{work}/R_{free}	24.2/27.8	26.2/31.9	21.7/26.3	24.9/30.3	23.8/30.5
Average B-factor	42.4	54.7	32.2	38.0	53.1
Autoinhibitory loop	66.8	71.8	65.5	57.0	82.3
SAM	35.1	54.8	28.3	25.2	52.8
RMSDs					
Bond length (Å)	0.017	0.006	0.022	0.018	0.013
Bond angles (°)	1.769	0.973	1.940	1.645	1.479
Ramachandran favored (%)	93.3	90.0	92.3	91.3	89.7
Ramachandran allowed (%)	5.7	9.0	7.5	7.5	8.5
Ramachandran outliers (%)	1.0	1.0	0.2	1.2	1.8
MolProbity clash score	11.86	4.74	10.7	16.6	7.41

All diffraction data were obtained from a single crystal. Values in parentheses are for highest-resolution shell. CC_{1/2} is the Pearson correlation coefficient of two half data sets as defined by Karplus and Diederichs.[174]

Structural insight into SET-I loop conformational dynamics

We found that the Q2265A mutant crystallized in a different crystal form than the other ASH1L variants (P1 versus P3₂21, respectively) (**Table 2.1**). There are four ASH1L Q2265A molecules per asymmetric unit, and interestingly the Q2265A crystal traps two different conformations of the SET-I and autoinhibitory loops, with each conformation represented by a pair of ASH1L monomers (**Figure 2.5A and B**). The most interesting difference between the two conformations is the two orientations of His2193, with the histidine side chain forming a hydrogen bond either with the 3' hydroxyl group of SAM or with the hydroxyl of Tyr2207 (**Figure 2.5A**). Such conformational exchange is consistent with the WT structure, in which we observe poor electron density for His2193 (**Figure 2.1C**). Thus, the structure of the Q2265A mutant provides snapshots of two different conformations of the SET-I loop. We also observed significant differences in the autoinhibitory loop between the two different ASH1L conformations (**Figure 2.5B**), further emphasizing the conformational dynamics of this loop.

We tested whether the Q2265A crystal structure could be used to corroborate the dynamics in ASH1L that we observed by NMR. We plotted pairwise distances between C α positions for the two different ASH1L conformations observed in the crystal structure of this mutant. Compared to the rest of the core SET domain, we observed large structural differences for the SET-I and autoinhibitory loops (**Figure 2.5C**). The two molecules in the asymmetric unit of WT ASH1L also show structural differences for these loops, but the differences between molecules in the asymmetric unit are more significant and better resolved in the Q2265A mutant. Interestingly, residues which show these large structural differences are not observed in the NMR spectrum of the SET domain (both WT and Q2265A mutant). Therefore, intermediate dynamics that lead to broadening of NMR signals are consistent with conformational

heterogeneity observed in the crystal structure of Q2265A.

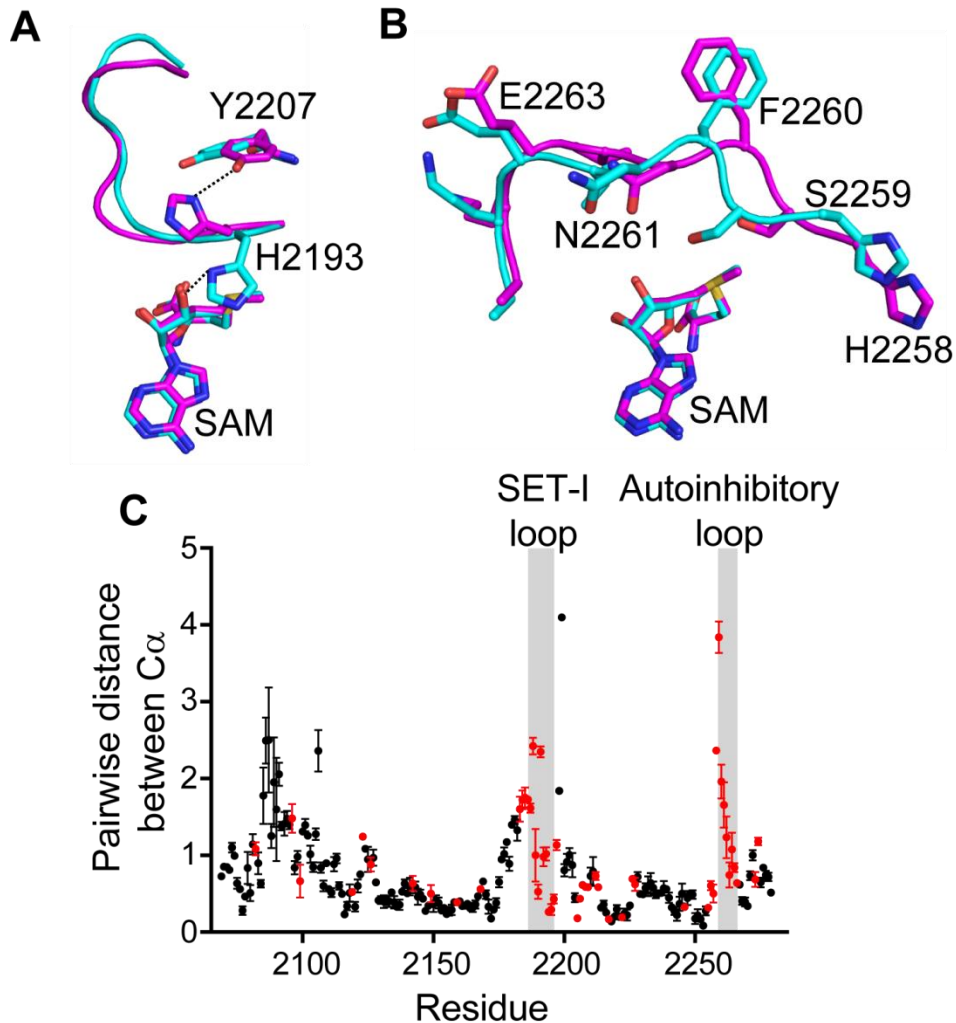


Figure 2.5. Conformational dynamics observed in the crystal structure of Q2265A. (A) Superposition of SET-I loops for two different ASH1L molecules in asymmetric unit. His2193 can form a hydrogen bond either with the 3' hydroxyl group of SAM or with the sidechain of Tyr2207. (B) Superposition of autoinhibitory loops for two different ASH1L molecules in asymmetric unit. (C) Pairwise distances between α carbons for the two ASH1L conformations observed in the Q2265A crystal. SET-I and autoinhibitory loop regions are highlighted by grey fields, residues unassigned in NMR spectra are shown in red.

Conformational exchange in the SET-I loop contributes to ASH1L activity

Analysis of the crystal structure of WT and Q2265A ASH1L indicated that the SET-I loop samples different conformations. We expected that such conformational heterogeneity might be partially regulated by His2193 which we found in two different conformations (**Figure**

2.5A). To probe the role of H2193 we introduced an H2193F mutation to disrupt its potential to form hydrogen bonds. We found that the H2193F mutation decreases activity of SET-BAH by more than 80% (**Figure 2.6A**). NMR studies of ASH1L H2193F showed that the mutation causes widespread chemical shift perturbations extending in a shell surrounding the dynamic SET-I and autoinhibitory loops (**Figure 2.6B**). We determined the crystal structure of H2193F and found a large ~ 8 Å conformational shift in the SET-I loop, with minimal perturbations to other regions when compared to wild-type protein (**Figure 2.6C**). Phe2193 in this mutant forms new hydrophobic contacts with Tyr2255, Val2262, and Ile2279 (**Figure 2.6C and D**). Interestingly, despite the significantly different conformation of the SET-I loop, the conformation of the autoinhibitory loop is unaffected. Altogether, these results demonstrate that conformational dynamics of the SET-I subdomain also play a significant role in modulating catalytic activity of the ASH1L SET domain. Furthermore, the dynamics of the SET-I subdomain occur in concert with those of the autoinhibitory loop (**Figure 2.6E**) to affect regulation of enzymatic activity.

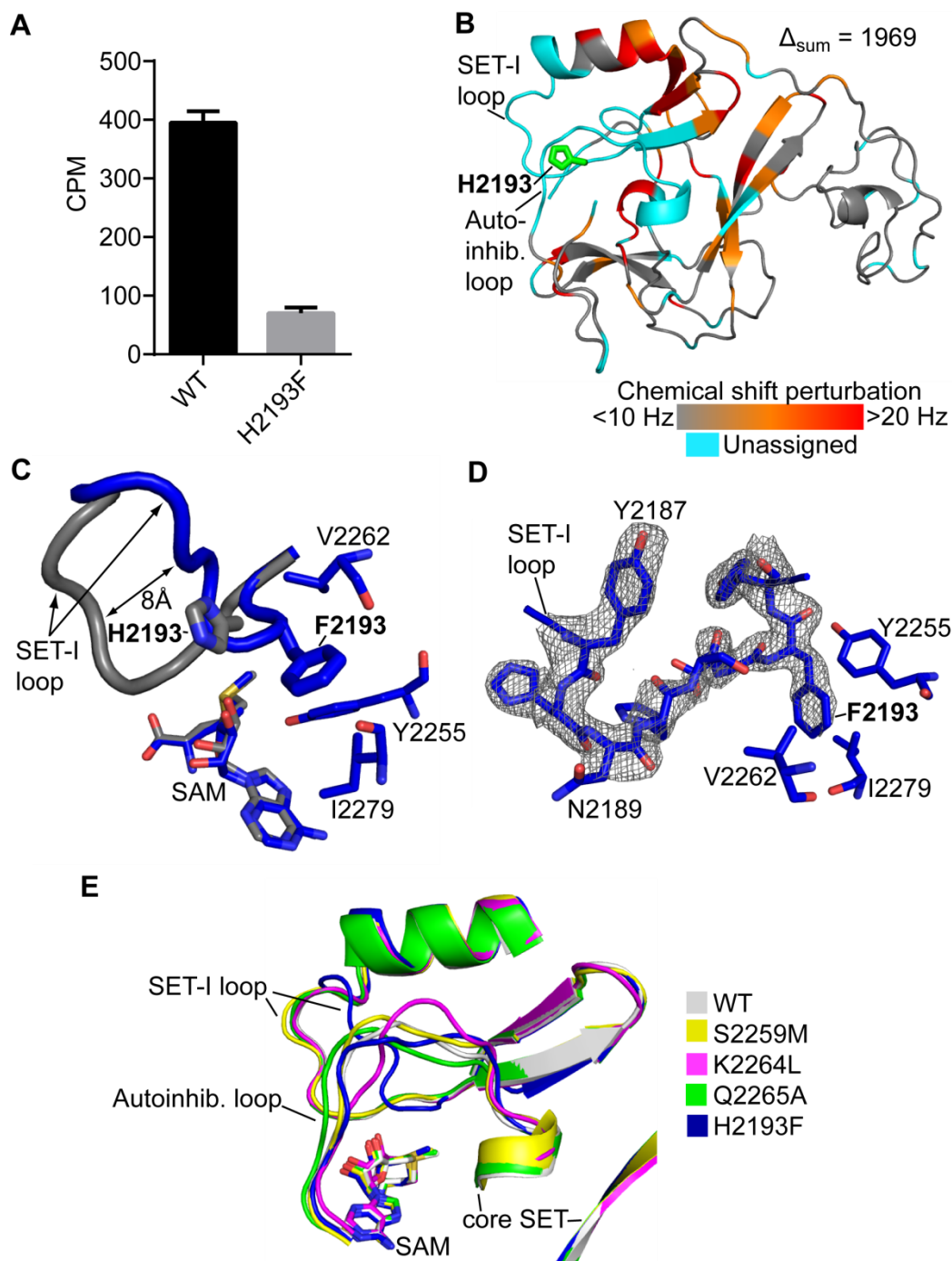


Figure 2.6. Structural perturbation in SET-I loop induced by the H2193F mutation. **(A)** Enzymatic activity of wild-type and H2193F ASH1L with 0.25 μM ASH1L and 1.7 μM chicken mono/di nucleosome substrate. **(B)** Chemical shift perturbations in TROSY spectra induced by H2193F, color coding as in Figure 4. **(C)** Superposition of the SET-I loops of H2193F (blue) and wild-type (grey). Residues involved in hydrophobic contacts with Phe2193 are shown and labeled. **(D)** The SET-I loop of H2193F mutant with mFo-DFc omit map contoured at 2.5σ . Residues forming hydrophobic contacts to Phe2193 are shown. **(E)** Superposition of crystal structures of WT ASH1L and four mutants.

Discussion

SET domain-containing KMTases are an important class of enzymes that constitute attractive targets for new therapeutics [175], but many structural aspects of SET domains remain poorly understood. The autoinhibitory loop is a striking feature observed in the crystal structure of the ASH1L SET domain. However, whether the loop functions as a simple swinging gate to allow access to the active site, or instead plays a more active role in substrate binding, is unknown. In this study we provide experimental evidence supporting the latter model. First, point mutations that destabilize the autoinhibitory loop did not increase enzymatic activity, as would be expected if the loop simply blocks access for substrate binding. Second, point mutations in the autoinhibitory loop that cause larger structural perturbations in the SET domain as judged from NMR and crystallography caused a substantial loss of enzyme activity, while point mutations that had smaller effects on the structure caused no change or even resulted in an increase in enzymatic activity compared to wild-type ASH1L. This suggests that the proper structure of the autoinhibitory loop is necessary for substrate binding, and the loop itself represents an important regulatory feature of the SET domain.

We discovered that in addition to the autoinhibitory loop, the SET-I loop plays an important role in ASH1L SET domain function. The dynamics of this conformationally heterogeneous loop are partly regulated by His2193, a residue in the SET-I loop that adopts two different conformations and mutation of which significantly impairs enzymatic activity. In other SET domain-containing proteins, the SET-I region is poorly conserved and forms key contacts with histone substrates [132,176–179], indicating that it likely functions in determining substrate specificity among different KMTases [180]. In the structure of the related Pr-Set7 KMTase bound to a histone H4 fragment, the SET-I loop forms extensive contacts with the H4 peptide

[132,176]. Our results indicate that SET-I is also important to the enzymatic activity of ASH1L, likely through forming analogous contacts with histone H3 residues.

Crystallographic and NMR studies of the SET domain and multiple mutants showed that the autoinhibitory loop and SET-I loop experience conformational dynamics occurring on the microsecond-millisecond timescale. Importantly, this conformational heterogeneity experienced in the absence of bound substrate does not appear to include an open, substrate-accessible form of the SET domain. The fully open conformation of ASH1L would require a much larger conformational change than any we observed and most likely such conformation can be achieved in the presence of nucleosome. However, the conformational dynamics we observed probably facilitate the major structural change leading to opening of the substrate binding site. Indeed, dynamic loops important for nucleosome binding have been characterized in a variety of chromatin-interacting domains, such as the Sir3 BAH domain [181], the PCAF/GCN5 histone acetyltransferase [182], and bromodomains [183]. For example, multiple crystal structures of the Sir3 BAH domain showed that residues in two flexible loops are completely disordered in the free BAH domain structure [184,185], but these residues become ordered or partially ordered upon binding to nucleosome [181].

We also found that *in vitro* ASH1L is much more active on native nucleosomes than recombinant nucleosomes. Furthermore, ASH1L constructs including chromatin binding domains have significantly enhanced activity compared to the isolated SET domain. Together these findings suggest a model whereby chromatin-binding domains of ASH1L recognize posttranslational modifications on native nucleosomes and facilitate catalytic activity of the SET domain via substrate recruitment. Interestingly, the results of Eram et al. showing enhanced activity of the isolated ASH1L SET domain on native nucleosomes[160] suggest that the SET

domain itself may also have increased affinity for native nucleosomes. In previous studies, ASH1L has been reported to methylate both H3K4 and H3K36 [44,93,127,157–160]. While the precise targets of ASH1L *in vivo* are still under investigation, such promiscuity suggests that ASH1L specificity depends on the nature of the substrate, similar to the NSD family [79]. The number of methyl groups transferred by ASH1L is an additional unanswered question with important implications for control of gene expression. In an *in vitro* enzyme assay, ASH1L can mono- and dimethylate, but not trimethylate, H3K36 [160]. On the other hand, there is *in vivo* evidence in mouse that ASH1L may trimethylate H3K36 [93]. The autoinhibitory loop may function to regulate the number of methyl groups transferred, because the Q2265A mutation leads to H3K36me₃ in addition to the original H3K36me₂ product [127].

In summary, we discovered concerted dynamics in two loops of the ASH1L SET domain and showed that these loops regulate ASH1L enzymatic activity. Conformational dynamics in the substrate binding region of ASH1L may create transient pockets into which small molecule ligands could bind. Thus, it may be feasible to exploit the conformational dynamics to design small molecule inhibitors of ASH1L.

Chapter 3. Development of small molecule inhibitors for the ASH1L histone methyltransferase*

Abstract

The ASH1L histone methyltransferase is overexpressed in multiple cancers and activates oncogenic *HOX* genes. Small molecule inhibitors of ASH1L would be excellent tools to further investigate the role of ASH1L in cancer and could be the precursors of novel therapeutics; however, ASH1L inhibitors have not been reported to date. The ASH1L SET domain contains a conformationally heterogeneous autoinhibitory loop that blocks access of histone substrate to the enzyme active site. We used fragment based screening to identify a fragment-like ligand that binds to the autoinhibitory loop region of ASH1L. Structure-guided optimization of the ligand resulted in over 1000-fold improvement in binding affinity compared to the fragment screening hit. The optimized compound SK-1 inhibits ASH1L enzymatic activity with an IC_{50} value of $4.6 \pm 1.9 \mu\text{M}$ and binds to ASH1L with K_d of $1.7 \mu\text{M}$. Crystallographic and NMR studies indicate that SK-1 and related compounds inhibit ASH1L activity by stabilizing the autoinhibited conformation of the SET domain. This study demonstrates that fragment screening coupled with structure-based drug design can be used to develop potent and specific inhibitors for challenging epigenetic targets.

*Rogawski DS, et al. *ASH1L inhibitors block leukemogenesis through downregulation of Hox gene expression*. Manuscript in preparation.

Introduction

ASH1L is a histone 3, lysine 36 (H3K36)-specific methyltransferase (KMTase) that is an emerging cancer drug target due to its overexpression in multiple cancers and role in activating oncogene expression. Twenty-seven percent of aggressive, basal-like breast cancers have high-level copy number amplifications of the ASH1L gene [95]. In addition, high levels of ASH1L mRNA are associated with shorter survival in breast cancer patients [95]. In thyroid cancer, ASH1L is overexpressed in tumor-specific truncated forms and is downregulated by a tumor suppressor microRNA [94]. ASH1L activates genes in the *HOX-A*, *-B*, *-C*, and *-D* clusters [15,93,96,159], which are overexpressed in multiple cancers and correlated with metastasis and aggressive disease [89]. Notably, ASH1L activates *HOXA9* and its collaborator *MEIS1* [15], which are oncogenes in leukemia [168].

Small molecule inhibitors of ASH1L KMTase activity would be invaluable tools to investigate the role of ASH1L in cancer, but such compounds have not been reported to date. To develop potent and specific inhibitors of ASH1L, we recalled our and others' biophysical studies of the catalytic SET domain of ASH1L, which showed that an autoinhibitory loop blocks access of histone substrate to the active site [127,134]. Interestingly, the autoinhibitory loop and surrounding region experience conformational heterogeneity in the absence of nucleosome substrate [134]. We hypothesized that conformational dynamics in the autoinhibitory loop region could create transient pockets into which small molecule ligands could bind. Moreover, these ligands could possibly function as allosteric inhibitors by preferentially interacting with and stabilizing the autoinhibited form of ASH1L SET. Because the autoinhibitory loop is poorly conserved in primary sequence and structure, inhibitors that bind to the ASH1L autoinhibitory loop region would be predicted to have excellent selectivity over related KMTases.

As there are no known small molecule ligands of the ASH1L autoinhibitory loop region, we chose to take a fragment-based drug discovery (FBDD) approach [186], which would allow us to probe ASH1L SET for potential small molecule binding sites and then construct new ligands specific to ASH1L. FBDD is further advantageous because it requires relatively small compound libraries to sample chemical space, and often results in improved physicochemical and pharmacokinetic properties of lead compounds [187]. Our FBDD strategy identified a ligand that binds to the ASH1L autoinhibitory loop region. We then used information from NMR and crystallographic studies to optimize the fragment-like ligand into a potent and specific ASH1L inhibitor. This study identifies the first ASH1L inhibitor and provides the first example of successfully applying FBDD to KMTases.

Results

FBDD identifies a novel ligand of ASH1L SET domain

We conducted fragment screening of the ASH1L SET domain using ^{15}N - ^1H TROSY NMR, a highly sensitive method for detecting ligand binding to any site on ASH1L. We screened 1500 fragment-like compounds ($\text{MW} < 300$) in the presence of twofold excess SAM cofactor over ASH1L to find ligands that do not compete with SAM. This screen produced two hits that caused perturbations to the same TROSY peaks, suggesting the two ligands bind to the same site. The low hit rate indicates that ASH1L SET does not generally bind to fragment-like compounds, but only to a privileged class of compounds that form specific interactions with the SET domain. The stronger of the two hits, thioamide **1**, causes chemical shift perturbations or broadening of multiple TROSY peaks (**Figure 3.1A**). We determined from NMR titrations that **1** binds to ASH1L SET with affinity weaker than 1 mM (data not shown).

To identify the binding site of **1**, we took advantage of our previously reported assignment of the ASH1L TROSY spectrum [134]. Notably, we do not observe TROSY peaks corresponding to the autoinhibitory loop region of ASH1L due to conformational heterogeneity and peak broadening. We mapped chemical shift perturbations caused by binding of **1** onto the ASH1L SET crystal structure (**Figure 3.1B**). We found that the perturbations mapped to a shell surrounding the unassigned autoinhibitory loop region. This result suggests that **1** binds to the conformationally dynamic autoinhibitory loop region, causing observable perturbations to residues surrounding this region. To confirm that **1** binds to the autoinhibitory loop region, we mutated Ser2259 in the autoinhibitory loop to methionine (**Figure 3.1C**). Introduction of the bulky methionine side chain in the autoinhibitory loop abrogated binding of **1** while causing minor changes to the overall TROSY spectrum (**Figure 3.1D**), indicating that **1** binds to the autoinhibitory loop region of ASH1L.

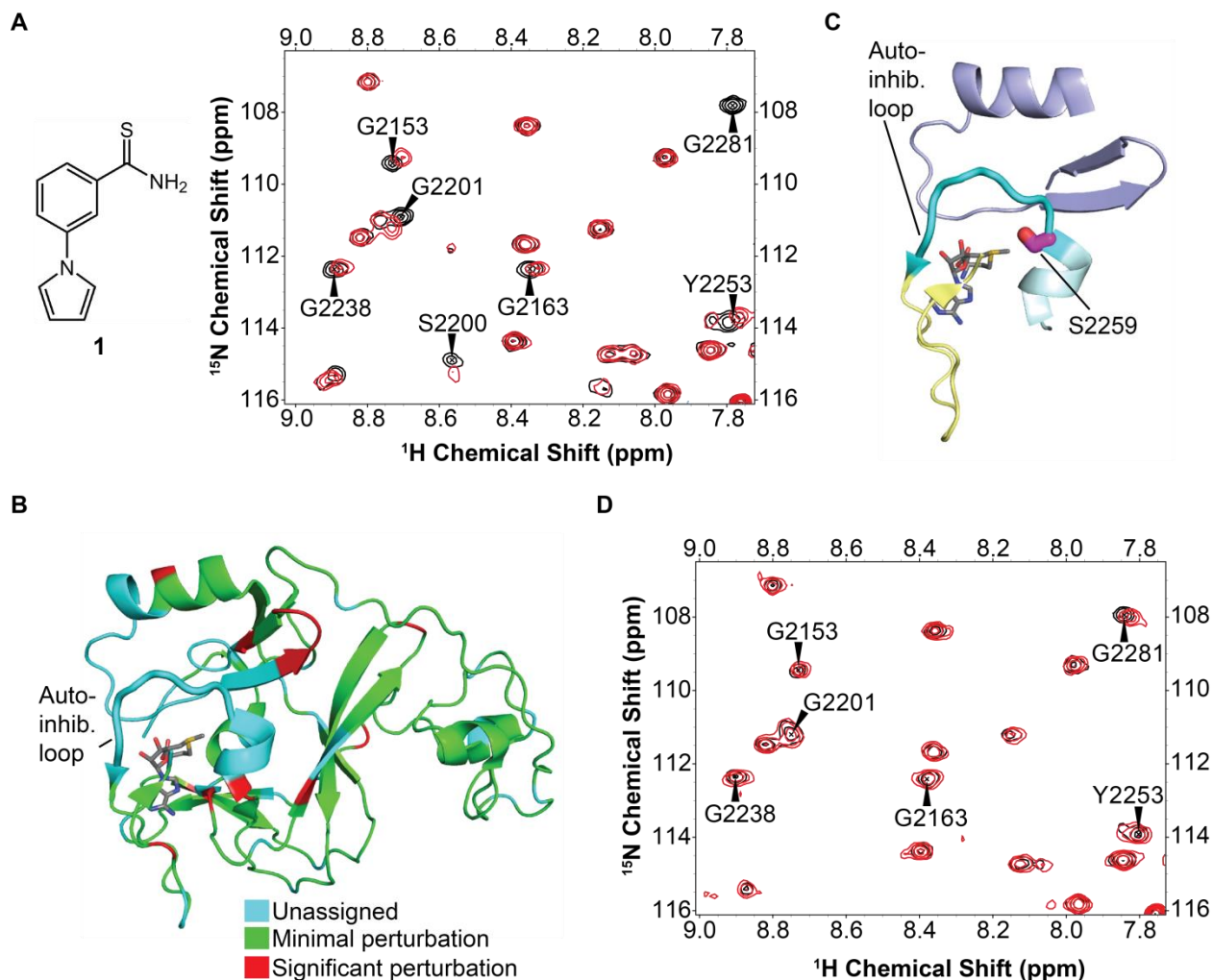


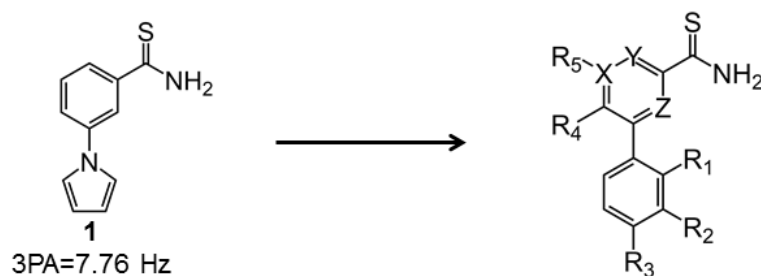
Figure 3.1. Identification of a fragment-like ligand of the ASH1L SET domain. (A) A region of the ^{15}N - ^1H TROSY spectrum of ASH1L SET (100 μM) with no compound (black) and 1 mM **1** (red). Structure of **1** shown at left. (B) Perturbations to the TROSY spectrum induced by binding of **1** mapped onto the SET domain crystal structure (PDB code 4YNM). (C) Location of Ser2259 in the autoinhibitory loop (cyan). SET-I subdomain colored light purple, core SET subdomain colored light blue, and post-SET subdomain colored pale yellow. (D) S2259M mutation prevents binding of **1** to ASH1L SET, as shown by lack of chemical shift perturbations in the presence of 1 mM **1**.

Medicinal chemistry optimization of fragment screening hit 1

To improve the potency of **1**, we undertook a systematic medicinal chemistry campaign, synthesizing over 130 compounds total (Appendix). To compare weak compounds with affinities in the high micromolar to millimolar range, we developed an NMR-based estimation of affinity called 3 Peak Affinity (3PA). The 3PA is based on the premise that for NMR peaks showing a

fast exchange binding mode and at a compound concentration within 2-3-fold of the K_d , the relative affinity of different compounds correlates with the size of the induced chemical shift perturbations. The 3PA is calculated by taking the average of the chemical shift perturbations in hertz at 3 TROSY peaks (G2163, K2228, and I2278) in the presence of 500 μ M compound (see Methods for details). More potent compounds induce larger chemical shift perturbations and receive a higher 3PA score.

We began optimization of **1** by exploring changes to the pyrrole group. Substitution of the pyrrole with phenyl in **1a** improved the 3PA to 25.38 Hz compared to 7.78 Hz for **1** (**Table 3.1**). A hydroxyl derivative **1b** had improved solubility and potency with a 3PA of 29.76. Subsequent methylation of the hydroxyphenyl at the R_1 and R_2 positions further improved binding in **1c** and **1d**, with 3PA values of 48.71 and 36.31, respectively. We were able to analyze binding of **1c** to ASH1L by isothermal titration calorimetry (ITC) and found a K_d value of 135 μ M (data not shown). In the context of the hydroxyphenyl (**1b**) we performed nitrogen scanning in the thioamide-conjugated ring (**Table 3.1**). We found that pyridine derivatives **1e** and **1f** showed dramatically reduced binding compared to **1b**, but **1g** had increased potency with a 3PA of 40.45. We next performed methylation of the thioamide-conjugated ring at the R_4 and R_5 positions in **1h** and **1i**, but found that both compounds had decreased potency by NMR compared to **1b**.

Table 3.1. Medicinal chemistry optimization of fragment screening hit **1**.

Compound	R ₁	R ₂	R ₃	R ₄	R ₅	X	Y	Z	3PA (Hz)	K _d (μM)
1a	H	H	H	H	H	C	C	C	25.38	n.d.
1b	H	H	OH	H	H	C	C	C	29.76	n.d.
1c	CH ₃	H	OH	H	H	C	C	C	48.71	135
1d	H	CH ₃	OH	H	H	C	C	C	36.31	n.d.
1e	H	H	OH	H	H	N	C	C	2.07	n.d.
1f	H	H	OH	H	H	C	N	C	no binding	n.d.
1g	H	H	OH	H	H	C	C	N	40.45	n.d.
1h	H	H	OH	CH ₃	H	C	C	C	16.35	n.d.
1i	H	H	OH	H	CH ₃	C	C	C	21.08	n.d.

The SAR thus far indicated that substitutions to the thioamide-conjugated ring were in general not tolerated (**Table 3.1** and Appendix). In contrast, expansion of the pyrrole to indole **2** dramatically improved affinity as evidenced by a slow exchange binding mode observed by NMR. For example, as **2** is titrated into ASH1L SET, we observe two TROSY peaks for G2163, with the relative peak heights changing with increasing amounts of **2** (**Figure 3.2A**). To test whether the more potent **2** could inhibit ASH1L enzymatic activity, we used an *in vitro* enzyme assay with nucleosomes as substrate [134]. Consistent with our hypothesis that ligands of the autoinhibitory loop region should block enzymatic activity, we found that **2** inhibits ASH1L with IC₅₀ value of 76 μM (**Figure 3.2A**). An isomer of **2** with indole attached at the 3 position to the

thioamide-conjugated ring (**3**) had a similar IC_{50} of 100 μM (**Figure 3.2B and C**) and proved a more tractable scaffold for further synthesis.

We next focused our medicinal chemistry efforts on optimization of the indole of **3**. Synthesis of over 50 compounds with this scaffold indicated that substitutions at the 1 and 6 positions of the indole generally improved potency while substitutions at other positions were not tolerated (Appendix). Substitution of the indole at position 1 with hydroxyethyl to yield **4** improved potency fivefold to an IC_{50} of 22 μM . In addition, substitution at position 6 with methylamine **5** improved potency threefold to an IC_{50} of 36 μM . We found that combining these two substituents in **6** had an additive effect, with IC_{50} of 13 μM (**Figure 3.2C**). The K_d of **6** as measured by ITC is 4.6 μM (**Figure 3.2D**).

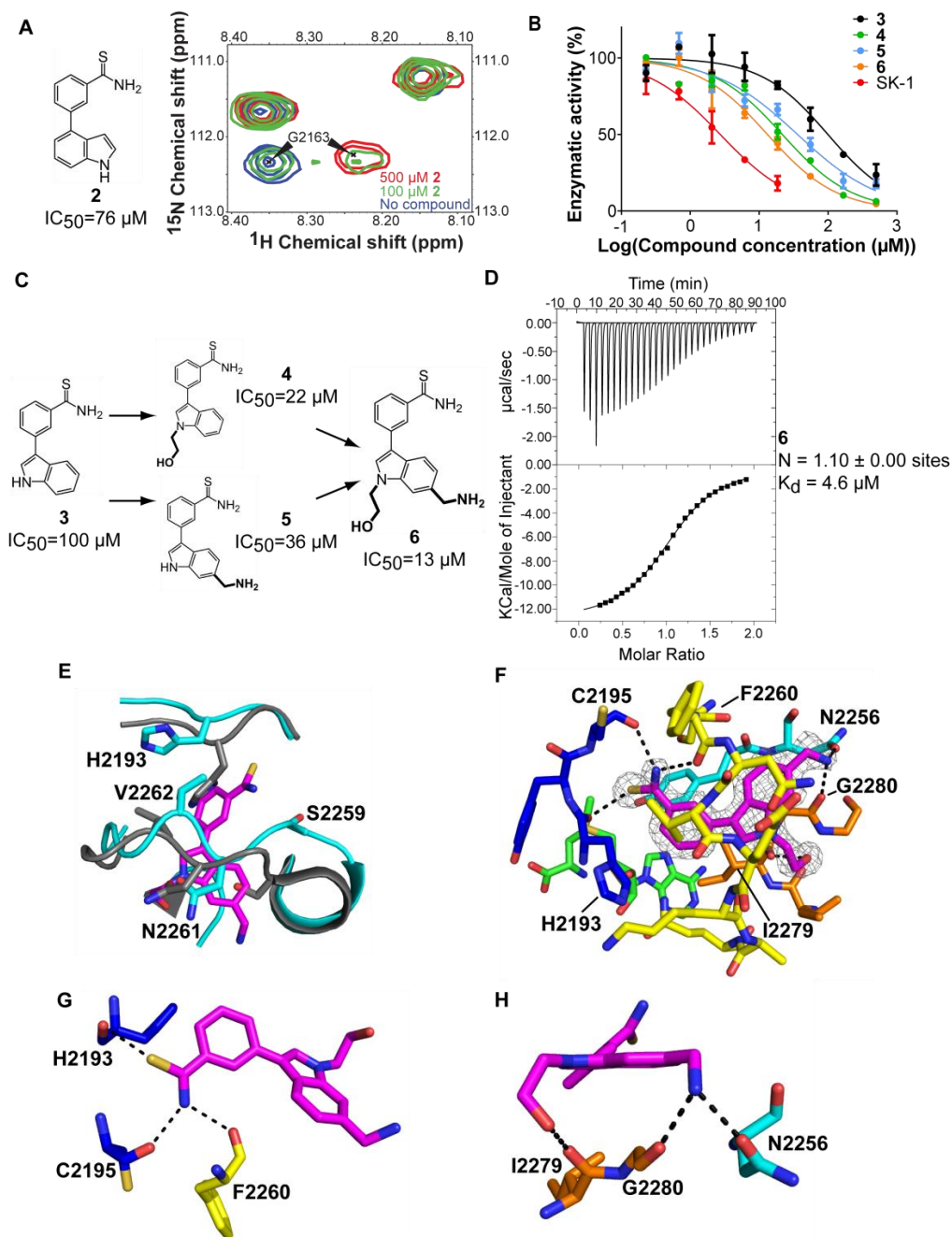


Figure 3.2. Biochemical and structural characterization of ASH1L inhibitors. (A) Compound **2** binds to ASH1L with slow exchange kinetics on the NMR timescale. (B) Comparison of ASH1L inhibitor potency in *in vitro* KMTase assay. (C) Substitutions at the 1 and 6 positions of the indole have an additive effect on inhibitor potency. (D) Binding of **6** to ASH1L by ITC. (E) Crystal structure of **6** (magenta) bound to ASH1L. Inhibitor binding causes a conformational change in the autoinhibitory and SET-I loops. (F) Crystal structure of **6** (magenta) bound to ASH1L with 2Fo-Fc map contoured at 1.0σ . Autoinhibitory loop shown in yellow, SET-I loop in blue, core SET in cyan, post-SET in orange, and SAM cofactor in green. (G) Interactions made by thioamide group in structure of **6** bound to ASH1L. (H) Interactions made by hydroxyethyl and methylamine groups of **6**.

Crystal structure of 6 in complex with ASH1L SET shows details of protein-ligand interaction and stabilization of autoinhibitory loop region

To further characterize binding of bi-substituted **6** and obtain structural information for inhibitor design, we crystallized the ASH1L SET domain in complex with **6**. Crystals of the wild-type SET domain grown in the presence of inhibitors consistently diffract to low resolution (~ 3 Å), precluding detailed structural analysis. However, we found that crystals of **6** in complex with Q2265A mutant ASH1L diffracted to 1.7 Å and showed excellent electron density for the inhibitor. Consistent with our previous NMR mapping and mutagenesis studies, **6** binds to the autoinhibitory loop region of ASH1L, with the autoinhibitory loop wrapping around **6** on three sides (**Figures 3.2E and F**). The loop undergoes a conformational change in which the sidechains of Ser2259 and His2193 rotate away from the binding site to accommodate the inhibitor, while Val2262 and Asn2261 adjust to make van der Waals contacts with the phenyl and indole of **6**, respectively (**Figure 3.2E**). Binding of **6** also induces formation of a β -turn at the C-terminus of the post-SET subdomain, and Ile2279 and Gly2280 within this turn make hydrophobic contacts to the inhibitor (**Figure 3.2F**). The thioamide group of **6** binds deep inside the protein, with the thioamide forming a sulfur-oxygen chalcogen bond [188] to the backbone carbonyl of His2193 (**Figure 3.2G**). The thioamide nitrogen forms hydrogen bonds with the backbone carbonyls of Cys2195 and Phe2260. Closer to the surface of the protein, the hydroxyethyl substituent on the indole forms a hydrogen bond with the backbone carbonyl of Ile2279, while the methylamine forms hydrogen bonds to the Asn2256 sidechain and the Gly2280 backbone carbonyl (**Figure 3.2H**).

Compared to the crystal structure of ASH1L without ligand, the resolution of the crystal structure of ASH1L bound to **6** is significantly higher (1.7 versus 2.3 Å) and the average protein B factor is substantially lower (16.6 ± 10.1 Å² versus 38.3 ± 16.0 Å², mean and standard

deviation) (**Table 3.2**), suggesting that **6** stabilizes dynamics of the ASH1L SET domain. To determine if the autoinhibitory loop region is particularly stabilized by inhibitor binding relative to the rest of the SET domain, we converted raw B factors to normalized B factors, thereby correcting for the large difference in average B factor between the free SET domain and SET domain/inhibitor complex (see Methods). We found that normalized B factors for the autoinhibitory loop and the SET-I loop were significantly lower in the crystal structure of ASH1L in complex with **6** compared to the structure of ASH1L alone (**Figures 3.3A and B**). In addition, by NMR we found that binding of **6** perturbs the majority of TROSY peaks for ASH1L and causes appearance of several new peaks (**Figure 3.3C**). Further inspection of ITC parameters revealed a thermodynamic signature of binding for **6** and related compounds, characterized by a favorable change in enthalpy ($\Delta H = -13.0$ kcal/mol) and an unfavorable entropic component ($-T\Delta S = 5.7$ kcal/mol) (**Figure 3.3D**) likely due to conformational changes of the SET domain upon ligand binding [189]. Together, these results suggest that **6** and related compounds inhibit ASH1L activity by stabilizing the autoinhibited conformation of the SET domain.

Table 3.2. Crystallographic data collection and refinement statistics for crystals of ASH1L Q2265A in complex with inhibitors.^a

	No compound	6	9	SK-1
Data Collection				
Space group	P1	P1	P1	P1
Cell dimensions				
a, b, c (Å)	53.7, 61.8, 73.2	54.1, 62.3, 73.0	53.9, 62.1, 72.9	53.1, 61.3, 71.8
α , β , γ (°)	91.6, 93.8, 90.5	87.8, 85.6, 90.0	92.0, 94.4, 90.0	88.1, 85.7, 89.6
No. protein molecules in asymmetric unit	4	4	4	4
Resolution (Å)	50–2.30 (2.34–2.30)	50–1.70 (1.73–1.70)	50–1.77 (1.80–1.77)	50–1.65 (1.68–1.65)
R_{merge} (%)	6.0 (25.5)	9.6 (59.1)	11.9 (63.4)	8.2 (57.9)
R_{meas} (%)	8.6 (36.2)	11.1 (69.0)	13.8 (74.5)	9.3 (66.3)
CC_{1/2}^b in outer shell	0.85	0.73	0.68	0.79
I/σI	15.5 (2.8)	14.4 (2.1)	15.9 (2.1)	26.0 (2.1)
Completeness (%)	96.0 (97.3)	96.5 (94.1)	97.0 (95.2)	97.1 (95.2)
Redundancy	2.0 (2.0)	3.9 (3.7)	3.9 (3.7)	4.2 (4.2)
Refinement				
Resolution (Å)	73.04–2.30	40.86–1.69	36.05–1.78	40.13–1.63
No. of reflections	38087	96238	83517	103276
No. of atoms	7254	8349	8226	8098
Protein	6920	7027	7008	7108
SAM	108	108	108	108
Zn ²⁺	12	12	12	12
Ligand	--	92	96	100
Water	214	1110	1002	770
R_{work}/R_{free}	24.9/30.3	19.0/24.1	22.2/28.5	17.8/22.1
Average B-factor	38.0	18.1	22.6	28.4
Protein	38.3	16.6	21.4	27.8
SAM	25.2	8.9	10.4	17.6
Zn ²⁺	34.0	12.6	19.3	25.0
Ligand	--	15.5	20.2	25.0
Water	36.2	29.0	32.3	37.1
RMSDs				
Bond length (Å)	0.018	0.030	0.023	0.026
Bond angles (°)	1.645	2.228	2.143	2.411
Ramachandran favored (%)	91.3	93.9	92.1	95.5
Ramachandran allowed (%)	7.5	5.1	6.9	3.3
Ramachandran outliers (%)	1.2	1.0	1.0	1.2
MolProbity clash score	16.6	10.96	14.8	11.32

^aAll diffraction data were obtained from a single crystal. Values in parentheses are for the highest-resolution shell. ^bCC1/2 is the Pearson correlation coefficient of two half-data sets as defined by Karplus and Diederichs.[174]

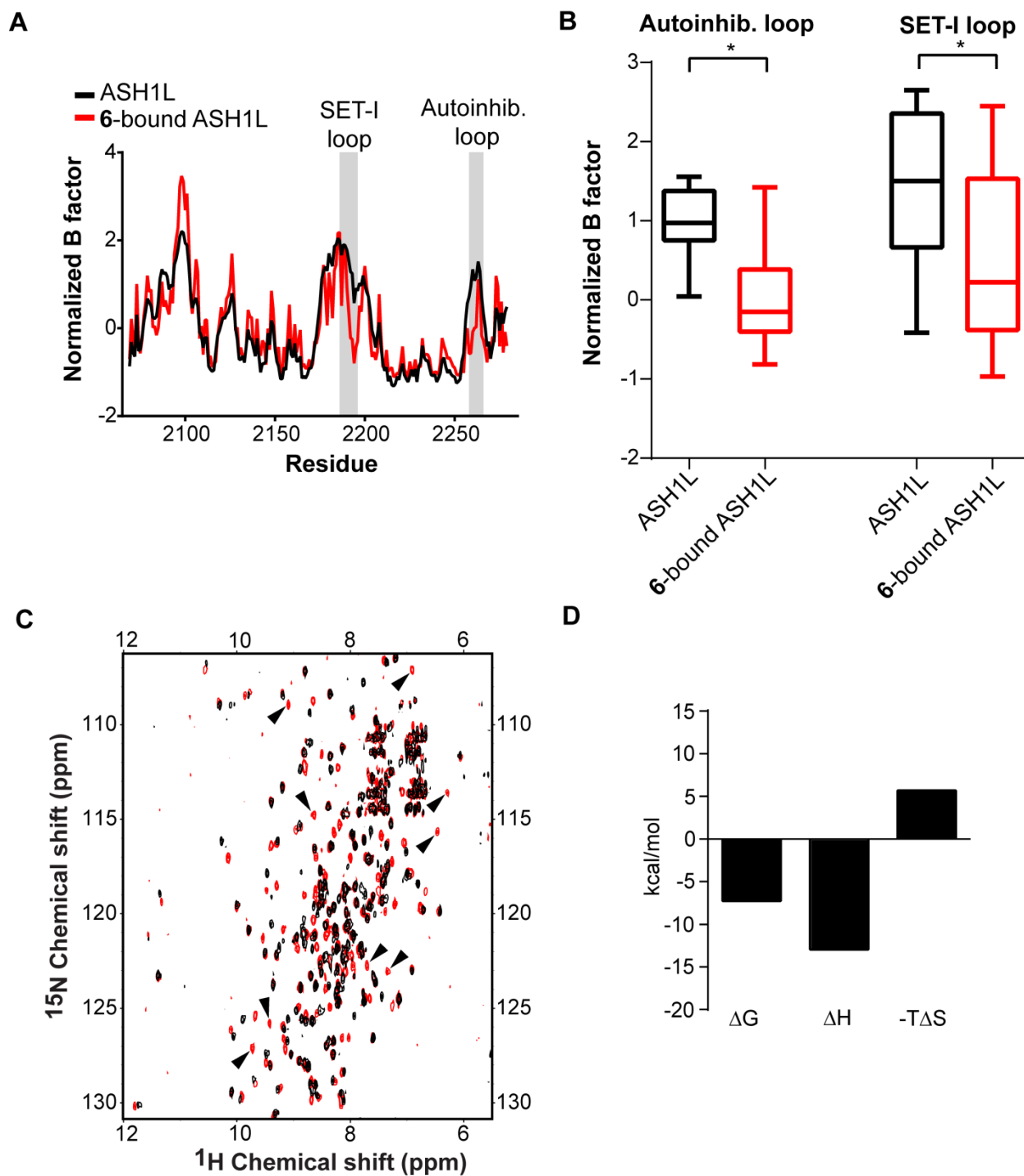


Figure 3.3. Compound **6** stabilizes autoinhibited conformation of ASH1L SET domain. **(A)** Normalized B factors for ASH1L without ligand and ASH1L bound to **6**. **(B)** Box and whiskers plot of normalized B factors for the autoinhibitory and SET-I loops of ASH1L without ligand and ASH1L bound to **6**. Whiskers show min to max. * $P < 0.0001$, t test. **(C)** ^{15}N - ^1H TROSY spectra of ASH1L without ligand (black) and ASH1L saturated with **6** (red). Arrowheads mark peaks in the red spectrum lacking counterparts in the black spectrum. **(D)** Thermodynamic parameters of binding of **6** to ASH1L SET, determined from ITC.

Structure-based design of SK-1

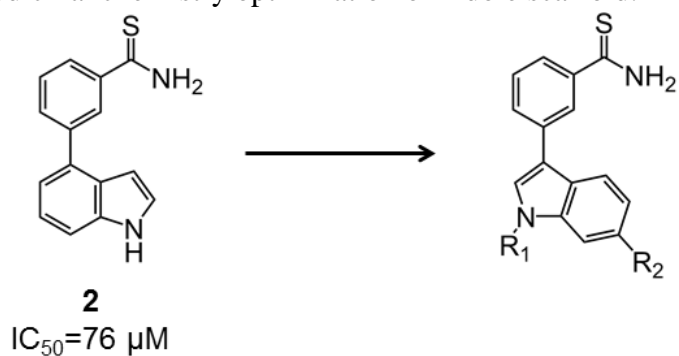
Based on the high resolution structure of ASH1L with **6**, we designed another series of compounds with substitutions at positions 1 and 6 of the indole. First we tested whether hydrogen bonding to the protein could be optimized by modifying the methylamine at position 6 to amine **7** (**Table 3.3**). This modification led to a small decrease in potency in the enzyme assay but a small improvement in K_d measured by ITC from 4.6 μM in **6** to 4.0 μM in **7**. Next, in the structure with **6** we observed several backbone carbonyls in the vicinity of the hydroxyethyl substituent. Although the carbonyl groups are not oriented optimally for formation of hydrogen bonds with the inhibitor, we hypothesized that fluorinated substituents at position 1 of the indole might engage in orthogonal C-F \cdots C=O interactions with the carbonyls [190]. Therefore, in the context of amine at position 6, we synthesized trifluorobutyl **8** and difluoropropyl **9**, which had K_d values of 3.4 and 2.0 μM , respectively (**Table 3.3** and **Figure 3.4A**, ITC data not shown). Compound **9** also has an enzyme assay IC_{50} value of 3.6 ± 0.1 μM , making it the most potent ASH1L inhibitor thus far. A crystal structure of **9** in complex with ASH1L revealed two different conformations of the difluoropropyl group in different ASH1L monomers of the asymmetric unit, with the fluorines capable of interacting either with the backbone carbonyl of Lys2264 or that of Gly2280 (**Figure 3.4B**).

To enhance potency by forming two fluorine-backbone interactions simultaneously, we synthesized a branched fluoropropyl derivative SK-1 (**Figure 3.4A**). SK-1 inhibited ASH1L enzymatic activity with an IC_{50} of 4.6 ± 1.9 μM and bound to ASH1L with a K_d of 1.7 μM (**Figure 3.4C**), values similar to **8**. The ITC experiment showed thermodynamic parameters similar to **6**, with a favorable change in enthalpy ($\Delta H = -17.2$ kcal/mol) and an unfavorable entropic component ($-T\Delta S = 9.3$ kcal/mol) (**Figure 3.4C**), underscoring a binding event driven by

enthalpy, with formation of hydrogen and chalcogen bonds and van der Waals interactions within the binding pocket. Interestingly, a co-crystal structure of SK-1 showed that the branched fluoropropyl substituent does not interact strongly with the backbone carbonyls of Lys2264 or Gly2280. Instead, the branched substituent fills the concave surface formed by residues Ile2278, Ile2279, Gly2280, and Gly2281 as they make a β -turn (**Figure 3.4D**).

To evaluate selectivity of SK-1 for ASH1L, we tested the effect of SK-1 on the enzymatic activity of related SET domains including NSD1-3, SETD2, and EHMT2. The most closely related SET domain, NSD1, shares 46% sequence identity to ASH1L SET. We found that low micromolar concentrations of SK-1 that inhibit ASH1L by 50% or more have minimal or no effect on related SET domains, demonstrating excellent selectivity of SK-1 for ASH1L (**Figure 3.4E**).

Table 3.3. Medicinal chemistry optimization of indole scaffold.



Compound	R ₁	R ₂	IC ₅₀	K _d (μM)
3	H	H	100	n.d.
4	CH ₂ CH ₂ OH	H	22	n.d.
5	H	CH ₂ NH ₂	36	n.d.
6	CH ₂ CH ₂ OH	CH ₂ NH ₂	13	4.6
7	CH ₂ CH ₂ OH	NH ₂	23	4.0
8	CH ₂ CH ₂ CH ₂ CF ₃	NH ₂	7	3.4
9	CH ₂ CH ₂ CF ₂ H	NH ₂	3.6 ± 0.1	2.0
10	CH ₂ CH ₂ CH ₂ CF ₃		4.4 ± 0.3	2.1
11	CH ₂ CH ₂ CH ₂ CF ₃	CH ₂ OH	3.8 ± 1.8	2.1
SK-1	CH ₂ CH(CFH ₂) ₂	NH ₂	4.6 ± 1.9	1.7
12	CH ₂ CH(CFH ₂) ₂	CH ₂ OH	n.d.	1.1
13	CH ₂ CH ₂ CF ₂ H	CH ₂ OH	n.d.	0.9

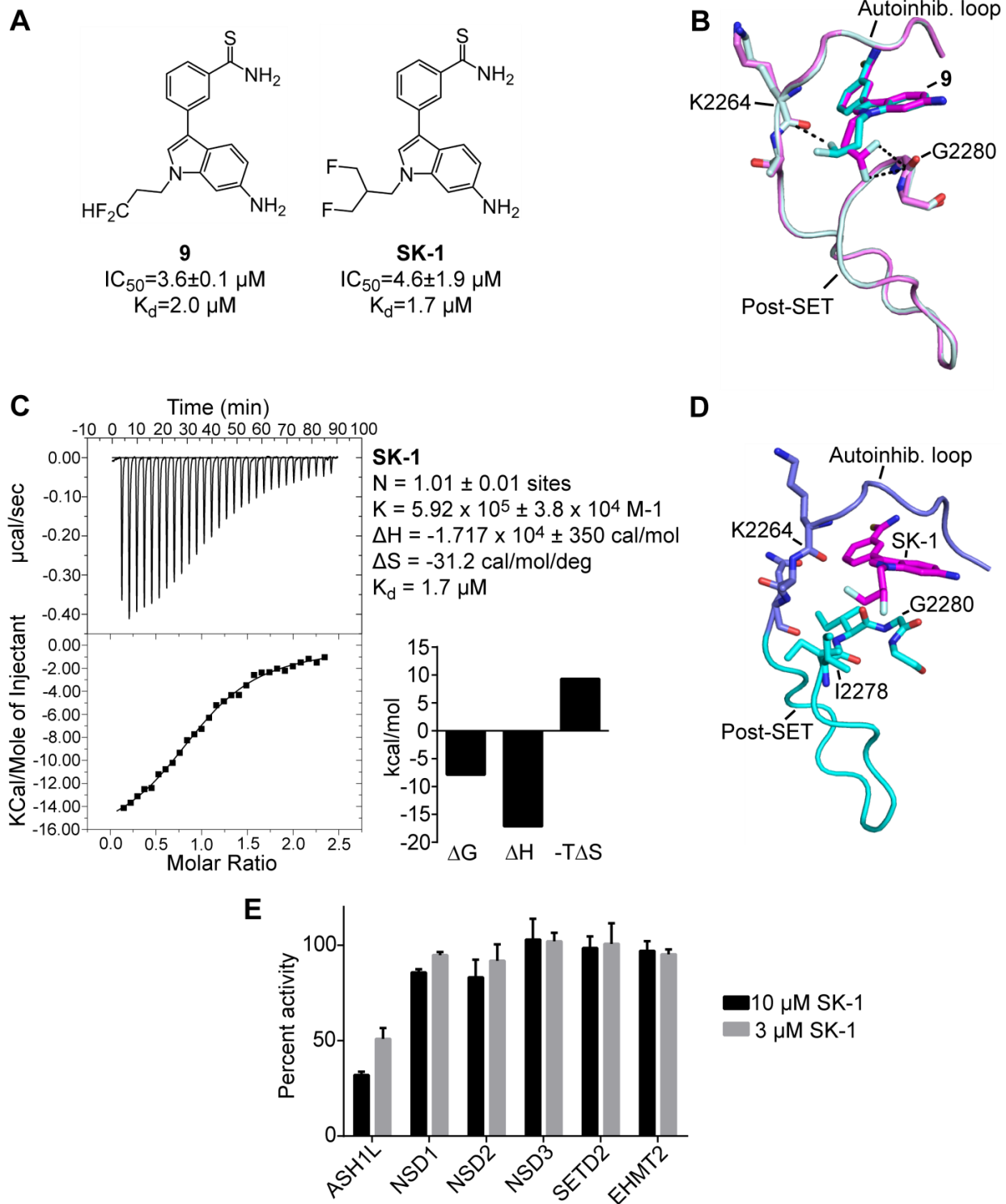


Figure 3.4. Development of SK-1. (A) Chemical structures of **9** and SK-1. (B) Crystal structure of ASH1L bound to **9**. Two different molecules of ASH1L SET in the asymmetric unit bound to **9** are shown in cyan and magenta. (C) Binding of SK-1 to ASH1L SET by ITC. Thermodynamic parameters of binding are listed and graphed. (D) Crystal structure of ASH1L bound to SK-1. (E) Selectivity testing of SK-1 across a panel of related KMTases.

Optimizing substituents at position 6 of indole and combining with substituents at position 1

In the context of the trifluorobutyl group at position 1, we attempted to optimize the substituent at position 6 by introducing more bulky substituents as well as a different hydrogen bond donor. The resulting imidazole with amine linker **10** and hydroxymethyl **11** compounds both had K_d values of 2.1 μM (**Table 3.3**, ITC data not shown), approaching a 2-fold improvement over the amine (compound **8**). At this point the SAR suggested that the best indole substituents were the difluoropropyl or branched monofluoropropyl at position 1 and imidazole with amine linker or hydroxymethyl at position 6. We combined branched fluoropropyl and hydroxymethyl substituents in **12**, which has a K_d of 1.1 μM (**Figure 3.5A**, ITC data not shown). We combined difluoropropyl and hydroxymethyl substituents in **13**, which has a K_d of 0.9 μM (**Figure 3.5A and B**). These results confirm that combining substituents at the 1 and 6 positions has an additive effect on binding affinity.

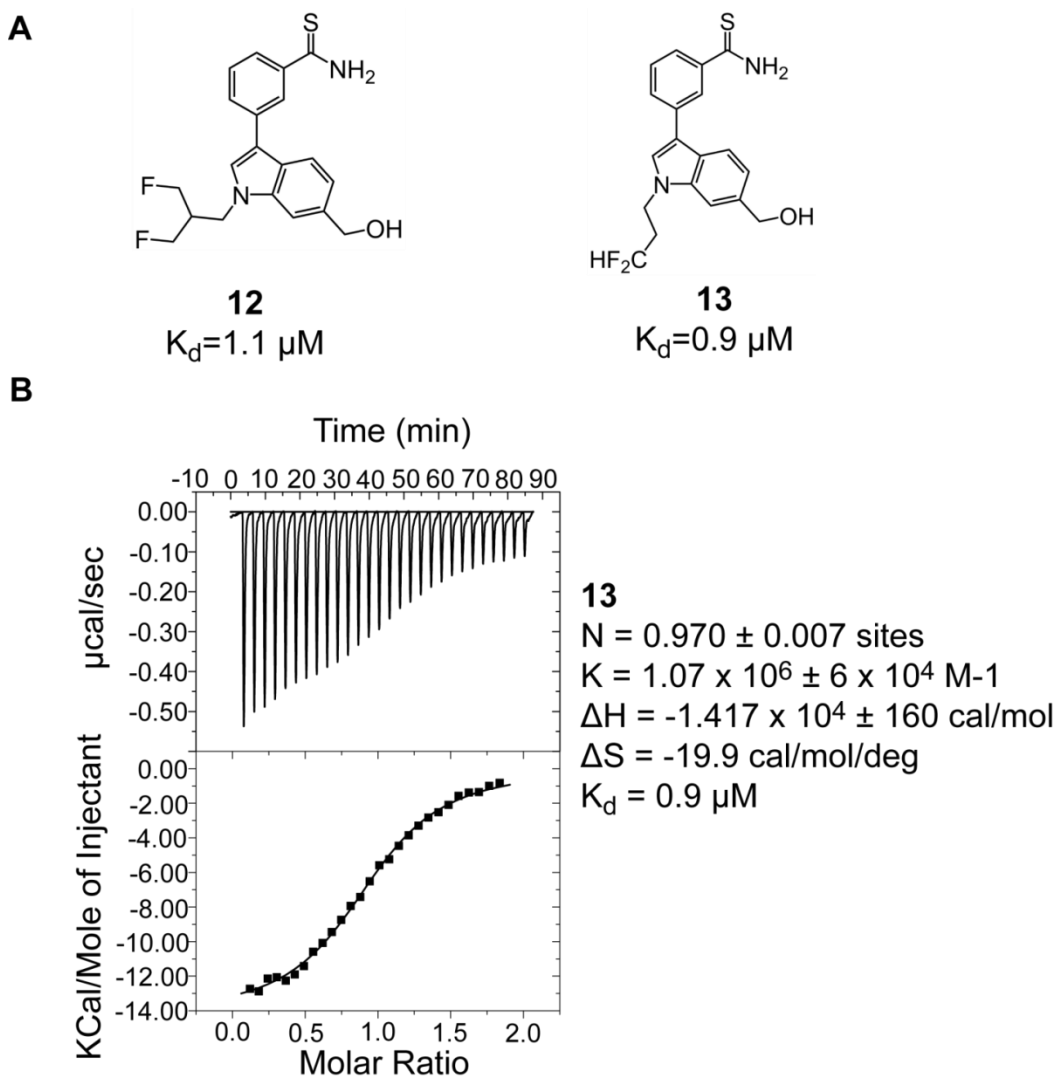


Figure 3.5. Combination of optimized substituents at 1 and 6 positions. (A) Chemical structures of **12** and **13**, with K_d values listed. (B) Binding of **13** to ASH1L by ITC. Thermodynamic parameters listed.

Sulfur-oxygen chalcogen bonding in ASH1L-inhibitor complexes

In all the co-crystal structures with ASH1L and inhibitors, we observed a sulfur-oxygen chalcogen bond between the thioamide sulfur of the inhibitor and the carbonyl oxygen of His2193. For example, in the SK-1 co-crystal structure the thioamide sulfur is positioned 3.0 \AA from the His2193 carbonyl oxygen (**Figure 3.6**). Interestingly, many methyltransferases including SET domains use sulfur-oxygen chalcogen bonding to mediate recognition of the SAM cofactor, which contains a sulfonium cation [188]. Although chalcogen bonding to SAM is not

observed in structures of ASH1L SET in the absence of inhibitor ($\leq 3.2 \text{ \AA}$ required between sulfonium and an oxygen), the backbone carbonyl oxygen of His2193 is intriguingly close ($3.3\text{-}3.9 \text{ \AA}$) to the SAM sulfonium [134]. Given that His2193 is located in a conformationally heterogeneous region of ASH1L SET [134], it remained possible that this residue engages in $\text{S}\cdots\text{O}$ chalcogen bonding transiently in solution, but the bond was not captured in the static crystal structure. Remarkably, the ASH1L inhibitors trigger a protein conformational change resulting in the His2193 carbonyl forming a chalcogen bond with SAM. For example, in the SK-1 co-crystal structure, the His2193 carbonyl oxygen is 3.2 \AA from the sulfonium of SAM (**Figure 3.6**). Thus, binding of SK-1 and related inhibitors creates a chalcogen bonding network involving the inhibitor, His2193, and SAM (**Figure 3.6**). By stabilizing the conformationally dynamic region of ASH1L, the inhibitors reveal a propensity for $\text{S}\cdots\text{O}$ chalcogen bonding between His2193 and SAM. Further crystallographic, NMR, and quantum mechanical studies may help determine if this bond plays a role in enzymatic activity.

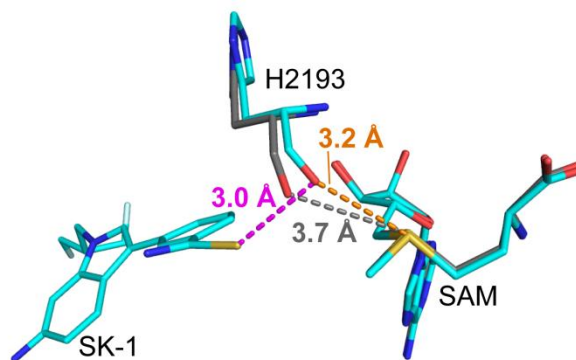


Figure 3.6. Chalcogen bonds induced by binding of SK-1 to ASH1L SET. SK-1 forms a chalcogen bond with the H2193 carbonyl oxygen, which in turn forms a chalcogen bond with the sulfonium of SAM. ASH1L Q2265A structure shown in grey, SK-1-bound ASH1L Q2265A structure shown in cyan.

Discussion

An autoinhibitory loop that blocks access of histone substrate to the active site is a unique feature of H3K36-specific SET domains. Although this loop was an interesting structural element, it was previously unclear how to exploit this loop for inhibitor development because there were no natural or synthetic small molecule ligands that bound to the loop region. It seemed likely that new chemical scaffolds would be required to target this site, perhaps which did not exist in HTS libraries. Here we have shown that FBDD is an effective method for probing the druggability of the autoinhibitory loop region and identifying novel compounds that interact with the loop. Fragment screening identified a compound that formed multiple specific interactions with ASH1L SET but had weak affinity due to its small size. We optimized the compound to increase potency but maintained the original interactions made by the fragment. Expansion from pyrrole to bi-substituted indole resulted in a 1000-fold improvement in binding affinity. The advanced ASH1L inhibitors have a unique scaffold that has not been previously reported in the SciFinder database. NMR and crystallographic evidence suggests that these compounds inhibit ASH1L activity by stabilizing the autoinhibited conformation of the SET domain. This work represents the first ASH1L inhibitors reported to date.

SK-1 is a potent and selective ASH1L inhibitor developed through iterative structure-based drug design. The thioamide group represents a unique chemical warhead, forming an elegant network of chalcogen and hydrogen bonds deep inside the protein. An additional chalcogen bond between His2193 and SAM induced by inhibitor binding may contribute to stabilization of the autoinhibitory loop region. The phenyl and indole rings of SK-1 form a scaffold that fills the pocket surrounded by the autoinhibitory loop. The branched fluoropropyl and amino substituents of the indole form additional contacts with the protein, together

contributing a greater than 20-fold increase in potency over the unsubstituted scaffold. Thus, SK-1 represents an efficient ASH1L inhibitor with a well-characterized structure-activity relationship. This study provides some intriguing evidence for the role of sulfur-oxygen chalcogen bonding in protein and medicinal chemistry and suggests that this type of interaction might be considered when designing other small molecule inhibitors. We have also shown how fluorination can be used to improve inhibitor potency, either by orthogonal C–F \cdots C=O interactions or van der Waals interactions.

In conclusion, in this work we used an innovative FBDD strategy to develop the first small molecule inhibitors of ASH1L. Structure-based medicinal chemistry optimization improved inhibitor potency greater than 1000-fold compared to the original fragment screening hit. These compounds are specific to ASH1L, are predicted to be cell-penetrant, and will be valuable tools for investigating the function of ASH1L in cancer and other diseases.

Chapter 4. Genetic and pharmacologic approaches identify novel functions of ASH1L in breast cancer and leukemia*

Abstract

ASH1L is an epigenetic regulator and histone lysine methyltransferase (KMTase) that is overexpressed in multiple forms of cancer and activates oncogenic *HOX* genes. To further investigate the role of ASH1L in cancer we performed genetic knockdown of the *ASH1L* gene and pharmacologic inhibition of ASH1L KMTase activity in cancer cells. We found that ASH1L is required for growth of breast cancer and leukemia cells, and that ASH1L activates *HOXA* genes and *MEIS1* in leukemia. Moreover, targeted in-frame deletion of the ASH1L SET domain and pharmacologic inhibition of ASH1L KMTase activity revealed a key role for the ASH1L SET domain in leukemic transformation mediated by MLL-AF9 and CALM-AF10. The ASH1L inhibitor SK-1 selectively prevents growth and induces apoptosis and differentiation of MLL-AF9-transformed mouse bone marrow cells. Conversely, SK-1 has no effect on differentiation of cells transformed by *Hoxa9/Meis1*. SK-1 also inhibits growth of MLL-rearranged human leukemia cell lines. Our results demonstrate cellular efficacy of ASH1L inhibitors and simultaneously uncover a new role for the ASH1L SET domain in acute leukemia.

*Rogawski DS, et al. *ASH1L inhibitors block leukemogenesis through downregulation of Hox gene expression*. Manuscript in preparation.

Introduction

Epigenetic pathways are disrupted in cancer, causing aberrations in chromatin structure, gene expression, and genome integrity [8]. ASH1L is an epigenetic regulator and histone 3, lysine 36 (H3K36)-specific methyltransferase (KMTase) with an emerging role in cancer. ASH1L is overexpressed in a variety of solid tumors, including thyroid and breast cancer [94,161]. In thyroid cancer, ASH1L is overexpressed in tumor-specific truncated forms. The tumor suppressor microRNA miR-142-3p inhibits ASH1L protein expression by binding to the *ASH1L* 3'UTR, an effect correlated with inhibition of colony formation and slowing of thyroid cancer cell growth [94]. In addition, the *ASH1L* gene frequently undergoes copy number amplification in aggressive basal-like breast cancer, and high expression of *ASH1L* mRNA is associated with shorter survival of breast cancer patients [161]. Finally, in hepatocellular carcinoma (HCC), structural variations are found near the *ASH1L* gene, and knockdown of *ASH1L* in HCC cells slows proliferation [191].

In multiple developmental and oncogenic contexts, ASH1L activates *HOXA*, *-B*, *-C*, and *-D* genes and *MEIS1* [15,88,93,96]. ASH1L's KMTase activity is required for at least some of its gene activating function, as deletion of the ASH1L SET domain in differentiating mouse embryonic stem cells leads to decreased expression of 152 genes, including members of the *Hox* and *Wnt* families [93]. These findings are highly relevant because *HOX* genes are oncogenic drivers in many different blood and solid tumors [89]. For example, overexpression of *HOXA9* is highly associated with a poor prognosis in acute myeloid leukemia (AML) [90], and *HOXA9* and its collaborator *MEIS1* are required for survival of MLL-rearranged leukemia cells [91,92]. Interestingly, ASH1L deficiency causes a major reduction in long-term hematopoietic stem cells (HSCs) in mouse bone marrow, but surprisingly has very modest effects on peripheral blood

counts due to increased proliferation of progenitors downstream of HSCs [15]. *ASH1L*-deficient HSCs are also unable to reconstitute bone marrow output when transplanted into lethally irradiated mice [15]. These findings indicate that *ASH1L* maintains quiescence and self-renewal potential of long-term HSCs, but whether *ASH1L* regulates the stemness properties of leukemic stem cells is unknown.

Despite the correlations between *ASH1L* and cancer, definitive evidence demonstrating the oncogenic potential of *ASH1L* has remained elusive. Here we undertook a combined genetic and pharmacologic approach to investigate *ASH1L*'s role in breast cancer and leukemia. We identified an important role for the *ASH1L* SET domain in activating *HOX* genes and maintaining leukemia in multiple cellular models.

Results

ASH1L regulates proliferation of breast cancer cells

Breast cancer cell lines mirror many of the molecular features found in patients and are useful model systems to investigate genes that contribute to oncogenesis [192]. Because *ASH1L* is frequently amplified in aggressive basal-like breast cancer, which includes the “triple-negative” immunophenotype [161], we aimed to determine the extent to which *ASH1L* regulates proliferation of breast cancer cells. We selected two cell lines, MDA-MB-468, a triple negative cell line, and MCF-7, an estrogen receptor-positive line, and then used shRNA to knockdown *ASH1L* expression, resulting in a reduction in *ASH1L* mRNA of approximately 40-50% in both cell lines (**Figure 4.1A and B**). We found that MDA-MB-468 cells transduced with the *ASH1L*-targeting shRNA grew more slowly over 12 days compared to cells transduced with a non-targeting control shRNA (**Figure 4.1C**). In contrast, MCF-7 cells transduced with the *ASH1L* and non-targeting shRNAs grew at the same rate (**Figure 4.1D**). These results suggest that

ASH1L regulates growth of triple-negative breast cancer. Additional experiments are ongoing to determine genes regulated by ASH1L in breast cancer.

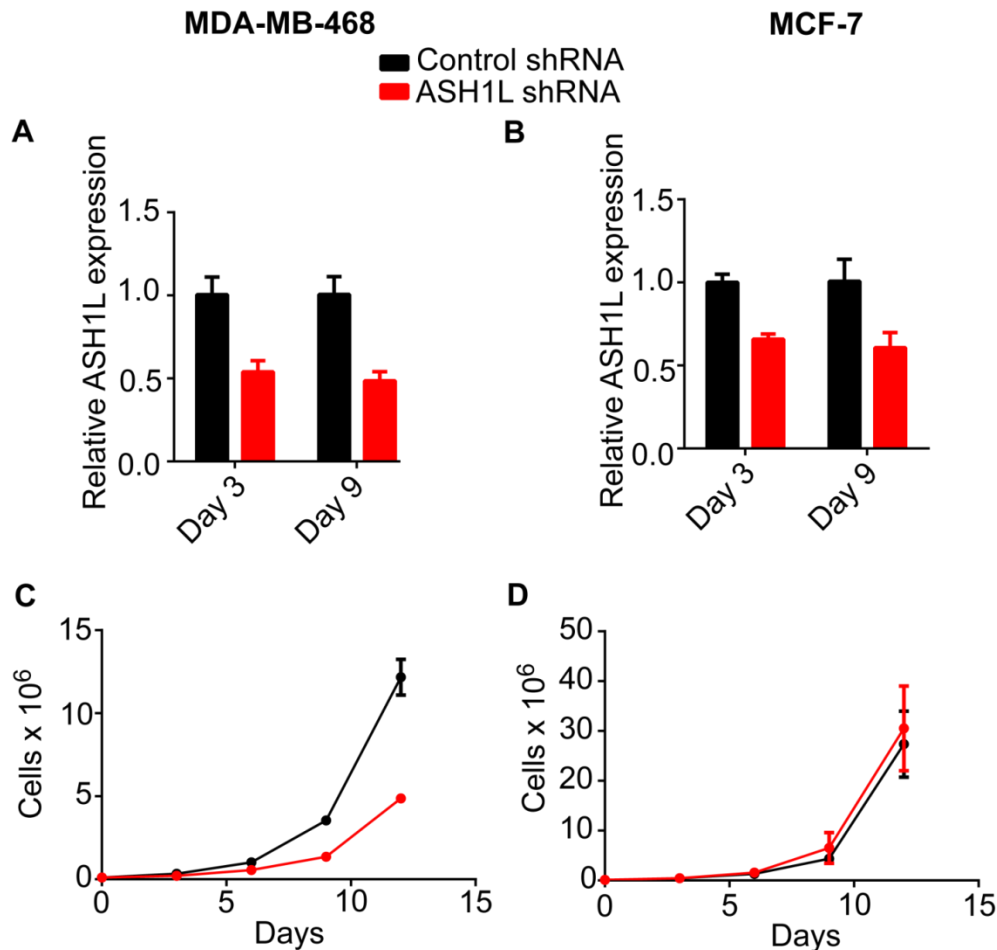


Figure 4.1. Knockdown of ASH1L slows proliferation of MDA-MB-468 cells. (A) and (B) Relative expression of ASH1L in MDA-MB-468 (A) and MCF-7 (B) cells treated with non-targeting shRNA (black) or ASH1L shRNA (red). Expression was normalized to HPRT1 and referenced to control shRNA-transduced cells. (C) and (D) Growth of MDA-MB-468 (C) and MCF-7 (D) cells treated with non-targeting shRNA (black) or ASH1L shRNA (red).

ASH1L regulates HOX gene expression in human leukemia cells

ASH1L's role in *HOX* gene activation in multiple developmental contexts [15,93] prompted us to determine whether ASH1L activates *HOX* genes in leukemia. We selected two human AML cell lines, THP-1 and KG1, for ASH1L knockdown experiments. THP-1 cells harbor the MLL-AF9 translocation, while KG1 cells lack MLL rearrangements but express high levels of *HOXA* genes. Transduction of cells with ASH1L shRNA led to a decrease in *ASH1L*

expression of 50-60% compared to cells transduced with a control shRNA (**Figure 4.2A, B**). In THP1 cells, we observed concomitant decreases in expression of *HOXA5*, *HOXA9*, *HOXA10*, and *MEIS1* by 40-70%. In KG1 cells, *ASH1L* knockdown caused similar decreases in *HOXA* genes and *MEIS1* of 20-60%. These data indicate that *ASH1L* regulates *HOX* and *MEIS1* expression in multiple AML cell lines.

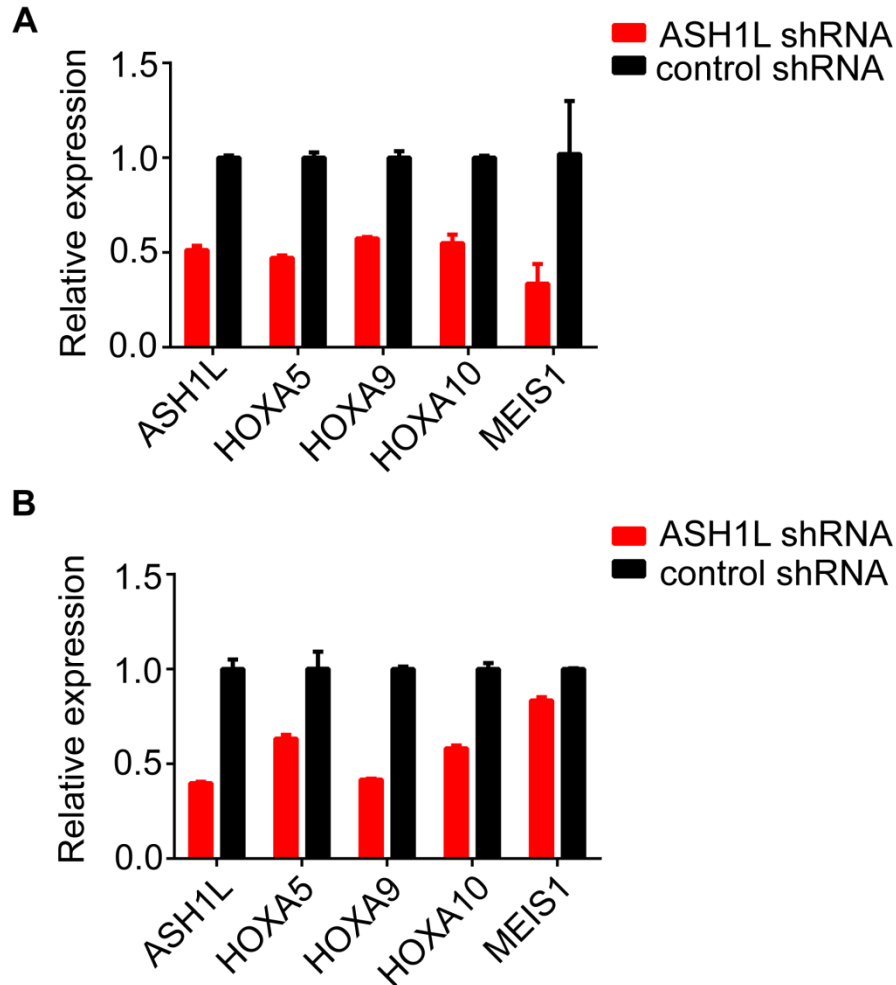


Figure 4.2. Relative expression of *ASH1L* and *HOX* genes in THP1 (**A**) and KG1 (**B**) cells 6 days after transduction with an shRNA targeting *ASH1L* (red) or a nontargeting control shRNA (black). Data are representative of three independent experiments. Expression was normalized to GAPDH and referenced to control shRNA-transduced cells.

ASH1L regulates proliferation of CALM-AF10-transformed mouse bone marrow cells

ASH1L's function in activating *HOX* genes suggested that ASH1L could regulate cell proliferation in acute leukemias that depend on *HOX* gene expression. Therefore, we chose to further study ASH1L's role in leukemia by using mouse bone marrow cells transformed with the *CALM-AF10* fusion gene (CALM-AF10 BMCs). CALM-AF10 is a rare but recurrent translocation observed in acute leukemia patients, and when expressed in mice the fusion gene induces a long-latency leukemia characterized by marked upregulation of *Hoxa* genes [193,194]. In collaboration with Andy Muntean's laboratory (Department of Pathology, University of Michigan), we transduced CALM-AF10 BMCs with a doxycycline-inducible shRNA targeting *Ash1l*. After induction of the shRNA with doxycycline, *Ash1l* expression decreased approximately 50% compared to un-induced control cells (**Figure 4.3A**). To determine the extent to which *Ash1l* regulates growth of CALM-AF10 BMCs, we performed a competitive proliferation assay by culturing CALM-AF10 BMCs harboring the inducible *Ash1l* shRNA together with parental CALM-AF10 BMCs in the presence of doxycycline. We observed a steady decrease in DsRed⁺, *Ash1l* shRNA-expressing cells over the course of two weeks, with near complete elimination of DsRed⁺ cells at 17 days (**Figure 4.3B**). These results indicate that CALM-AF10 BMCs require *Ash1l* for proliferation.

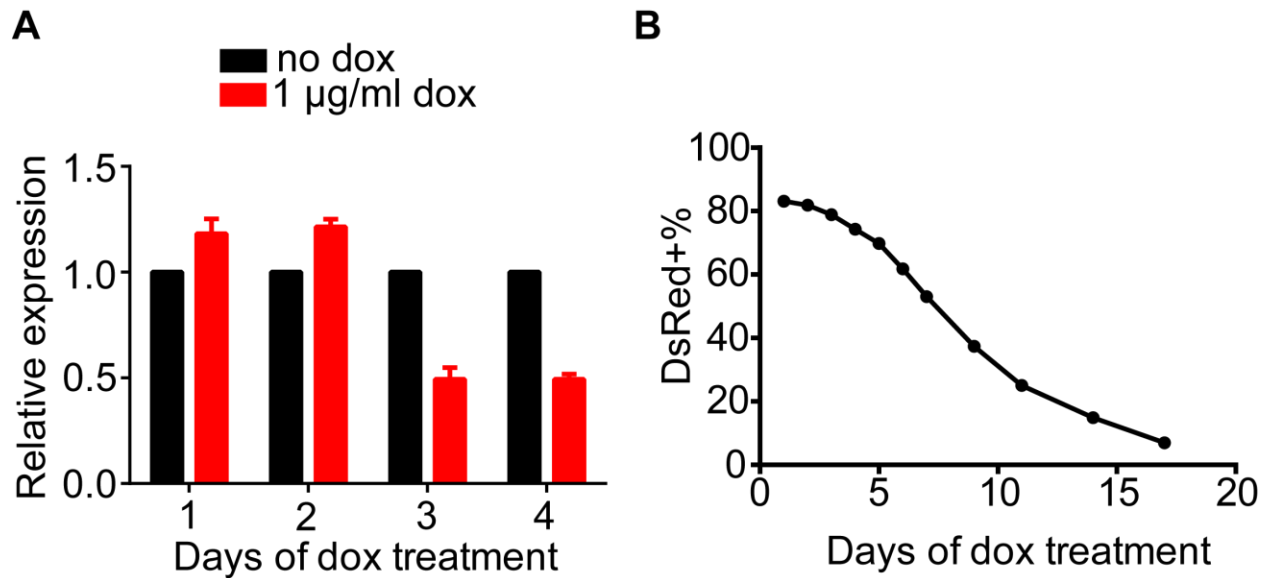


Figure 4.3. Knockdown of Ash11 inhibits proliferation of CALM-AF10 BMCs. **(A)** qRT-PCR for *Ash11* in CALM AF-10 BMCs expressing doxycycline-inducible shRNA targeting Ash11. Expression was normalized to Gapdh and referenced to uninduced cells. **(B)** Percentage of DsRed+, Ash11 shRNA-expressing cells in 17-day growth competition assay with parental CALM-AF10 BMCs.

The ASH1L SET domain is required for leukemic transformation

ASH1L is a large epigenetic regulator containing a catalytic SET domain as well as chromatin-interacting bromo, PHD, and BAH domains and an unannotated N-terminus. The function of different ASH1L regions and domains is poorly understood. Given that ASH1L is required for proliferation of CALM-AF10 BMCs, we aimed to determine the extent to which the ASH1L SET domain contributes to leukemogenesis. We obtained knock-in mice expressing mutant Ash11 containing a short in-frame deletion within the AWS-SET domain (Δ SET) (mice obtained from Ivan Maillard’s laboratory, Life Sciences Institute, University of Michigan, originally a kind gift from Gang Huang, Cincinnati Children’s Hospital, and Kenichi Nishioka, Saga University) [93]. We transduced Δ SET and wild-type BMCs with the CALM-AF10 and MLL-AF9 oncogenes and performed colony assays to determine whether the ASH1L SET domain is required for transformation by the oncogenes. After five rounds of plating, Δ SET

BMCs transduced with CALM-AF10 formed less than 20% the number of colonies that wild-type BMCs formed (**Figure 4.4A**). In addition, after four rounds of plating, Δ SET BMCs transduced with MLL-AF9 formed no visible colonies compared to robust colony formation by wild-type BMCs transduced with MLL-AF9 (**Figure 4.4A**). Microscopy revealed dense and compact colonies for wild-type cells transformed by MLL-AF9, but only diffuse small clusters of Δ SET cells, indicating a failure of transformation in the Δ SET cells (**Figure 4.4B**). Wright-Giemsa staining showed substantial morphological differences between Δ SET and wild-type BMCs transduced with the CALM-AF10 and MLL-AF9 oncogenes at the last round of the colony assay (**Figure 4.4C**). The wild-type cells had a blast-like morphology, but the Δ SET cells had increased cell size, lower nuclear-to-cytoplasmic ratio, and highly vacuolated cytoplasm, indicative of differentiation to a macrophage-like phenotype. These results indicate that the ASH1L SET domain is required for transformation by multiple leukemic oncogenes.

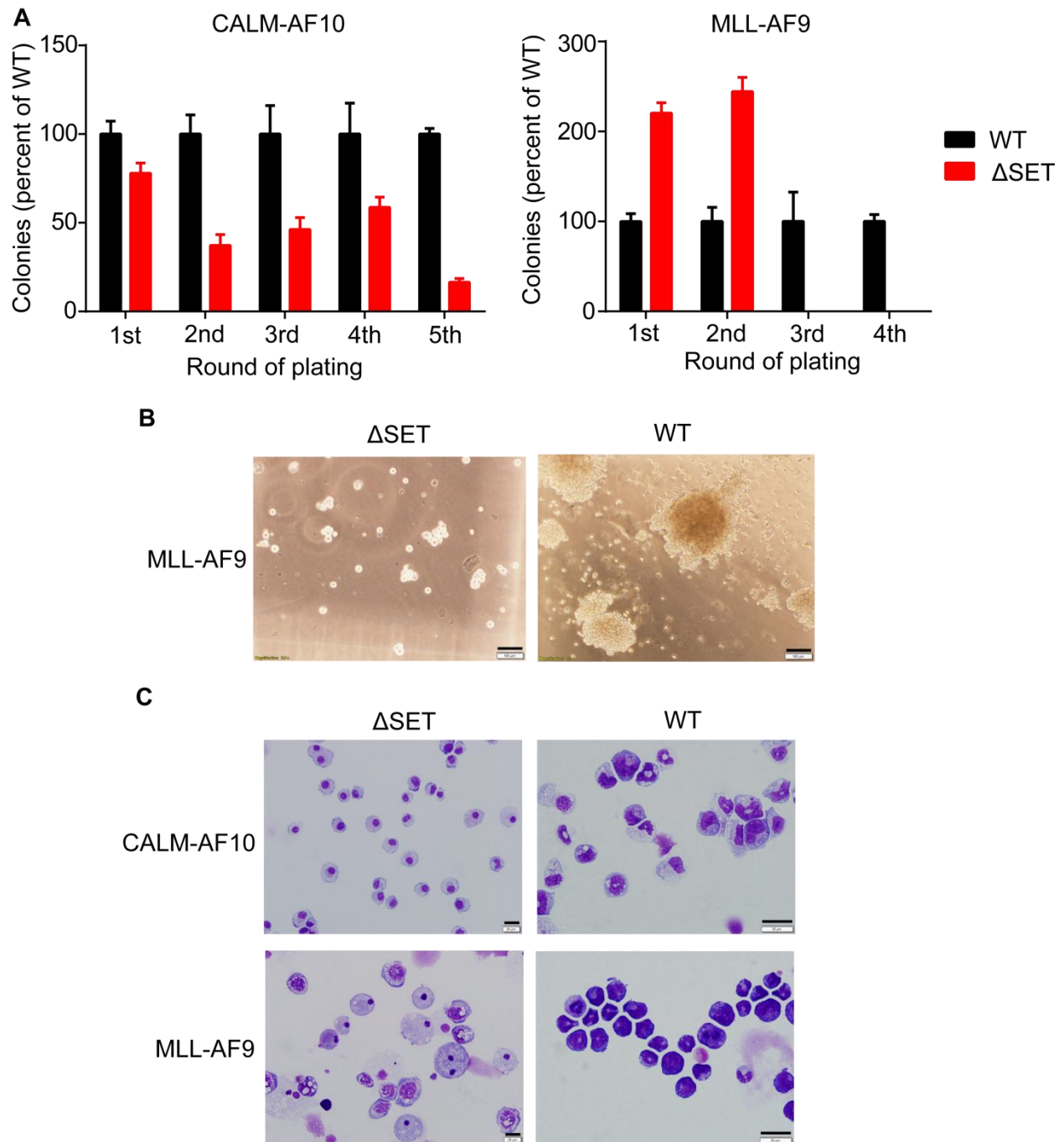


Figure 4.4. Colony formation by Δ SET and wild-type mouse BMCs. **(A)** Colonies formed by Δ SET and wild-type BMCs transduced with CALM-AF10 and MLL-AF9, presented as a percentage of wild-type colonies. **(B)** Colony morphology of Δ SET and wild-type BMCs transduced with oncogenes. Scale bars are 100 μ m. **(C)** Wright-Giemsa-stained Δ SET and wild-type BMCs transduced with oncogenes, at the final round of plating. Scale bars are 20 μ m.

SK-1 inhibits growth of MA9 but not HM2 mouse bone marrow cells

The failure of Δ SET BMCs to be transformed by MLL-AF9, together with the function of ASH1L in activating *HOXA9* and *MEIS1*, suggested that the KMTase activity of ASH1L might be involved in activating an oncogenic transcriptional program in MLL-rearranged leukemia. To directly investigate the role of ASH1L KMTase activity, we used the ASH1L KMTase inhibitor we developed, SK-1, and the negative control compound SK-nc. SK-1 inhibits ASH1L with an *in vitro* IC₅₀ of 4.6±1.9 μ M, while SK-nc binds to ASH1L with more than 100-fold weaker affinity. We also tested two other ASH1L inhibitors, **10** (IC₅₀=4.4±0.3 μ M) and **11** (IC₅₀=3.8±1.8 μ M). We treated mouse BMCs transformed by MLL-AF9 (MA9) or Hoxa9/Meis1 (HM2) with the inhibitors for 12 days. At 4 days of treatment, SK-1 inhibited growth of MA9 cells with GI₅₀ of 14 μ M compared to 50 μ M for HM2 cells, suggesting a specific inhibitory effect on MA9 proliferation (**Figure 4.5A and B**). **10** also inhibited growth of MA9 cells with nearly twofold greater potency than HM2 cells (GI₅₀ of 11 versus 20 μ M). In contrast, the negative control compound SK-nc had much weaker activity in both cell lines (GI₅₀ of 48 and 70 μ M in MA9 and HM2 cells, respectively). At 8 days of treatment, SK-1 and **10** showed enhanced activity in MA9 cells with GI₅₀ of 6 and 10 μ M, respectively, while their activity in HM2 cells remained weaker with GI₅₀ of 33 and 20 μ M, respectively (**Figure 4.5C and D**). By 12 days of treatment, SK-1 inhibited growth of MA9 cells with GI₅₀ of 4 μ M, more than 7-fold more potent than HM2 cells (29 μ M) (**Figure 4.5E and F**). In addition, **10** displayed a greater than 3-fold increased potency in MA9 cells (GI₅₀ of 7 μ M) compared to HM2 cells (23 μ M). At all the time points, inhibitor **11** showed weak growth inhibition for both MA9 and HM2 cells, suggesting that **11** may be poorly cell membrane permeable. The negative control compound SK-nc also maintained weak activity in MA9 and HM2 cells at 12 days (GI₅₀ of 34 and 59 μ M, respectively). These results indicate

that ASH1L inhibitors selectively inhibit growth of MA9 cells over HM2 cells, with SK-1 exhibiting the most potent and selective effects (**Figure 4.5G, H, and I**).

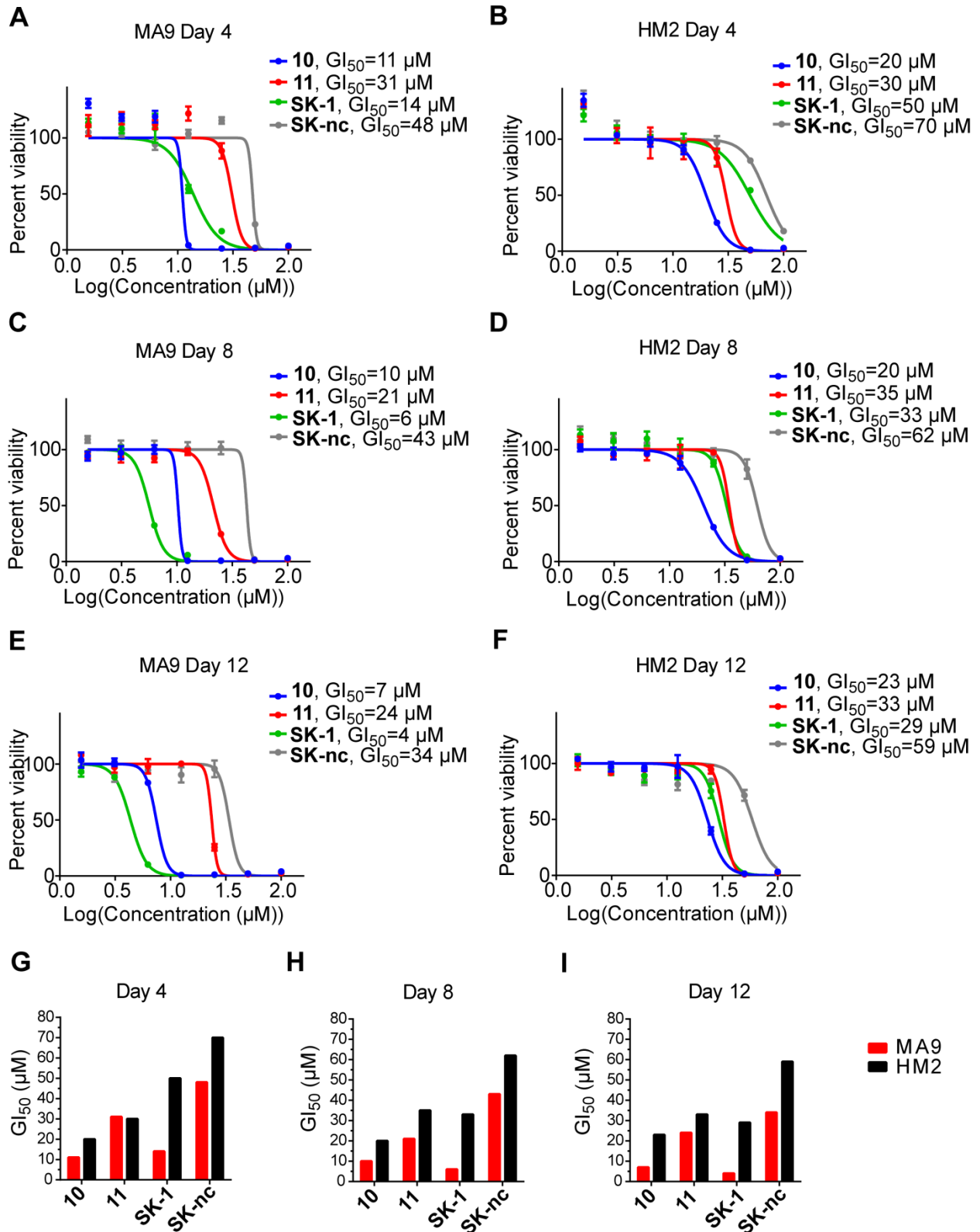


Figure 4.5. Effect of ASH1L inhibitors on cell viability of MA9 (A, C, and E) and HM2 (B, D, and F) cells, after 4 days (A and B), 8 days (C and D), and 12 days (E and F) of inhibitor treatment. GI_{50} values are graphed after 4 days (G), 8 days (H) and 12 days (I).

SK-1 induces differentiation of MA9 cells

Given the selective growth inhibition of MA9 cells mediated by SK-1, we investigated whether SK-1 induces hematopoietic differentiation. We found that MA9 cells treated with 3.1 μ M SK-1 for 12 days had increased cell size, increased number of vacuoles, and decreased nuclear-to-cytoplasmic ratio compared to MA9 cells treated with SK-nc, indicating differentiation into macrophages (**Figure 4.6A**). Flow cytometry also showed increased cell size (forward scatter) and granularity (side scatter) of MA9 cells treated with SK-1 compared to SK-nc-treated cells (**Figure 4.6B**). In addition, in cells treated with 3.1 μ M SK-1 we observed a slight increase in expression of CD11b (**Figure 4.6B**), a differentiation marker of myeloid cells, and higher concentrations of SK-1 are likely needed to observe a larger effect. MA9 cells treated with SK-1 had an increased percentage of cells undergoing apoptosis compared to DMSO- or SK-nc-treated cells (**Figure 4.6C**). Finally, SK-1 but not SK-nc caused a decrease in expression of *Hoxa* genes and *Meis1* (**Figure 4.6D**). Together, these results suggest that SK-1 inhibits growth of MA9 cells by blocking ASH1L SET-dependent activation of *Hoxa* genes and *Meis1*.

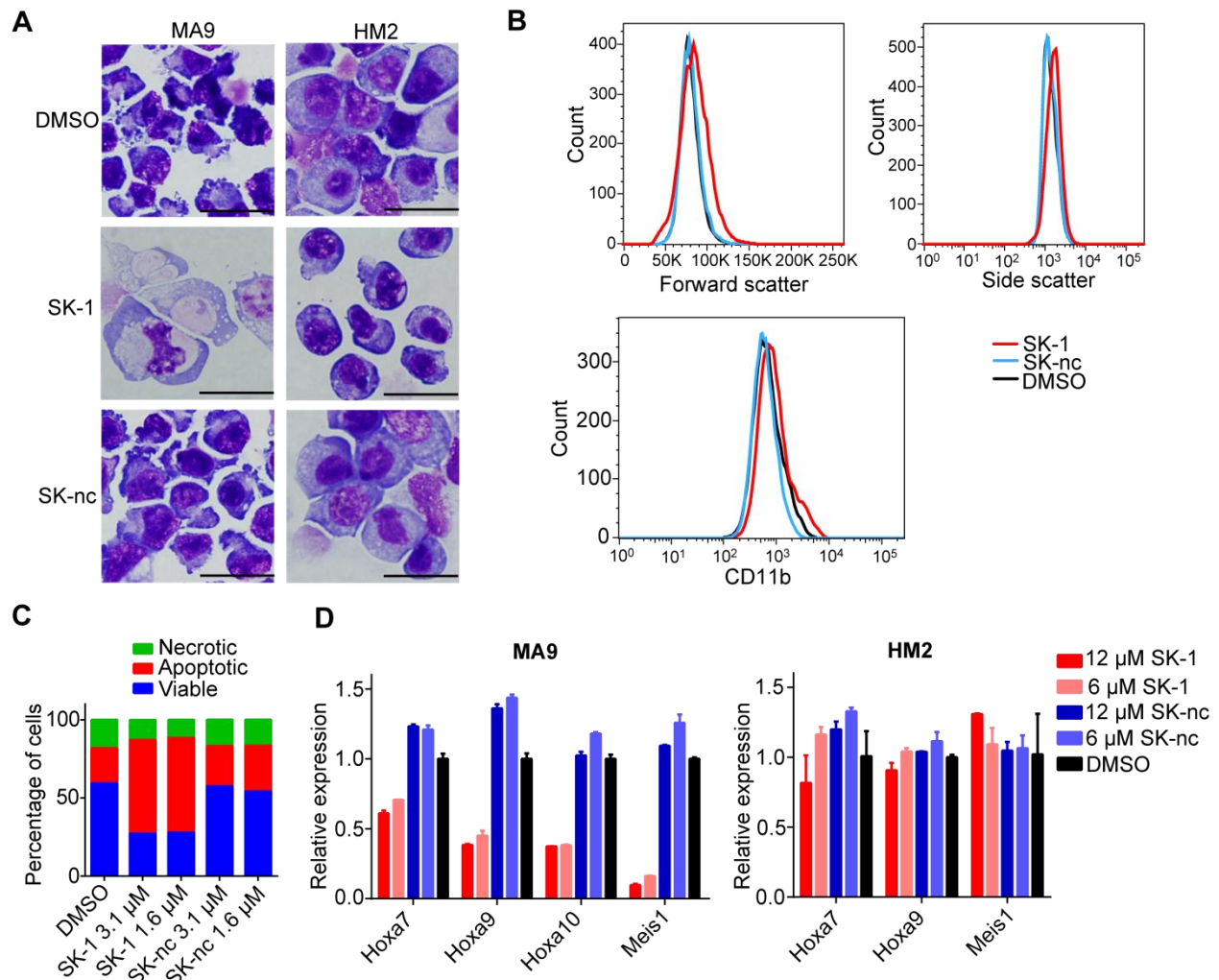


Figure 4.6. SK-1 induces differentiation of MA9 cells. **(A)** Wright-Giemsa-stained MA9 and HM2 cells treated for 12 days with 3.1 μ M SK-1, 3.1 μ M SK-nc or DMSO. Scale bars are 50 μ m. **(B)** Histograms of forward and side scatter and CD11b expression for MA9 cells treated for 12 days with 3.1 μ M SK-1, 3.1 μ M SK-nc, or DMSO, as detected by flow cytometry. **(C)** Apoptosis and necrosis induced by 12 days of SK-1, SK-nc, or DMSO treatment in MA9 cells, as detected by Annexin V and propidium iodide staining and flow cytometry. **(D)** qRT-PCR for *Hox* genes and *Meis1* in MA9 and HM2 cells treated for 4 days with SK-1, SK-nc, or DMSO. Expression was normalized to β -actin and referenced to DMSO-treated cells.

SK-1 blocks proliferation of MLL-rearranged human leukemic cell lines

We next tested the effect of SK-1 on proliferation of human leukemia cell lines. For this experiment we chose MV4;11 cells, which contain an MLL-AF4 translocation, and MOLM-13 cells, which harbor an MLL-AF9 translocation. We found that after 4 days of compound treatment, SK-1 prevents growth of MV4;11 cells with GI_{50} of 2 μ M, compared to 27 μ M for the

negative control compound SK-nc (**Figure 4.7A**). In addition, SK-1 prevents growth of MOLM-13 cells with GI₅₀ of 9 μM, compared to 38 μM for SK-nc (**Figure 4.7B**). These results suggest that the ASH1L SET domain is required for proliferation of human MLL-rearranged leukemias and that targeting ASH1L KMTase activity might represent a promising therapeutic strategy in acute leukemia.

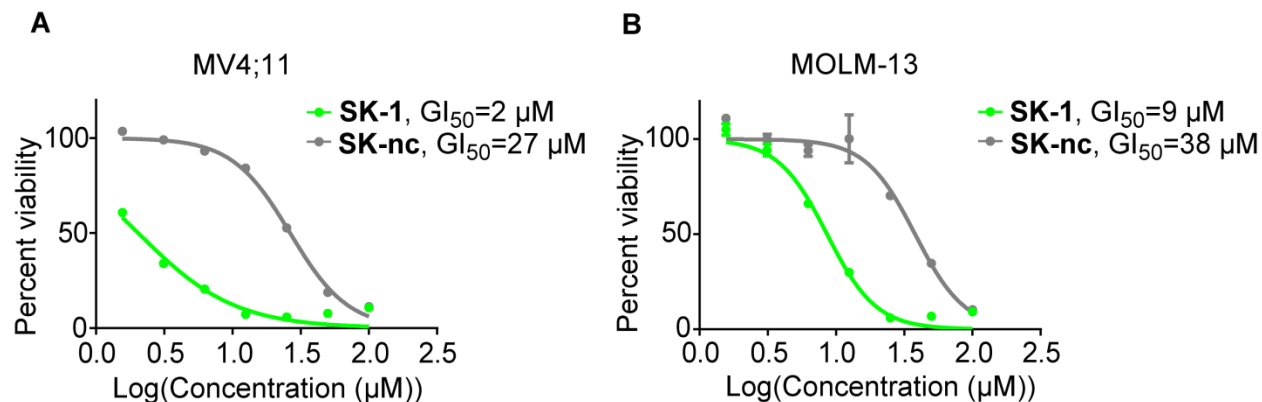


Figure 4.7. Growth inhibition by ASH1L inhibitor SK-1 and negative control compound SK-nc in human leukemia cell lines MV4;11 (**A**) and MOLM-13 (**B**) after 4 days of compound treatment.

Discussion

Numerous studies have provided circumstantial evidence for a role of ASH1L in oncogenesis. For example, ASH1L is overexpressed in multiple cancer types and activates *Hox* and *Wnt* genes, which are dysregulated in many varied tumors [89,195]. Moreover, the ASH1L SET domain is a key region of the protein that dimethylates H3K36 and activates *Hox* gene expression during development [93]. Nevertheless, definitive proof of ASH1L's oncogenic activities had previously remained elusive. Here we used genetic and pharmacologic approaches to demonstrate that ASH1L activates *HOX* genes and regulates cell growth in a subset of acute leukemia. Our results show that ASH1L is required for proliferation in leukemias driven by MLL rearrangements and the CALM-AF10 fusion, but not by *Hoxa9/Meis1*. Furthermore, using a targeted in-frame deletion of the ASH1L AWS-SET domain and a small molecule inhibitor of

ASH1L KMTase activity, we provide evidence suggesting that the SET domain in particular is important for ASH1L's oncogenic function. This finding is highly significant given that ASH1L is a large epigenetic regulator with multiple chromatin-interacting domains. Although we demonstrated a key role for the SET domain in leukemogenesis, the function of the other domains remains to be explored. Future studies could test whether combination of ASH1L knockdown and an ASH1L SET inhibitor has a more pronounced effect on differentiation of leukemia cells than the SET inhibitor alone, which would suggest that other regions of ASH1L contribute to leukemogenesis distinctly from the SET domain.

We found that the ASH1L SET domain is required for leukemic transformation mediated by two different fusion proteins, MLL-AF9 and CALM-AF10. AF9 and AF10 are nuclear proteins that participate in the same protein interaction network [196], but MLL and CALM function in quite different pathways. Wild-type MLL regulates *Hox* gene expression and is essential for hematopoietic development [197,198], while wild-type CALM plays roles in clathrin-mediated endocytosis, erythropoiesis, and iron metabolism [199,200]. Nevertheless, a common feature of leukemias driven by MLL fusion proteins and the CALM-AF10 fusion is the high expression of *HOXA* genes and *MEIS1* [201]. Our finding that ASH1L is important in both MLL-AF9- and CALM-AF10-driven leukemia suggests that ASH1L might be an important lynchpin in *HOX* gene activation. In support of this hypothesis, we found that ASH1L activates *HOX* genes in human leukemia THP1 and KG1 cells. THP1 cells harbor an MLL-AF9 translocation, while KG1 cells lack MLL rearrangements, yet both cell lines have comparably high levels of *HOXA9* expression relative to the K562 chronic myeloid leukemia cell line (data not shown). We have also found that human leukemia cell lines with MLL-AF4 and MLL-AF9 translocations are sensitive to ASH1L inhibition, suggesting that ASH1L KMTase activity is

important in human MLL-rearranged leukemia with varying translocation partner genes. Further testing of ASH1L inhibitors in a larger panel of leukemia cell lines will reveal whether the ASH1L SET domain is universally important for *HOX* gene activation and cell proliferation in leukemia with high *HOX* expression.

Compared to similar ASH1L inhibitors, SK-1 showed the most potent and selective activity on MA9 cells versus HM2 cells (GI₅₀ of 4 μM versus 29 μM at 12 days). Inhibitor **10** had a similar potency of growth inhibition (GI₅₀ of 7 μM at 12 days) but was not as selective for MA9 cells as SK-1, suggesting that **10** may cause some off-target toxicity. Inhibitor **11** showed weak growth inhibition for both MA9 and HM2 cells, suggesting that **11** may be poorly cell membrane permeable. Cellular testing of additional ASH1L inhibitors and control compounds will help further characterize the toxicity and membrane permeability profiles of these molecules. Optimization of the pharmacokinetic parameters of ASH1L inhibitors for *in vivo* experiments will also be the subject of future studies.

In conclusion, we have shown that ASH1L inhibitors developed in our lab have potent cellular activity, and at the same time we used the inhibitors to discover a novel role for ASH1L in leukemia. ASH1L inhibitors will be valuable tools to investigate the function of the ASH1L SET domain in many types of cancer. For example, we have shown that ASH1L regulates proliferation of triple negative breast cancer cells, and ASH1L inhibitors can be used to determine if the ASH1L SET domain is involved in breast cancer. The ASH1L inhibitors we developed have low molecular weight (<400 Da) and are readily accessible synthetically. Future work will focus on honing potency and pharmacokinetic properties for *in vivo* studies with animal models of leukemia and breast cancer.

Chapter 5. Conclusions and perspectives

Stephen Frye described five principles of a quality chemical probe [202]:

- 1) The probe is sufficiently potent and selective *in vitro* to confidently associate its *in vitro* profile to its cellular or *in vivo* profile.
- 2) Activity in a cell-based or cell-free assay influences a physiologic function of the target in a dose-dependent manner.
- 3) Has sufficient chemical and physical property data to interpret results as due to its intact structure or a well-characterized derivative.
- 4) Cellular activity data allow one to confidently address at least one hypothesis about the role of the molecular target in a cell's response to its environment.
- 5) The probe is freely available to the academic community with no restrictions on use.

In this thesis work, we made major strides toward developing the first quality chemical probe and small molecule inhibitor of ASH1L. Regarding principle 1, our current most potent ASH1L inhibitors bind with 1 μ M affinity and show an excellent selectivity profile over related histone lysine methyltransferases (KMTases). For principle 2, the inhibitors block ASH1L-mediated methylation of the physiologically relevant nucleosome substrate, and ChIP assays will be conducted to determine the effect of ASH1L inhibitors on histone methylation in cells. In regards to principle 3, crystal structures of ASH1L in complex with inhibitors reveal specific interactions made by the intact compounds, in agreement with the structure-activity relationship. For principle 4, we used the ASH1L inhibitor SK-1 and negative control compound SK-nc to discover a new role for ASH1L in leukemia. As for principle 5, we plan to make ASH1L

inhibitors freely available for researchers around the world to investigate ASH1L's role in cancer and other biological systems. Altogether, we made substantial progress toward fulfilling Frye's principles for a quality chemical probe, and further synthetic modifications in the near future will hone potency and cellular activity. Eventually, ASH1L inhibitors could be developed into novel therapeutics for human disease.

The first inhibitors of ASH1L KMTase activity represent a new paradigm in the KMTase inhibitor field

Small molecule inhibitors have not been reported for most H3K36-specific KMTases, and no inhibitors have been reported for ASH1L or the related NSD proteins, despite strong evidence for their oncogenic function (Chapter 1). On the other hand, academic and industrial groups have developed multiple inhibitors for KMTases that methylate H3K4, H3K9, H3K27, and H3K79 [9,203–205]. The KMTase inhibitors reported to date fall into two categories: SAM-competitive and histone substrate-competitive. Two of the best characterized SAM-competitive inhibitors are EZH2-targeted tazemetostat (EPZ-6438) and DOT1L-targeted pinometostat (EPZ-5676), both of which are currently in clinical trials. EPZ-6438 was optimized from a high-throughput screening hit [203], inhibits EZH2 with K_i of 2.5 nM, and has 35-fold selectivity for EZH2 over the very closely related EZH1 and >4,500-fold selectivity relative to a panel of 14 other KMTases [206]. In contrast, development of EPZ-5676 for DOT1L proceeded through modification of SAM and was guided by mechanistic and structural understanding of the DOT1L-catalyzed reaction [9]. EPZ-5676 inhibits DOT1L with K_i of ≤ 0.08 nM [207] and has >37,000-fold selectivity for DOT1L relative to a panel of other KMTases [207]. The excellent selectivity profiles for EPZ-6438 and EPZ-5676 show that while all KMTases utilize SAM, potent and selective SAM-competitive inhibitors can be developed for at least a subset of the enzymes. On the other hand, histone substrate-competitive inhibitors of SET domain KMTases

with superb potency and selectivity have also been reported. As an example, UNC0642 inhibits the H3K9 KMTases G9a and GLP with $IC_{50} < 2.5$ nM and is $>2,000$ -fold selective relative to other KMTases [205]. Nevertheless, development of SAM- or histone substrate-competitive inhibitors may not be the optimal approach for all KMTases, particularly for autoinhibited H3K36-specific KMTases that lack a defined histone binding pocket in existing crystal structures [134,208].

In this thesis work, we developed the first small molecule inhibitors of the H3K36 KMTase ASH1L. These inhibitors represent a new paradigm in the KMTase inhibitor field. Rather than binding to a well-defined SAM or histone binding pocket on the SET domain, the ASH1L inhibitors bind to the autoinhibitory loop region and stabilize the autoinhibited conformation of the SET domain. Mechanism of action studies are ongoing to determine whether the ASH1L inhibitors are competitive with SAM or nucleosome; competition with SAM is unlikely given the tertiary ASH1L/SAM/inhibitor complexes observed by crystallographic studies, while competition with nucleosome remains a possibility as the inhibitor and nucleosome may be unable to simultaneously bind to the SET domain. Regardless, the ASH1L inhibitors are better viewed as allosteric inhibitors because rather than bind to a substrate binding pocket, they modify the structure and dynamics of the ASH1L autoinhibitory loop region to block enzymatic activity. It is intriguing to speculate that ASH1L enzymatic activity *in vivo* might normally be controlled by physiologic allosteric regulation involving the autoinhibitory loop. For instance, ASH1L enzymatic activity might be regulated by post-translational modification of the autoinhibitory loop, or by binding of the loop to a cofactor molecule. Indeed, ostensibly novel small molecule binding sites on proteins discovered in biochemical screens sometimes later reveal themselves to be important physiologic regulatory sites that bind to

endogenous ligands or cofactors [209,210]. Our studies motivate further research to identify allosteric regulatory sites on ASH1L and other SET domains. Moreover, development of allosteric inhibitors of SET domain-containing KMTases could be an effective alternative strategy compared to the previously reported inhibitors that bind to the SAM or histone substrate binding pockets.

Our studies are additionally highly innovative because we used a fragment-based drug discovery approach (FBDD) to develop the first ASH1L inhibitors. Our work represents the first successful application of FBDD to KMTases and suggests that FBDD may be a worthwhile strategy for developing inhibitors of other KMTases. FBDD may be particularly well-suited to identify new allosteric regulatory sites on SET domain-containing proteins. In sum, the development of allosteric ASH1L inhibitors using an FBDD strategy represents a highly innovative and significant breakthrough in the KMTase inhibitor field.

Advantages and disadvantages of the FBDD approach

In this work we used FBDD to design potent, selective, and cell-active ASH1L inhibitors. We started with a fragment screen of 1500 compounds and selected one ligand for medicinal chemistry optimization. We did not obtain crystal structures of the ligand/ASH1L complex during the initial medicinal chemistry effort, but we still substantially improved binding affinity by expanding the pyrrole in the screening hit to an indole. With sufficiently potent compounds and an optimized crystallization strategy, we later successfully obtained crystal structures of nine compounds in complex with ASH1L, which provided valuable information for additional inhibitor design. The FBDD approach resulted in a ligand-efficient, drug-like lead compound with cellular activity.

The FBDD approach has several advantages over high-throughput screening (HTS) for developing inhibitors of H3K36-specific KMTases. To begin, an HTS campaign that measures KMTase activity is susceptible to artifacts, including compounds that alter the substrate or cause aggregation of the enzyme. In contrast, FBDD ensured that we found screening hits that form specific interactions with the ASH1L SET domain. Second, KMTase inhibitors appear to be rare in existing HTS libraries, as three different pharmaceutical companies reported SAM-competitive inhibitors of the EZH2 KMTase with similar pyridone-containing scaffolds [11]. FBDD allowed us to construct a new KMTase inhibitor scaffold starting from a “foothold” fragment-like ligand. Finally, FBDD not only identified a novel series of ligands that bind to ASH1L but also identified a novel ligand binding site. There were previously no natural or synthetic small molecule ligands of the ASH1L autoinhibitory loop region, and static crystal structures showed scant space to accommodate a ligand. However, our NMR studies showed conformational heterogeneity of the autoinhibitory loop region (Chapter 2), suggesting that a ligand-binding pocket might exist transiently in solution, and FBDD confirmed that this site is capable of binding small molecules. While many previous efforts were focused on developing SAM-competitive inhibitors of KMTases, often by modifying SAM or the natural product sinefungin [129,143,144], FBDD led to the discovery of an additional small molecule binding region that may be exploited for inhibitor development.

On the other hand, FBDD has some disadvantages, mainly associated with time and cost. Due to their small size, fragment-like ligands generally have weak binding affinity and must be extensively optimized for biological experiments. A major synthetic chemistry effort is therefore required. Crystal structures of the ligand bound to the target protein are very important for optimization but are not always easily obtained. Even with a co-crystal structure and medicinal

chemistry expertise on hand, some fragment binding pockets are too small to allow growth of the fragment. In summary, FBDD is a rigorous step-wise approach for designing novel, ligand-efficient inhibitors, but the time and cost required to obtain cell-active compounds is often larger than that required with HTS.

Improving potency, physicochemical, and pharmacokinetic properties of ASH1L inhibitors

From the original fragment screening hit **1** that had millimolar binding affinity for ASH1L SET, we developed cell-active ASH1L inhibitors with 1 μ M potency. Nevertheless, information from the structure-activity relationship and crystal structures demonstrates that many opportunities remain for further improving the potency of ASH1L inhibitors. For example, our early SAR suggested that binding affinity could be improved by modifying the thioamide conjugated phenyl to other heterocyclic rings. In addition, crystal structures suggest that other fused ring systems may be preferable over the indole scaffold, due to the potential for a hydrogen bond with the backbone carbonyl of K2264. Perhaps the most promising locations for further optimization are positions 1 and 6 of the indole. The crystal structure with SK-1 reveals a groove formed by the C-terminal residues that could potentially interact with a variety of substituents at position 1 (**Figure 5.1A**). At position 6 of the indole, the SAR demonstrated that hydroxymethyl and imidazole with amine linker substituents are preferred over aminomethyl and amino substituents (Chapter 3 and Appendix). Interestingly the crystal structure with **11** showed that hydroxymethyl at position 6 forms a water-mediated hydrogen bond with the N2261 sidechain, whereas the crystal structure with **6** showed that aminomethyl at position 6 points in the opposite direction and forms hydrogen bonds with the backbone carbonyl of G2280 and the sidechain of N2256 (**Figure 5.1B**). Therefore, incorporating a substituent at this position capable of interacting with both regions of the protein may increase potency. In addition, there is a

possibility of expanding the substituent toward residues 2257-2259, which form the C-terminus of an α -helix (**Figure 5.1B**). Finally, we have consistently found that combining substituents at the 1 and 6 positions results in an additive effect on binding affinity (Chapter 3 and Appendix). As we identify more potent substituents at each position, we will continue to combine them to maximize potency. In addition to improving potency, we will simultaneously work to maintain adequate physicochemical and pharmacokinetic properties. Fortunately, SK-1 already shares characteristics of orally active drugs [211]. SK-1 has molecular weight of 359 Da, four hydrogen bond donors, three hydrogen bond acceptors, and clogP of 3.2 (calculated in ChemDraw), supporting its further optimization for *in vivo* studies.

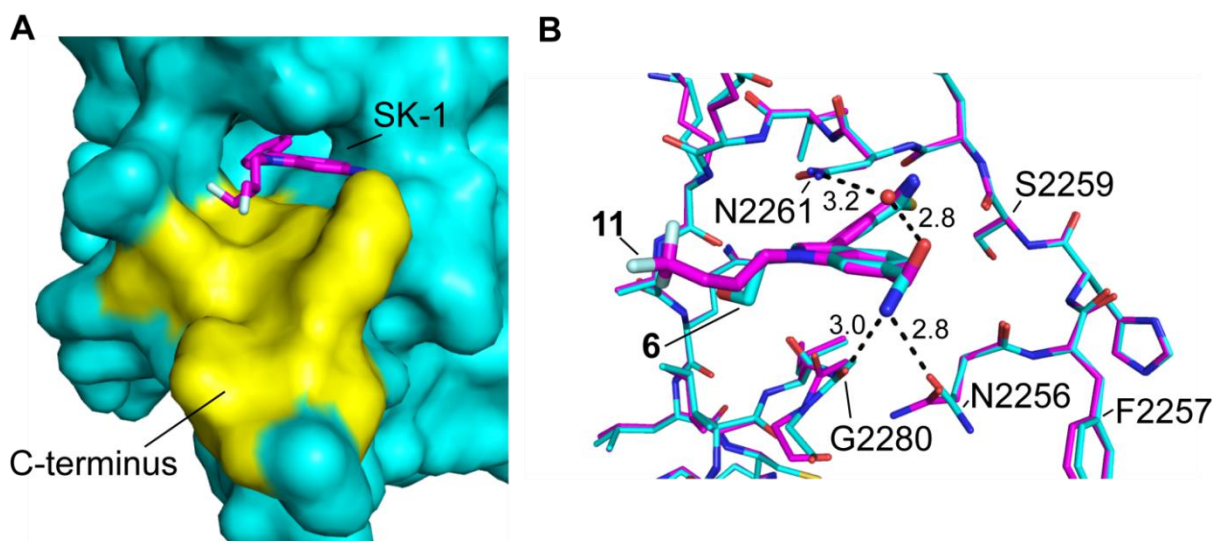


Figure 5.1. Regions for optimization of ASH1L inhibitors. (A) Surface representation of crystal structure of ASH1L SET in complex with SK-1, with C-terminal residues highlighted in yellow. (B) Overlay of crystal structures of ASH1L in complex with **6** and **11**. The different interactions made by **6** and **11** are shown.

The role of ASH1L in leukemia

Our studies revealed a new role for ASH1L in activating *Hox* genes and regulating cell growth in leukemia. Because mouse BMCs transformed by *Hoxa9/Meis1* are not sensitive to ASH1L inhibitors, our results suggest that ASH1L's leukemogenic function depends on activation of *Hoxa9* and *Meis1*. We further showed that the ASH1L SET domain in particular

plays a key role in activating *Hox* gene expression and driving proliferation of leukemic cells, a highly significant finding given that ASH1L is a large epigenetic regulator with multiple chromatin-interacting domains. Our finding that ASH1L regulates leukemic cell growth in CALM-AF10- and MLL-AF9-driven leukemia suggests that ASH1L may play an important role in leukemias with high *Hox* expression.

Future studies will be needed to investigate the details of ASH1L's role in leukemia. Given that ASH1L KMTase activity is important for *Hox* gene activation, we hypothesize that ASH1L activates *Hox* genes by methylating H3K36 at *Hox* loci. We failed to detect a change in H3K36 dimethylation globally by Western blot upon treatment with ASH1L inhibitors (data not shown), likely due to the presence of multiple H3K36-specific KMTases. However, in the near future we will more thoroughly investigate ASH1L's effect on chromatin modifications by performing chromatin immunoprecipitation (ChIP) experiments with ASH1L inhibitors.

Another key topic for future research is the relationship between ASH1L and MLL in leukemia, as a partnership between ASH1L and MLL during development is conserved from flies to mammals. ASH1 and the MLL homolog TRX are members of the *trithorax* (TrxG) group of epigenetic activators that maintain *Hox* gene expression during *Drosophila* development. ASH1 and TRX interact genetically, colocalize on salivary gland polytene chromosomes, and coimmunoprecipitate each other in embryonic nuclear extracts [212,213]. In mammals, ASH1L and MLL cooperate to activate *Hox* genes and maintain quiescence of long-term hematopoietic stem cells [15,198,214,215].

MLL has a well-characterized role in leukemia, as MLL fusion proteins drive disease in 10% of human acute leukemia [216], and upregulation of *Hox* genes is essential for MLL's leukemogenic function [168]. Although ASH1L's role in leukemia was not previously

characterized, our work demonstrates that ASH1L regulates leukemic cell proliferation and activates *Hox* genes, similar to MLL. Further biochemical studies are necessary to determine the nature of the relationship between ASH1L and MLL in leukemia. In the near-term, we are particularly interested whether combination of ASH1L inhibitors with menin-MLL inhibitors [217] or WDR5-MLL inhibitors [204] will result in more potent effects on leukemic cells than either treatment alone. We envision three simple models of the ASH1L and MLL relationship that would affect the outcome of combination therapy. In the first scenario, ASH1L and MLL participate in different pathways that independently contribute to target gene activation (**Figure 5.2, scenario 1**). In this case, combination of ASH1L and MLL-targeted inhibitors would be expected to have an approximately additive effect on target gene expression. In the second scenario, ASH1L and MLL participate in overlapping pathways that together cause a much greater activation of target genes than either ASH1L or MLL alone (**Figure 5.2, scenario 2**). For example, MLL-mediated H3K4 methylation could cooperate with ASH1L-mediated H3K36 methylation to cause a marked upregulation of transcription. In this scenario, inhibiting ASH1L or MLL individually would be nearly as effective as inhibiting both proteins. In the third scenario, ASH1L and MLL are each individually capable of activating their target genes to near-maximal levels (**Figure 5.2, scenario 3**). In this scenario, combining ASH1L and MLL inhibitors would have a synergistic effect on downregulating target gene expression. We will perform combination therapy experiments in the near future to test these different models.

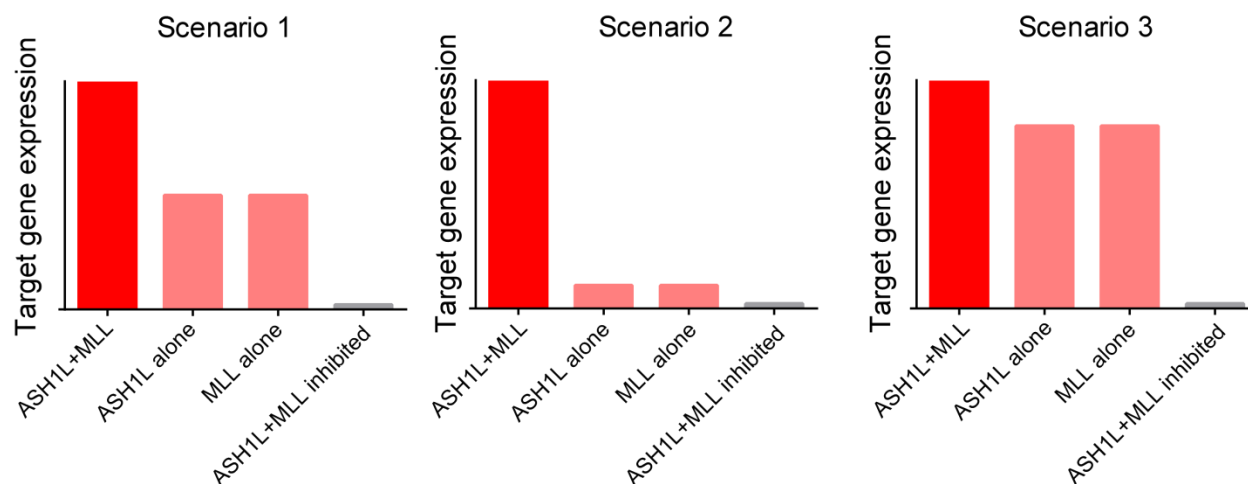


Figure 5.2. Three models of the relationship between ASH1L and MLL in activating target genes. In scenario 1, ASH1L and MLL function in separate, independent pathways that each contribute to target gene activation. In this case, combining ASH1L and MLL inhibitors would be expected to have an approximately additive effect. In scenario 2, ASH1L and MLL cooperate to achieve a marked increase in target gene expression, but the individual proteins are poorly effective. In this case, combining ASH1L and MLL inhibitors would not have a significantly greater effect than either inhibitor alone. In scenario 3, both ASH1L and MLL are individually capable of near-maximal target gene activation. In this case, combining ASH1L and MLL inhibitors would have a synergistic effect on decreasing target gene expression.

Future directions

This work demonstrates the successful development of potent and cell-active ASH1L inhibitors using an FBDD approach. As outlined above, many exciting opportunities remain for further enhancing the potency and optimizing the pharmacokinetic properties of the ASH1L inhibitors we developed. Nonetheless, the inhibitors presented here already represent valuable tools to investigate many questions surrounding ASH1L's function in cancer and development. ASH1L inhibitors could be used to identify the transcriptional targets of ASH1L on a genome-wide scale, to elucidate the *in vivo* methylation targets of ASH1L, and to determine in what cell types and developmental stages ASH1L is active. Further biochemical studies will also be needed to determine the mechanism whereby ASH1L is recruited to its target loci and what co-activators cooperate with ASH1L. In addition to cancer, ASH1L has been implicated in the pathogenesis of facioscapulohumeral muscular dystrophy and liver fibrosis [86,87], and ASH1L

inhibitors could be used to investigate these diseases. We predict that ASH1L inhibitors will facilitate many important discoveries surrounding ASH1L's role in biology.

Chapter 6. Materials and Methods

ASH1L constructs

ASH1L SET (amino acids 2069–2288), ASH1L SET-PHD (amino acids 2069–2636), and ASH1L SET-BAH (amino acids 2069–2833) were cloned from the full-length human ASH1L cDNA. ASH1L N-SET (amino acids 2003–2303) was codon-optimized for expression in *Escherichia coli*, and the DNA was purchased from Life Technologies. Point mutations were made in SET and SET-BAH constructs by polymerase chain reaction (PCR) and confirmed by DNA sequencing.

Protein purification

ASH1L SET and N-SET proteins were expressed as MOCR fusion proteins in *E. coli* BL21(DE3) T1R cells at 22 °C. Transformed cells were lysed in buffer A containing 50 mM Tris (pH 7.5), 500 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP), and 20 mM imidazole. Cell debris was pelleted by centrifugation, and the supernatant was loaded on a column packed with nickel-nitrilotriacetic acid beads. The column was washed with buffer A and protein eluted with a 100 mL linear gradient up to buffer A containing 500 mM imidazole. The MOCR tag was cleaved with tobacco etch virus (TEV) protease during overnight dialysis against 50 mM Tris (pH 7.5), 100 mM NaCl, and 1 mM TCEP. Cleaved ASH1L was isolated from MOCR by repeating the nickel column purification and collecting ASH1L in the flow-through and low-imidazole fractions. ASH1L was further purified by gel filtration chromatography using a Superdex-75 column running in buffer B containing 50 mM Tris (pH 7.5), 100 mM NaCl, and 1 mM TCEP. ASH1L SET-PHD and SET-BAH proteins were purified similarly, with the

following differences. Expression was performed at 18 °C; cleavage with TEV and the second nickel column were omitted to maintain protein stability, and gel filtration was performed on a Superdex-200 column.

NMR studies

¹⁵N-¹³C ASH1L SET was prepared at 300 μM concentration in buffer containing 50 mM HEPES pH 7.5, 50 mM NaCl, 1 mM TCEP, and 300 μM S-adenosyl methionine (SAM). HSQC Spectra were acquired at 25 and 30°C with a Bruker Avance III spectrometer equipped with cryoprobe, running Topspin version 2.1. Backbone assignment was performed using triple resonance experiments: HNCA, HN(CO)CA, HNCACB, and CBCA(CO)NH. For analysis of ASH1L mutants, wild-type and mutant ASH1L SET proteins were prepared at 100 μM concentration in 50 mM Tris pH 7.5, 50 mM NaCl, 1 mM TCEP, 200 μM SAM, 5% DMSO, and 5% D₂O. ¹H-¹⁵N TROSY spectra were acquired at 30°C. All NMR processing and spectral visualization were performed using NMRPipe [218] and Sparky [219].

Chemical shift perturbations (Δ) in ¹⁵N-¹H TROSY spectra caused by mutations compared to wild-type were calculated as the square root of the sum of the squares of the differences in the ¹H and ¹⁵N chemical shifts [220],

$$\Delta = [(\delta_H \times 600)^2 + (\delta_N \times 60.8)^2]^{1/2} \quad (1)$$

where δ_H and δ_N are the chemical shift differences in ppm between mutant and wild-type for ¹H and ¹⁵N, respectively. Global structural perturbations caused by mutations were determined by calculating the sum of Δ for all amides (Δ_{sum}).

NMR fragment screening

¹⁵N-labeled ASH1L SET was prepared at a 100 μM concentration in 50 mM Tris pH 7.5, 50 mM NaCl, 1 mM TCEP, and 200 μM SAM. Approximately 1500 fragment-like ligands from

an internal collection in Dr. Cierpicki's lab were screened in 10 or 20-compound mixtures at either 250 or 500 μM concentration in a final DMSO concentration of 5%. TROSY spectra were recorded at 30°C. Compound mixtures causing chemical shift perturbations compared to DMSO were deconvoluted to identify ASH1L ligands.

NMR-based 3PA Measurements

The 3PA measurement for measuring relative binding affinity of compounds was calculated by taking the average of the chemical shift perturbations in hertz at three ^{15}N - ^1H TROSY peaks (G2163, K2228, and I2278) in the presence of 500 μM compound. Chemical shift perturbations caused by compound binding were calculated as in equation 1 above.

KMTase assay

Chicken mono/dinucleosomes (HMT-35-179), chicken oligo nucleosomes (HMT-35-177), and HeLa nucleosomes (HMT-35-123) were purchased from Reaction Biology. Recombinant nucleosomes were purified in house as described previously [221]. For testing different ASH1L constructs and nucleosome substrates, ASH1L (0.25 μM) was incubated with 0.7 μM SAM and 0.2 or 0.4 μM nucleosome in HMTase buffer containing 50 mM Tris (pH 8.5), 25 mM NaCl, 2 mM MgCl_2 , and 1 mM DTT in a total volume of 25 μL for 1 h at 30 °C. For assays with ASH1L mutants, ASH1L SET-BAH WT and mutant proteins (0.25 μM) were incubated with 20 μM SAM (5% radiolabeled [^3H]SAM) and 1.7 μM chicken mono/dinucleosomes in HMTase buffer in a total volume of 10 μL for 1 h at 30 °C. For testing compounds, ASH1L SET-BAH (0.25 μM) was incubated with 0.7 μM SAM, 0.2 μM chicken mono/dinucleosomes, and the compound in a concentration range from 500 to 0.2 μM in HMTase buffer in a total volume of 15 μL for 1 hr at 30 °C. For testing specificity of ASH1L inhibitors, assay conditions were identical except NSD1, NSD2, NSD3, SETD2, and EHMT2

(0.2 μ M) were substituted for ASH1L SET-BAH. The reactions were stopped by spotting 5 μ L of the reaction mixture on P81 cellulose squares (Reaction Biology). The P81 squares were dried for 45 min and washed five times with 50 mM sodium bicarbonate (pH 9.0), 10 min per wash. The P81 squares were then dried for 1 h, added to 10 mL of Ultima Gold scintillation cocktail (PerkinElmer), and analyzed using a Beckman scintillation counter.

Isothermal titration calorimetry

ASH1L SET was extensively dialyzed at 4°C against ITC buffer (50 mM sodium phosphate pH 7.5, 50 mM NaCl, 1 mM TCEP). Compounds were dissolved in DMSO and diluted with ITC buffer to final concentrations of 0.1-1 mM in 5% DMSO. The protein solution was adjusted to 5% DMSO final concentration. Both protein and compound solutions were adjusted to 50-100 μ M SAM, to maintain ASH1L stability. The titrations were performed using a VP-ITC titration calorimetric system (MicroCal) at 25°C. The calorimetric cell, containing ASH1L (concentration ranging from 10-50 μ M) was titrated with the compounds (concentration ranging from 0.1-1 mM) injected in 10 μ l volumes. Data was analyzed using Origin 7.0 (OriginLab) to obtain thermodynamic parameters.

Crystallization and structure determination

ASH1L wild-type and mutant SET domain proteins in 50 mM Tris (pH 7.5), 100 mM NaCl, and 1 mM TCEP were concentrated to 10 mg/mL. Crystals of ASH1L SET domain wild-type as well as S2259M, K2264L, and H2193F mutants were obtained using the sitting drop method in 20 mM Tris (pH 7.5) and 25% PEG3350 at 4 °C. Q2265A crystals were obtained using the sitting drop method in 0.1 M Tris (pH 8.5), 0.2 M MgCl₂, and 30% PEG4000 at 17 °C. For cryoprotection, crystals of WT, S2259M, K2264L, and H2193F were soaked in a

crystallization solution containing 20% glycerol, while Q2265A crystals were soaked in a crystallization solution containing 25% PEG400.

For crystallization of ASH1L Q2265A with compounds, 7 mg/ml protein was mixed with equimolar compound in a final DMSO concentration of 5%. Crystal seeds were produced by vortexing ASH1L Q2265A crystals with a Seed Bead (Hampton). Crystals were obtained using the sitting drop method by adding the ASH1L Q2265A/compound complex solution 1:1 with crystal seeds in 0.1 M Tris (pH 8.5), 0.2 M MgCl₂, and 30% PEG4000 and incubating at 17 °C.

All data were collected under cryogenic conditions at Life Sciences-Collaborative Access Team beamlines 21ID-D, -F, and -G at the Advanced Photon Source at Argonne National Laboratory (Argonne, IL). Data were processed with HKL2000 [222]. Structures were determined by molecular replacement using MOLREP [223] with the wild-type ASH1L structure [Protein Data Bank (PDB) entry 3OPE] as a search model in molecular replacement. Models were built and refined using REFMAC5 [224], Coot [225], and the CCP4 package [226]. Validation of structures was performed using MolProbity [227].

Average crystallographic B factors per residue were calculated as the average of the B-factors for all the atoms of each residue. Then the residue B factors were normalized using the ‘z-score normalization’ [228],

$$B_{x-zscore(i)} = [B_{x(i)} - \langle B \rangle_{(i)}] / s_{(i)}, \quad (2)$$

where $B_{x-zscore(i)}$ is the normalized z-score for residue x in structure i , $B_{x(i)}$ is the B-factor for residue x , $\langle B \rangle_{(i)}$ is the average residue B-factor for structure i , and $s_{(i)}$ is the corresponding standard deviation.

Cell culture

MDA-MB-468 cells were cultured in DMEM supplemented with 10% FBS, MEM non-essential amino acids (Thermo Fisher), and 1% penicillin/streptomycin. MCF-7 cells were cultured in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin. MLL-AF9 (MA9) transformed murine bone marrow cells (BMCs) were prepared as described previously [229]. MA9 BMCs were cultured in Iscove's modified Dulbecco's medium (IMDM) supplemented with 15% FBS (Stem Cell Technologies), 10 ng/ml IL-3, and 1% P/S. MV4;11 and MOLM-13 cells were cultured in RPMI-1640 medium with 10% FBS and 1% penicillin/streptomycin. THP-1 cells were cultured in RPMI-1640 medium with 10% FBS and 1% penicillin/streptomycin. KG1 cells were cultured in IMDM with 20% FBS and 1% penicillin/streptomycin.

Cell viability assays

MA9 and HM2 cells were plated at 1×10^5 cells/ml in 24-well plates, treated with 0.25% DMSO or compounds and cultured at 37°C for 12 days. Every four days, the volume corresponding to 1×10^5 cells of DMSO-treated cells was spun down and resuspended in fresh media with fresh compound. At day 0 and each four day interval, 100 μ l aliquots of the cell suspension were transferred to 96-well plates in quadruplicates. The quadruplicate samples were incubated for 4 days at 37°C, and then an MTT cell proliferation assay kit (Roche) was used to measure viable cells. Absorbance was read at 570 nm using a PHERAstar (BMG) microplate reader.

Quantitative RT-PCR

Total RNA was extracted from cells using the RNeasy mini kit (QIAGEN), and then 100-2000 ng of total RNA was reverse transcribed using the High Capacity cDNA Reverse

Transcription Kit (Applied Biosystems) according to the manufacturer's protocol. Real-time PCR was performed using a CFX96 Real-Time PCR Detection System (Biorad). TaqMan Gene Expression Master Mix and TaqMan Gene Expression Assays for human *HPRT1* (Hs02800695), human *ASH1L* (Hs00218516), human *HOXA5* (Hs00430330), human *HOXA9* (Hs00365956), human *HOXA10* (Hs00172012), human *MEIS1* (Hs00180020), mouse *Gapdh* (Mm99999915), mouse *Ash1l* (Mm00467322), mouse *Hoxa7* (Mm00657963), mouse *Hoxa9* (Mm00439364), mouse *Hoxa10* (Mm00433966), mouse *Meis1* (Mm00487664), and mouse *B-actin* (Mm00607939) were purchased from Thermo Fisher. Relative quantification of each gene transcript was carried out using the $\Delta\Delta C_t$ method as described in the Biorad Real-Time PCR Applications Guide.

Cytospin/Wright-Giemsa staining

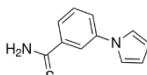
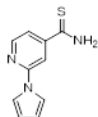
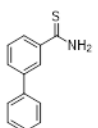
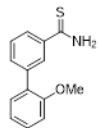
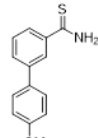
1 x 10⁵ MA9 mouse BMCs treated with compounds were harvested and placed in a Shandon EZ Single Cytofunnel (Thermo Fisher). Samples were centrifuged at 600 rpm for 5 min. The slides were air dried before staining with a Hema-3 kit (Thermo Fisher).

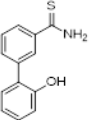
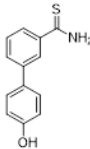
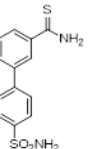
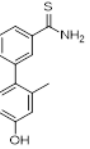
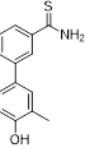
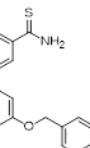
Expression of CD11b and apoptosis

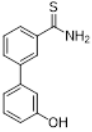
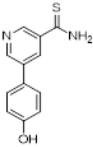
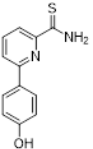
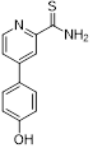
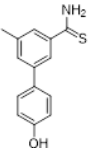
For CD11b expression analysis, 1.2 x 10⁵ MA9 mouse BMCs treated with compounds were harvested and washed with FACS buffer (PBS, 1% FBS, 0.1% NaN₃). Cells were resuspended in 100 μ l FACS buffer and incubated with 1 μ l Pacific Blue rat anti-mouse CD11b antibody (BioLegend cat# 101224) at 4°C for 30 min. Cells were then washed and analyzed by flow cytometry. For analysis of apoptosis and necrosis, cells were washed with Annexin V binding buffer and incubated with 4 μ l Annexin V-FITC (BD Biosciences) and 6 μ l propidium iodide (1mg/ml, Sigma-Aldrich) at 25°C for 10 min before being analyzed by flow cytometry.

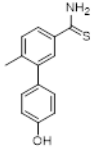
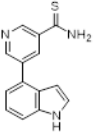
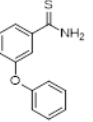
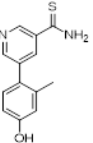
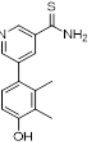
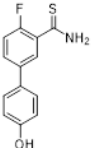
Appendix. Selected ASH1L inhibitors synthesized to date

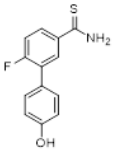
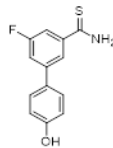
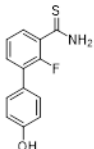
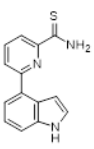
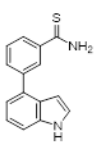
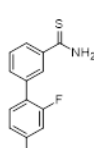
Table A.1. Selected ASH1L inhibitors synthesized to date.

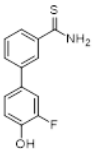
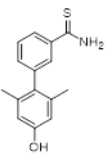
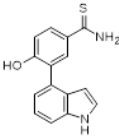
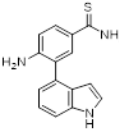
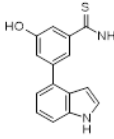
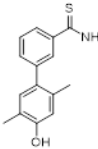
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μ M cmpd	Enz. Assay IC ₅₀ Trial 1 (μ M)	Enz. Assay IC ₅₀ Trial 2 (μ M)	Enz. Assay IC ₅₀ Mean \pm St dev. (μ M)	K _d by ITC (μ M)
	12B5	1	202.3	7.76	N/A				
	BD595		203.3	n.b.	N/A				
	BD597	1a	213.3	25.38	N/A				
	BD605		243.3	16.48	N/A				
	BD606		243.3	3.63	N/A				

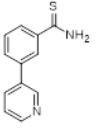
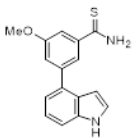
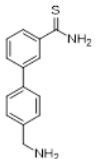
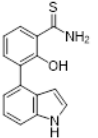
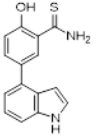
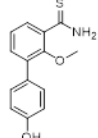
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μM compd	Enz. Assay IC ₅₀ Trial 1 (μM)	Enz. Assay IC ₅₀ Trial 2 (μM)	Enz. Assay IC ₅₀ Mean±St dev. (μM)	K _d by ITC (μM)
	BD608		229.3	16.30	N/A				
	BD609	1b	229.3	29.76	N/A				
	BD610		292.4	29.26	N/A				
	BD632	1c	243.3	48.71	N/A				135
	BD633	1d	243.3	36.31	N/A				
	BD634		319.4	n.b.	N/A				

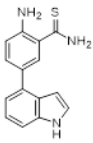
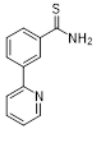
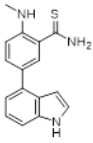
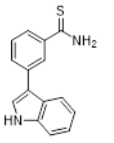
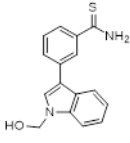
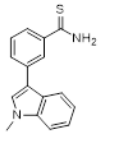
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μM compd	Enz. Assay IC ₅₀ Trial 1 (μM)	Enz. Assay IC ₅₀ Trial 2 (μM)	Enz. Assay IC ₅₀ Mean±St dev. (μM)	K _d by ITC (μM)
	BD635		229.3	18.14	N/A				
	BD638	1e	230.29	2.07	N/A				
	BD639	1g	230.29	40.45	N/A				
	BD640	1f	230.29	n.b.	N/A				
	BD660	1i	243.32	21.08	N/A				

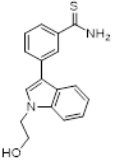
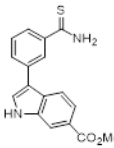
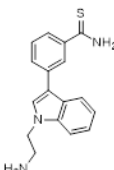
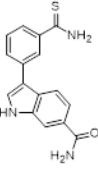
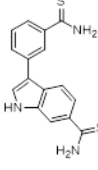
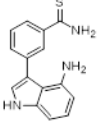
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μM compd	Enz. Assay IC ₅₀ Trial 1 (μM)	Enz. Assay IC ₅₀ Trial 2 (μM)	Enz. Assay IC ₅₀ Mean±St dev. (μM)	K _d by ITC (μM)
	BD661	1h	243.32	16.35	N/A				
	BD673		253.32	7.70	N/A				
	BD685		229.3	5.57	N/A				
	BD697		244.31	n.b.	N/A				
	BD698		258.34	n.b.	N/A				
	BD699		247.29	7.05	N/A				

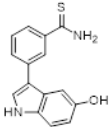
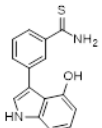
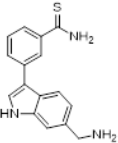
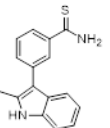
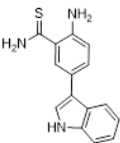
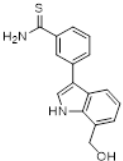
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μM compd	Enz. Assay IC ₅₀ (μM) Trial 1	Enz. Assay IC ₅₀ (μM) Trial 2	Enz. Assay IC ₅₀ Mean±St dev. (μM)	K _d by ITC (μM)
	BD700		247.29	23.28	N/A				
	BD701		247.29	19.75	N/A				
	BD702		247.29	n.b.	N/A				
	BD710		253.32	50.60	N/A				
	BD720	2	252.34	N/A	30%	76			
	BD725		247.29	43.93	N/A				

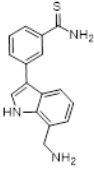
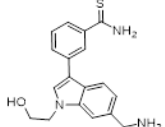
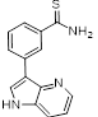
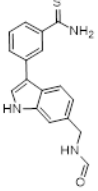
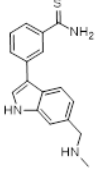
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μ M compd	Enz. Assay IC ₅₀ Trial 1 (μ M)	Enz. Assay IC ₅₀ Trial 2 (μ M)	Enz. Assay IC ₅₀ Mean \pm St dev. (μ M)	K _d by ITC (μ M)
	BD726		247.29	31.15	N/A				
	BD727		257.35	4.73	N/A				
	BD748		282.36	6.00	N/A				
	BD749		281.38	6.12	N/A				
	BD756		268.33	44.94	N/A				
	BD757		257.35	38.53	N/A				

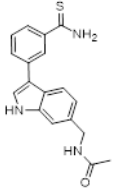
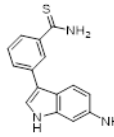
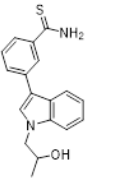
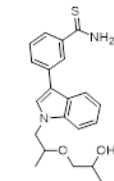
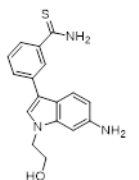
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μ M compd	Enz. Assay IC ₅₀ Trial 1 (μ M)	Enz. Assay IC ₅₀ Trial 2 (μ M)	Enz. Assay IC ₅₀ Mean \pm St dev. (μ M)	K _d by ITC (μ M)
	BD759		214.29	3.67	N/A				
	BD770		282.36	9.15	N/A				
	BD797		242.3	23.23	N/A				
	BD818		268.33	7.32	N/A				
	BD819		268.33	N/A	20%				
	BD840		267.35	n.b.	N/A				

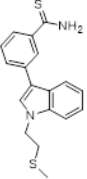
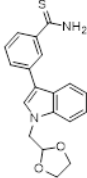
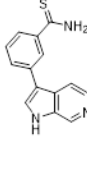
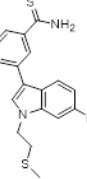
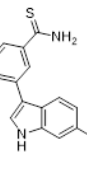
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μM compd	Enz. Assay IC ₅₀ (μM) Trial 1	Enz. Assay IC ₅₀ (μM) Trial 2	Enz. Assay IC ₅₀ Mean±St dev. (μM)	K _d by ITC (μM)
	BD842		259.32	N/A	10%				
	BD845		214.29	11.47	N/A				
	BD848		281.38	8.23	N/A				
	BD944	3	252.34	N/A	50%	100			
	BD954		282.36	N/A	90%	22			12.1
	BD955		266.36	N/A	80%				

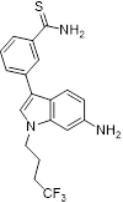
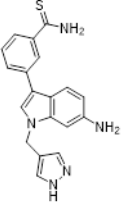
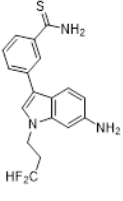
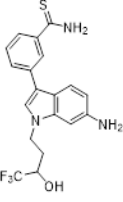
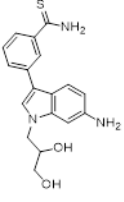
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μ M compd	Enz. Assay IC ₅₀ Trial 1 (μ M)	Enz. Assay IC ₅₀ Trial 2 (μ M)	Enz. Assay IC ₅₀ Mean \pm St dev. (μ M)	K _d by ITC (μ M)
	BDY-19	4	296.39	N/A	90%	22			
	BDY-20		310.37	N/A	50%				
	BD990		295.4	N/A	50%	26			
	BDY-29		295.36	p.s.	N/A				
	BDY-33		311.42	N/A	50%	86			
	BDY-64		267.35	w.b.	N/A				

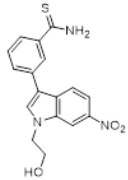
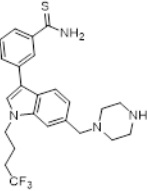
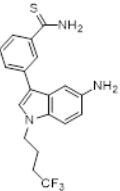
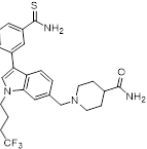
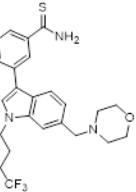
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μ M compd	Enz. Assay IC ₅₀ (μ M) Trial 1	Enz. Assay IC ₅₀ (μ M) Trial 2	Enz. Assay IC ₅₀ Mean \pm St dev. (μ M)	K _d by ITC (μ M)
	BDY-66	SK-nc	268.33	w.b.	N/A				
	BDY-67		268.33	w.b.	N/A				
	BDY-73	5	281.38	N/A	70%	36			
	BDY-78		266.36	N/A	20%				
	DA-24		267.35	w.b.	N/A				
	DA-51		282.36	N/A	50%	136			

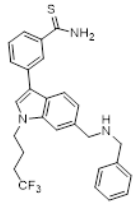
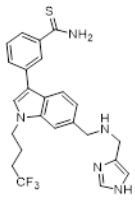
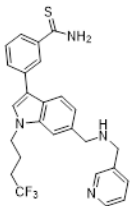
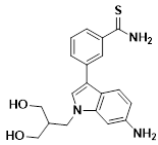
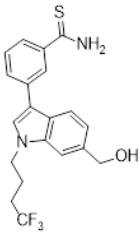
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μM compd	Enz. Assay IC ₅₀ (μM) Trial 1	Enz. Assay IC ₅₀ (μM) Trial 2	Enz. Assay IC ₅₀ Mean±St dev. (μM)	K _d by ITC (μM)
	DA-53		281.38	N/A	50%				
	BD1104	6	325.43	N/A	90%	13			4.6
	BD1132		253.32	w.b.	N/A				
	BD1143		309.39	N/A	80%	42			
	BD1151		295.4	N/A	80%	36			

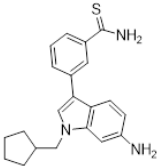
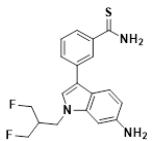
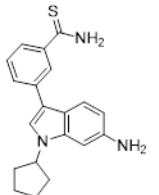
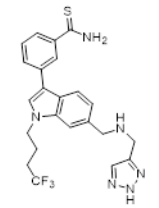
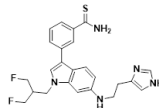
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μ M compd	Enz. Assay IC ₅₀ Trial 1 (μ M)	Enz. Assay IC ₅₀ Trial 2 (μ M)	Enz. Assay IC ₅₀ Mean \pm St dev. (μ M)	K _d by ITC (μ M)
	BD1155		323.41	N/A	80%	29			
	DJM10		267.35	N/A	100%	45			
	DJM7A		310.42	N/A	50%	43			
	DJM7B		368.5	N/A	50%				
	DJM23	7	311.4	N/A	100%	22.7			4.0

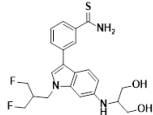
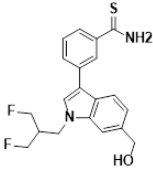
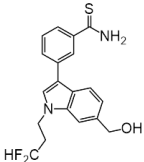
Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μ M compd	Enz. Assay IC ₅₀ Trial 1 (μ M)	Enz. Assay IC ₅₀ Trial 2 (μ M)	Enz. Assay IC ₅₀ Mean \pm St dev. (μ M)	K _d by ITC (μ M)
	DJM29		326.48	p.s.	N/A				
	DJM30		338.43	N/A	60%				
	DJM26		268.34	w.b.	N/A				
	DJM34		341.49	N/A	100%	15.3			
	DJM44		281.38	N/A	80%				

Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μ M compd	Enz. Assay IC ₅₀ Trial 1 (μ M)	Enz. Assay IC ₅₀ Trial 2 (μ M)	Enz. Assay IC ₅₀ Mean \pm St dev. (μ M)	K _d by ITC (μ M)
	DJM55	8	377.43	N/A	100%	7.5			3.4
	DJM58		347.44	N/A	100%	33			
	DJM69	9	345.41	N/A	100%	3.5	3.7	3.6 \pm 0.1	2.0
	DJM78		393.43	N/A	90%	28			3.7
	DJM-94		341.43	N/A	100%	11			14

Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μ M compd	Enz. Assay IC ₅₀ Trial 1 (μ M)	Enz. Assay IC ₅₀ Trial 2 (μ M)	Enz. Assay IC ₅₀ Mean \pm St dev. (μ M)	K _d by ITC (μ M)
	DJM-97		341.39	N/A	60%				
	JJ-158		460.56	N/A	40%				
	JJ-159		377.43	N/A	50%				
	JJ-164		502.6	N/A	80%				
	JJ-165		461.55	N/A	30%				

Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μ M compd	Enz. Assay IC ₅₀ Trial 1 (μ M)	Enz. Assay IC ₅₀ Trial 2 (μ M)	Enz. Assay IC ₅₀ Mean \pm St dev. (μ M)	K _d by ITC (μ M)
	JJ-161		481.58	N/A	100%	4.1			
	JJ-179	10	471.55	N/A	100%	4.1	4.7	4.4 \pm 0.3	2.1
	JJ-185		482.57	N/A	100%	6.5			
	SKA-125		355.46	N/A	50%	34			
	JJ-186	11	392.44	N/A	100%	2.0	5.6	3.8 \pm 1.8	2.1

Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: % sat. at 100 μ M compd	Enz. Assay IC ₅₀ Trial 1 (μ M)	Enz. Assay IC ₅₀ Trial 2 (μ M)	Enz. Assay IC ₅₀ Mean \pm St dev. (μ M)	K _d by ITC (μ M)
	JJ-189		349.5	N/A	80%	34			
	SKA-137	SK-1	359.44	N/A	100%	2.7	6.4	4.6 \pm 1.9	1.7
	JJ-195		335.47	N/A	100%	7.8			
	JJ-197		473.52	N/A	100%	5.4			2.5
	SKM-144		453.56	N/A	100%	5.6	6.7	6.2 \pm 0.6	1.4

Structure	Lab ID	Text ID	MW	NMR fast ex.: 3PA (Hz) ^a	NMR slow ex.: sat. at 100 μM cmpd	Enz. Assay IC ₅₀ Trial 1 (μM)	Enz. Assay IC ₅₀ Trial 2 (μM)	Enz. Assay IC ₅₀ Mean±St dev. (μM)	K _d by ITC (μM)
	SKA-153		433.52	N/A	100%	5.3			
	SKA-174	12	374.45	N/A	100%				1.1
	JJ-228	13	360.42	N/A	100%				0.9

^an.b. denotes no binding, w.b. denotes weak binding, and p.s. denotes poorly soluble.

References

1. Greer EL, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* 13(5), 343–57 (2012).
2. Herz H-M, Garruss A, Shilatifard A. SET for life: biochemical activities and biological functions of SET domain-containing proteins. *Trends Biochem. Sci.* 38(12), 621–39 (2013).
3. Barski A, Cuddapah S, Cui K, *et al.* High-resolution profiling of histone methylations in the human genome. *Cell.* 129(4), 823–37 (2007).
4. Vermeulen M, Mulder KW, Denissov S, *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell.* 131(1), 58–69 (2007).
5. Mozzetta C, Boyarchuk E, Pontis J, Ait-Si-Ali S. Sound of silence: the properties and functions of repressive Lys methyltransferases. *Nat. Rev. Mol. Cell Biol.* 16(8), 499–513 (2015).
6. Shen C, Ipsaro JJ, Shi J, *et al.* NSD3-Short Is an Adaptor Protein that Couples BRD4 to the CHD8 Chromatin Remodeler. *Mol. Cell.* 60(6), 847–859 (2015).
7. Sankaran SM, Wilkinson AW, Gozani O. A PWWP domain of histone-lysine N-methyltransferase NSD2 binds to dimethylated Lys36 of histone H3 and regulates NSD2 function at chromatin. *J. Biol. Chem.* (2016).
8. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. *Cell.* 150(1), 12–27 (2012).
9. Daigle SR, Olhava EJ, Therkelsen CA, *et al.* Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. *Cancer Cell.* 20(1), 53–65 (2011).
10. Kuo AJ, Cheung P, Chen K, *et al.* NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol. Cell.* 44(4), 609–20 (2011).
11. McCabe MT, Creasy CL. EZH2 as a potential target in cancer therapy. *Epigenomics.* 6(3), 341–51 (2014).
12. Waters NJ, Smith SA, Olhava EJ, *et al.* Metabolism and disposition of the DOT1L

- inhibitor, pinometostat (EPZ-5676), in rat, dog and human. *Cancer Chemother. Pharmacol.* 77(1), 43–62 (2016).
13. Kuntz KW, Campbell JE, Keilhack H, *et al.* The Importance of Being Me: Magic Methyls, Methyltransferase Inhibitors, and the Discovery of Tazemetostat. *J. Med. Chem.* 59(4), acs.jmedchem.5b01501 (2016).
 14. Gross S, Rahal R, Stransky N, Lengauer C, Hoeflich KP. Targeting cancer with kinase inhibitors. *J. Clin. Invest.* 125(5), 1780–9 (2015).
 15. Jones M, Chase J, Brinkmeier M, *et al.* Ash1l controls quiescence and self-renewal potential in hematopoietic stem cells. *J. Clin. Invest.* 125(5), 2007–2020 (2015).
 16. Wang GG, Cai L, Pasillas MP, Kamps MP. NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nat. Cell Biol.* 9(7), 804–12 (2007).
 17. Pei H, Zhang L, Luo K, *et al.* MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. *Nature.* 470(7332), 124–8 (2011).
 18. Sein H, Väriv S, Kristjuhan A. Distribution and maintenance of histone H3 lysine 36 trimethylation in transcribed locus. *PLoS One.* 10(3), e0120200 (2015).
 19. Rao B, Shibata Y, Strahl BD, Lieb JD. Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Mol. Cell. Biol.* 25(21), 9447–59 (2005).
 20. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* 28(8), 817–25 (2010).
 21. Sen P, Dang W, Donahue G, *et al.* H3K36 methylation promotes longevity by enhancing transcriptional fidelity. *Genes Dev.* 29(13), 1362–76 (2015).
 22. Kizer KO, Phatnani HP, Shibata Y, Hall H, Greenleaf AL, Strahl BD. A Novel Domain in Set2 Mediates RNA Polymerase II Interaction and Couples Histone H3 K36 Methylation with Transcript Elongation. *Mol. Cell. Biol.* 25(8), 3305–3316 (2005).
 23. Carrozza MJ, Li B, Florens L, *et al.* Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell.* 123(4), 581–92 (2005).
 24. Venkatesh S, Smolle M, Li H, *et al.* Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature.* 489(7416), 452–5 (2012).
 25. Smolle M, Venkatesh S, Gogol MM, *et al.* Chromatin remodelers Isw1 and Chd1 maintain

- chromatin structure during transcription by preventing histone exchange. *Nat. Struct. Mol. Biol.* 19(9), 884–92 (2012).
26. Carvalho S, Raposo AC, Martins FB, *et al.* Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. *Nucleic Acids Res.* 41(5), 2881–93 (2013).
 27. Xie L, Pelz C, Wang W, *et al.* KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. *EMBO J.* 30(8), 1473–84 (2011).
 28. Fang R, Barbera AJ, Xu Y, *et al.* Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. *Mol. Cell.* 39(2), 222–33 (2010).
 29. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. *Oncogene.* 33(46), 5311–5318 (2013).
 30. Schwartz S, Meshorer E, Ast G. Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* 16(9), 990–5 (2009).
 31. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifications. *Science.* 327(5968), 996–1000 (2010).
 32. Pradeepa MM, Sutherland HG, Ule J, Grimes GR, Bickmore WA. Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet.* 8(5), e1002717 (2012).
 33. Guo R, Zheng L, Park JW, *et al.* BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. *Mol. Cell.* 56(2), 298–310 (2014).
 34. Kim S, Kim H, Fong N, Erickson B, Bentley DL. Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci.* 108(33), 13564–13569 (2011).
 35. de Almeida SF, Grosso AR, Koch F, *et al.* Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat. Struct. Mol. Biol.* 18(9), 977–983 (2011).
 36. Pai C-C, Deegan RS, Subramanian L, *et al.* A histone H3K36 chromatin switch coordinates DNA double-strand break repair pathway choice. *Nat. Commun.* 5, 4091 (2014).
 37. Jha DK, Strahl BD. An RNA polymerase II-coupled function for histone H3K36

- methylation in checkpoint activation and DSB repair. *Nat. Commun.* 5, 3965 (2014).
38. Pfister SX, Ahrabi S, Zalmas L-P, *et al.* SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. *Cell Rep.* 7(6), 2006–18 (2014).
 39. Li F, Mao G, Tong D, *et al.* The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSa. *Cell.* 153(3), 590–600 (2013).
 40. Fnu S, Williamson EA, De Haro LP, *et al.* Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. *Proc. Natl. Acad. Sci.* 108(2), 540–545 (2010).
 41. Kim H-S, Kim S-K, Hromas R, Lee S-H. The SET Domain Is Essential for Metnase Functions in Replication Restart and the 5' End of SS-Overhang Cleavage. *PLoS One.* 10(10), e0139418 (2015).
 42. Carlson SM, Moore KE, Sankaran SM, Reynoird N, Elias JE, Gozani O. A Proteomic Strategy Identifies Lysine Methylation of Splicing Factor snRNP70 by the SETMAR Enzyme. *J. Biol. Chem.* 290(19), 12040–7 (2015).
 43. Dorigi KM, Tamkun JW. The trithorax group proteins Kismet and ASH1 promote H3K36 dimethylation to counteract Polycomb group repression in *Drosophila*. *Development.* 140(20), 4182–92 (2013).
 44. Yuan W, Xu M, Huang C, Liu N, Chen S, Zhu B. H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* 286(10), 7983–7989 (2011).
 45. Yuan G, Ma B, Yuan W, *et al.* Histone H2A ubiquitination inhibits the enzymatic activity of H3 lysine 36 methyltransferases. *J. Biol. Chem.* 288(43), 30832–30842 (2013).
 46. Cai L, Rothbart SB, Lu R, *et al.* An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Mol. Cell.* 49(3), 571–82 (2013).
 47. Popovic R, Martinez-Garcia E, Giannopoulou EG, *et al.* Histone methyltransferase MMSET/NSD2 alters EZH2 binding and reprograms the myeloma epigenome through global and focal changes in H3K36 and H3K27 methylation. *PLoS Genet.* 10(9), e1004566 (2014).
 48. Komatsu S, Ichikawa D, Hirajima S, *et al.* Overexpression of SMYD2 contributes to malignant outcome in gastric cancer. *Br. J. Cancer.* 112(2), 357–64 (2015).

49. Komatsu S, Imoto I, Tsuda H, *et al.* Overexpression of SMYD2 relates to tumor cell proliferation and malignant outcome of esophageal squamous cell carcinoma. *Carcinogenesis*. 30(7), 1139–46 (2009).
50. Ohtomo-Oda R, Komatsu S, Mori T, *et al.* SMYD2 overexpression is associated with tumor cell proliferation and a worse outcome in human papillomavirus-unrelated nonmultiple head and neck carcinomas. *Hum. Pathol.* 49, 145–55 (2016).
51. Sakamoto LHT, Andrade RV de, Felipe MSS, Motoyama AB, Pittella Silva F. SMYD2 is highly expressed in pediatric acute lymphoblastic leukemia and constitutes a bad prognostic factor. *Leuk. Res.* 38(4), 496–502 (2014).
52. Cho H-S, Hayami S, Toyokawa G, *et al.* RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation. *Neoplasia*. 14(6), 476–86 (2012).
53. Kurotaki N, Imaizumi K, Harada N, *et al.* Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat. Genet.* 30(4), 365–6 (2002).
54. Hersh JH, Cole TRP, Bloom AS, Bertolone SJ, Hughes HE. Risk of malignancy in Sotos syndrome. *J. Pediatr.* 120(4), 572–574 (1992).
55. Hollink IHIM, van den Heuvel-Eibrink MM, Arentsen-Peters STCJM, *et al.* NUP98/NSD1 characterizes a novel poor prognostic group in acute myeloid leukemia with a distinct HOX gene expression pattern. *Blood*. 118(13), 3645–56 (2011).
56. Ostronoff F, Othus M, Gerbing RB, *et al.* NUP98/NSD1 and FLT3/ITD coexpression is more prevalent in younger AML patients and leads to induction failure: a COG and SWOG report. *Blood*. 124(15), 2400–7 (2014).
57. Deshpande AJ, Deshpande A, Sinha AU, *et al.* AF10 Regulates Progressive H3K79 Methylation and HOX Gene Expression in Diverse AML Subtypes. *Cancer Cell*. 26(6), 896–908 (2014).
58. Lu T, Jackson MW, Wang B, *et al.* Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65. *Proc. Natl. Acad. Sci. U. S. A.* 107(1), 46–51 (2010).
59. Prasad S, Ravindran J, Aggarwal BB. NF-kappaB and cancer: how intimate is this relationship. *Mol. Cell. Biochem.* 336(1-2), 25–37 (2010).
60. Nakshatri H, Appaiah HN, Anjanappa M, *et al.* NF-κB-dependent and -independent

- epigenetic modulation using the novel anti-cancer agent DMAPT. *Cell Death Dis.* 6, e1608 (2015).
61. Lawrence MS, Sougnez C, Lichtenstein L, *et al.* Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature.* 517(7536), 576–582 (2015).
 62. Berdasco M, Ropero S, Setien F, *et al.* Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proc. Natl. Acad. Sci. U. S. A.* 106(51), 21830–5 (2009).
 63. Yan X-J, Xu J, Gu Z-H, *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat. Genet.* 43(4), 309–15 (2011).
 64. Bergemann AD, Cole F, Hirschhorn K. The etiology of Wolf-Hirschhorn syndrome. *Trends Genet.* 21(3), 188–95 (2005).
 65. Hajdu I, Ciccia A, Lewis SM, Elledge SJ. Wolf-Hirschhorn syndrome candidate 1 is involved in the cellular response to DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* 108(32), 13130–4 (2011).
 66. Hsiao K-Y, Mizzen CA. Histone H4 deacetylation facilitates 53BP1 DNA damage signaling and double-strand break repair. *J. Mol. Cell Biol.* 5(3), 157–65 (2013).
 67. Hartlerode AJ, Guan Y, Rajendran A, *et al.* Impact of histone H4 lysine 20 methylation on 53BP1 responses to chromosomal double strand breaks. *PLoS One.* 7(11), e49211 (2012).
 68. Mirabella F, Wu P, Wardell CP, *et al.* MMSET is the key molecular target in t(4;14) myeloma. *Blood Cancer J.* 3, e114 (2013).
 69. Hudlebusch HR, Santoni-Rugiu E, Simon R, *et al.* The histone methyltransferase and putative oncoprotein MMSET is overexpressed in a large variety of human tumors. *Clin. Cancer Res.* 17(9), 2919–33 (2011).
 70. Hudlebusch HR, Skotte J, Santoni-Rugiu E, *et al.* MMSET is highly expressed and associated with aggressiveness in neuroblastoma. *Cancer Res.* 71(12), 4226–35 (2011).
 71. Oyer JA, Huang X, Zheng Y, *et al.* Point mutation E1099K in MMSET/NSD2 enhances its methyltransferase activity and leads to altered global chromatin methylation in lymphoid malignancies. *Leukemia.* 28(1), 198–201 (2014).
 72. Jaffe JD, Wang Y, Chan HM, *et al.* Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat. Genet.* 45(11), 1386–91 (2013).

73. Min D-J, Ezponda T, Kim MK, *et al.* MMSET stimulates myeloma cell growth through microRNA-mediated modulation of c-MYC. *Leukemia*. 27(3), 686–694 (2012).
74. Asangani IA, Ateeq B, Cao Q, *et al.* Characterization of the EZH2-MMSET histone methyltransferase regulatory axis in cancer. *Mol. Cell*. 49(1), 80–93 (2013).
75. Kang H-B, Choi Y, Lee JM, *et al.* The histone methyltransferase, NSD2, enhances androgen receptor-mediated transcription. *FEBS Lett*. 583(12), 1880–6 (2009).
76. Yang P, Guo L, Duan ZJ, *et al.* Histone methyltransferase NSD2/MMSET mediates constitutive NF- κ B signaling for cancer cell proliferation, survival, and tumor growth via a feed-forward loop. *Mol. Cell. Biol*. 32(15), 3121–31 (2012).
77. Ezponda T, Popovic R, Shah MY, *et al.* The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial-mesenchymal transition and invasive properties of prostate cancer. *Oncogene*. 32(23), 2882–90 (2013).
78. Chen Y, McGee J, Chen X, *et al.* Identification of druggable cancer driver genes amplified across TCGA datasets. *PLoS One*. 9(5), e98293 (2014).
79. Li Y, Trojer P, Xu C-F, *et al.* The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J. Biol. Chem*. 284(49), 34283–34295 (2009).
80. Jacques-Fricke BT, Gammill LS. Neural crest specification and migration independently require NSD3-related lysine methyltransferase activity. *Mol. Biol. Cell*. 25(25), 4174–86 (2014).
81. Rosati R, La Starza R, Veronese A, *et al.* NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t(8;11)(p11.2;p15). *Blood*. 99(10), 3857–60 (2002).
82. Taketani T, Taki T, Nakamura H, Taniwaki M, Masuda J, Hayashi Y. NUP98-NSD3 fusion gene in radiation-associated myelodysplastic syndrome with t(8;11)(p11;p15) and expression pattern of NSD family genes. *Cancer Genet. Cytogenet*. 190(2), 108–12 (2009).
83. Kang D, Cho H-S, Toyokawa G, *et al.* The histone methyltransferase Wolf-Hirschhorn syndrome candidate 1-like 1 (WHSC1L1) is involved in human carcinogenesis. *Genes. Chromosomes Cancer*. 52(2), 126–39 (2013).
84. Yang Z-Q, Liu G, Bollig-Fischer A, Giroux CN, Ethier SP. Transforming properties of 8p11-12 amplified genes in human breast cancer. *Cancer Res*. 70(21), 8487–97 (2010).

85. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. Genome regulation by polycomb and trithorax proteins. *Cell*. 128(4), 735–45 (2007).
86. Cabianca DS, Casa V, Bodega B, *et al.* A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell*. 149(4), 819–31 (2012).
87. Perugorria MJ, Wilson CL, Zeybel M, *et al.* Histone methyltransferase ASH1 orchestrates fibrogenic gene transcription during myofibroblast transdifferentiation. *Hepatology*. 56(3), 1129–1139 (2012).
88. Tanaka Y, Kawahashi K, Katagiri ZI, Nakayama Y, Mahajan M, Kioussis D. Dual function of histone H3 lysine 36 methyltransferase ASH1 in regulation of hox gene expression. *PLoS One*. 6(11), e28171 (2011).
89. Shah N, Sukumar S. The Hox genes and their roles in oncogenesis. *Nat. Rev. Cancer*. 10(5), 361–371 (2010).
90. Andreeff M, Ruvolo V, Gadgil S, *et al.* HOX expression patterns identify a common signature for favorable AML. *Leukemia*. 22(11), 2041–7 (2008).
91. Faber J, Krivtsov A V, Stubbs MC, *et al.* HOXA9 is required for survival in human MLL-rearranged acute leukemias. *Blood*. 113(11), 2375–85 (2009).
92. Kroon E, Kros J, Thorsteinsdottir U, Baban S, Buchberg AM, Sauvageau G. Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *EMBO J*. 17(13), 3714–25 (1998).
93. Miyazaki H, Higashimoto K, Yada Y, *et al.* Ash1l Methylates Lys36 of Histone H3 Independently of Transcriptional Elongation to Counteract Polycomb Silencing. *PLoS Genet*. 9(11), e1003897 (2013).
94. Colamaio M, Puca F, Ragozzino E, *et al.* miR-142-3p downregulation contributes to thyroid follicular tumorigenesis by targeting ASH1L and MLL1. *J. Clin. Endocrinol. Metab*. 100(1), E59–E69 (2014).
95. Liu L, Kimball S, Liu H, Holowatyj A, Yang Z-Q. Genetic alterations of histone lysine methyltransferases and their significance in breast cancer [Internet]. *Oncotarget*. 6(4), 2466–2482.
96. Kanellopoulou C, Gilpatrick T, Kilaru G, *et al.* Reprogramming of Polycomb-Mediated Gene Silencing in Embryonic Stem Cells by the miR-290 Family and the

- Methyltransferase Ash11. *Stem cell reports*. (2015).
97. Edmunds JW, Mahadevan LC, Clayton AL. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J*. 27(2), 406–20 (2008).
 98. Luscan A, Laurendeau I, Malan V, *et al*. Mutations in SETD2 cause a novel overgrowth condition. *J. Med. Genet*. 51(8), 512–7 (2014).
 99. Al Sarakbi W, Sasi W, Jiang WG, Roberts T, Newbold RF, Mokbel K. The mRNA expression of SETD2 in human breast cancer: correlation with clinico-pathological parameters. *BMC Cancer*. 9, 290 (2009).
 100. Hakimi AA, Ostrovnaya I, Reva B, *et al*. Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators BAP1 and SETD2: a report by MSKCC and the KIRC TCGA research network. *Clin. Cancer Res*. 19(12), 3259–67 (2013).
 101. Dalglish GL, Furge K, Greenman C, *et al*. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature*. 463(7279), 360–3 (2010).
 102. Mar BG, Bullinger LB, McLean KM, *et al*. Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia. *Nat. Commun*. 5, 3469 (2014).
 103. Huang KK, McPherson JR, Tay ST, *et al*. SETD2 histone modifier loss in aggressive GI stromal tumours. *Gut*. (2015).
 104. Zhu X, He F, Zeng H, *et al*. Identification of functional cooperative mutations of SETD2 in human acute leukemia. *Nat. Genet*. 46(3), 287–93 (2014).
 105. Kandoth C, McLellan MD, Vandin F, *et al*. Mutational landscape and significance across 12 major cancer types. *Nature*. 502(7471), 333–9 (2013).
 106. Kovac M, Navas C, Horswell S, *et al*. Recurrent chromosomal gains and heterogeneous driver mutations characterise papillary renal cancer evolution. *Nat. Commun*. 6, 6336 (2015).
 107. Gerlinger M, Rowan AJ, Horswell S, *et al*. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med*. 366(10), 883–92 (2012).
 108. Xiang W, He J, Huang C, *et al*. miR-106b-5p targets tumor suppressor gene SETD2 to inactive its function in clear cell renal cell carcinoma [Internet]. *Oncotarget*. 6(6), 4066–

- 4079 (2015).
109. Kanu N, Grönroos E, Martinez P, *et al.* SETD2 loss-of-function promotes renal cancer branched evolution through replication stress and impaired DNA repair. *Oncogene*. 34(46), 5699–708 (2015).
 110. Grosso AR, Leite AP, Carvalho S, *et al.* Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. *Elife*. 4 (2015).
 111. Simon JM, Hacker KE, Singh D, *et al.* Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. *Genome Res*. 24(2), 241–50 (2014).
 112. Tiedemann RL, Hlady RA, Hanavan PD, *et al.* Dynamic reprogramming of DNA methylation in SETD2-deregulated renal cell carcinoma. *Oncotarget*. (2015).
 113. Abu-Farha M, Lambert J-P, Al-Madhoun AS, Elisma F, Skerjanc IS, Figeys D. The tale of two domains: proteomics and genomics analysis of SMYD2, a new histone methyltransferase. *Mol. Cell. Proteomics*. 7(3), 560–72 (2008).
 114. Xu G, Liu G, Xiong S, Liu H, Chen X, Zheng B. The histone methyltransferase Smyd2 is a negative regulator of macrophage activation by suppressing interleukin 6 (IL-6) and tumor necrosis factor α (TNF- α) production. *J. Biol. Chem*. 290(9), 5414–23 (2015).
 115. Nguyen H, Allali-Hassani A, Antonysamy S, *et al.* LLY-507, a Cell-active, Potent, and Selective Inhibitor of Protein-lysine Methyltransferase SMYD2. *J. Biol. Chem*. 290(22), 13641–53 (2015).
 116. Huang J, Perez-Burgos L, Placek BJ, *et al.* Repression of p53 activity by Smyd2-mediated methylation. *Nature*. 444(7119), 629–32 (2006).
 117. Saddic LA, West LE, Aslanian A, *et al.* Methylation of the retinoblastoma tumor suppressor by SMYD2. *J. Biol. Chem*. 285(48), 37733–40 (2010).
 118. Abu-Farha M, Lanouette S, Elisma F, *et al.* Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2. *J. Mol. Cell Biol*. 3(5), 301–8 (2011).
 119. Zhang X, Tanaka K, Yan J, *et al.* Regulation of estrogen receptor α by histone methyltransferase SMYD2-mediated protein methylation. *Proc. Natl. Acad. Sci. U. S. A*. 110(43), 17284–9 (2013).
 120. Piao L, Kang D, Suzuki T, *et al.* The histone methyltransferase SMYD2 methylates

- PARP1 and promotes poly(ADP-ribosyl)ation activity in cancer cells. *Neoplasia*. 16(3), 257–64, 264.e2 (2014).
121. Cordaux R, Udit S, Batzer MA, Feschotte C. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc. Natl. Acad. Sci. U. S. A.* 103(21), 8101–6 (2006).
 122. Apostolou P, Toloudi M, Kourtidou E, *et al.* Potential role for the Metnase transposase fusion gene in colon cancer through the regulation of key genes. *PLoS One*. 9(10), e109741 (2014).
 123. Williamson EA, Rasila KK, Corwin LK, *et al.* The SET and transposase domain protein Metnase enhances chromosome decatenation: regulation by automethylation. *Nucleic Acids Res.* 36(18), 5822–31 (2008).
 124. Eom GH, Kim K-B, Kim JH, *et al.* Histone methyltransferase SETD3 regulates muscle differentiation. *J. Biol. Chem.* 286(40), 34733–42 (2011).
 125. Pires-Luís AS, Vieira-Coimbra M, Vieira FQ, *et al.* Expression of histone methyltransferases as novel biomarkers for renal cell tumor diagnosis and prognostication. *Epigenetics*. 10(11), 1033–43 (2015).
 126. Cooper SE, Hodimont E, Green CM. A fluorescent bimolecular complementation screen reveals MAF1, RNF7 and SETD3 as PCNA-associated proteins in human cells. *Cell Cycle*. 14(15), 2509–19 (2015).
 127. An S, Yeo KJ, Jeon YH, Song JJ. Crystal structure of the human histone methyltransferase ASH1L catalytic domain and its implications for the regulatory mechanism. *J. Biol. Chem.* 286(10), 8369–8374 (2011).
 128. Qiao Q, Li Y, Chen Z, Wang M, Reinberg D, Xu R-M. The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J. Biol. Chem.* 286(10), 8361–8368 (2011).
 129. Zheng W, Ibáñez G, Wu H, *et al.* Sinefungin derivatives as inhibitors and structure probes of protein lysine methyltransferase SETD2. *J. Am. Chem. Soc.* 134(43), 18004–18014 (2012).
 130. Wang L, Li L, Zhang H, *et al.* Structure of human SMYD2 protein reveals the basis of p53 tumor suppressor methylation. *J. Biol. Chem.* 286(44), 38725–37 (2011).
 131. Ferguson AD, Larsen NA, Howard T, *et al.* Structural basis of substrate methylation and

- inhibition of SMYD2. *Structure*. 19(9), 1262–73 (2011).
132. Xiao B, Jing C, Kelly G, *et al.* Specificity and mechanism of the histone methyltransferase Pr-Set7. *Genes Dev*. 19(12), 1444–1454 (2005).
 133. Horowitz S, Yesselman JD, Al-Hashimi HM, Trievel RC. Direct evidence for methyl group coordination by carbon-oxygen hydrogen bonds in the lysine methyltransferase SET7/9. *J. Biol. Chem*. 286(21), 18658–63 (2011).
 134. Rogawski DS, Ndoj J, Cho HJ, Maillard I, Grembecka J, Cierpicki T. Two Loops Undergoing Concerted Dynamics Regulate the Activity of the ASH1L Histone Methyltransferase. *Biochemistry*. 54(35), 5401–13 (2015).
 135. Graham SE, Tweedy SE, Carlson HA. Dynamic Behavior of the Post-SET Loop Region of NSD1: Implications for Histone Binding and Drug Development. *Protein Sci*. (2016).
 136. Tempel, W., Yu, W., Dong, A., Cerovina, T., Bountra, C., Arrowsmith, C.H., Edwards, A.M., Brown, P.J., Wu, H. SGC. PDB ID: 4YZ8. Methyltransferase domain of human Wolf-Hirschhorn Syndrome Candidate 1-Like protein 1 (WHSC1L1). .
 137. Jiang Y, Sirinupong N, Brunzelle J, Yang Z. Crystal structures of histone and p53 methyltransferase SmyD2 reveal a conformational flexibility of the autoinhibitory C-terminal domain. *PLoS One*. 6(6), e21640 (2011).
 138. Xu S, Zhong C, Zhang T, Ding J. Structure of human lysine methyltransferase Smyd2 reveals insights into the substrate divergence in Smyd proteins. *J. Mol. Cell Biol*. 3(5), 293–300 (2011).
 139. Jiang Y, Trescott L, Holcomb J, *et al.* Structural insights into estrogen receptor α methylation by histone methyltransferase SMYD2, a cellular event implicated in estrogen signaling regulation. *J. Mol. Biol*. 426(20), 3413–25 (2014).
 140. Sweis RF, Wang Z, Algire M, *et al.* Discovery of A-893, A New Cell-Active Benzoxazinone Inhibitor of Lysine Methyltransferase SMYD2. *ACS Med. Chem. Lett*. 6(6), 695–700 (2015).
 141. Greiner D, Bonaldi T, Eskeland R, Roemer E, Imhof A. Identification of a specific inhibitor of the histone methyltransferase SU(VAR)3-9. *Nat. Chem. Biol*. 1(3), 143–5 (2005).
 142. Campagna-Slater V, Mok MW, Nguyen KT, Feher M, Najmanovich R, Schapira M. Structural Chemistry of the Histone Methyltransferases Cofactor Binding Site. *J. Chem*.

- Inf. Model.* 51(3), 612–623 (2011).
143. Devkota K, Lohse B, Liu Q, *et al.* Analogues of the Natural Product Sinefungin as Inhibitors of EHMT1 and EHMT2. *ACS Med. Chem. Lett.* 5(4), 293–7 (2014).
 144. Yao Y, Chen P, Diao J, *et al.* Selective inhibitors of histone methyltransferase DOT1L: design, synthesis, and crystallographic studies. *J. Am. Chem. Soc.* 133(42), 16746–9 (2011).
 145. Schramm VL. Transition States, analogues, and drug development. *ACS Chem. Biol.* 8(1), 71–81 (2013).
 146. Poulin MB, Schneck JL, Matico RE, *et al.* Transition state for the NSD2-catalyzed methylation of histone H3 lysine 36. *Proc. Natl. Acad. Sci. U. S. A.* 113(5), 1197–201 (2016).
 147. Zhang J, Zheng YG. SAM/SAH Analogs as Versatile Tools for SAM-Dependent Methyltransferases. *ACS Chem. Biol.* (2015).
 148. Wang Z, Song J, Milne TA, *et al.* Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to Cyp33 and HDAC-mediated repression. *Cell.* 141(7), 1183–94 (2010).
 149. Wen H, Li Y, Xi Y, *et al.* ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature.* 508(7495), 263–8 (2014).
 150. Sanchez R, Zhou M-M. The PHD finger: a versatile epigenome reader. *Trends Biochem. Sci.* 36(7), 364–372 (2011).
 151. Huang Z, Wu H, Chuai S, *et al.* NSD2 is recruited through its PHD domain to oncogenic gene loci to drive multiple myeloma. *Cancer Res.* 73(20), 6277–88 (2013).
 152. Eidahl JO, Crowe BL, North JA, *et al.* Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Res.* 41(6), 3924–36 (2013).
 153. French CA, Rahman S, Walsh EM, *et al.* NSD3-NUT fusion oncoprotein in NUT midline carcinoma: implications for a novel oncogenic mechanism. *Cancer Discov.* 4(8), 928–41 (2014).
 154. Winter GE, Buckley DL, Paulk J, *et al.* DRUG DEVELOPMENT. Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science.* 348(6241), 1376–81 (2015).
 155. Bondeson DP, Mares A, Smith IED, *et al.* Catalytic in vivo protein knockdown by small-

- molecule PROTACs. *Nat. Chem. Biol.* 11(8), 611–7 (2015).
156. Tripoulas N, LaJeunesse D, Gildea J, Shearn A. The *Drosophila ash1* gene product, which is localized at specific sites on polytene chromosomes, contains a SET domain and a PHD finger. *Genetics*. 143(2), 913–928 (1996).
 157. Gregory GD, Vakoc CR, Rozovskaia T, *et al.* Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes. *Mol. Cell. Biol.* 27(24), 8466–8479 (2007).
 158. Xia M, Liu J, Wu X, *et al.* Histone methyltransferase Ash1l suppresses interleukin-6 production and inflammatory autoimmune diseases by inducing the ubiquitin-editing enzyme A20. *Immunity*. 39(3), 470–481 (2013).
 159. Tanaka Y, Katagiri Z-I, Kawahashi K, Kioussis D, Kitajima S. Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene*. 397(1-2), 161–168 (2007).
 160. Eram MS, Kuznetsova E, Li F, *et al.* Kinetic characterization of human histone H3 lysine 36 methyltransferases, ASH1L and SETD2. *Biochim. Biophys. Acta*. 1850(9), 1842–1848 (2015).
 161. Liu L, Kimball S, Liu H, Holowatyj A, Yang Z-Q. Genetic alterations of histone lysine methyltransferases and their significance in breast cancer. *Oncotarget*. 6(4), 2466–2482 (2014).
 162. Gao J, Aksoy BA, Dogrusoz U, *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal*. 6(269), p11 (2013).
 163. Cerami E, Gao J, Dogrusoz U, *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2(5), 401–404 (2012).
 164. Oh HR, An CH, Yoo NJ, Lee SH. Somatic mutations of amino acid metabolism-related genes in gastric and colorectal cancers and their regional heterogeneity - a short report. *Cell. Oncol. (Dordr)*. 37(6), 455–461 (2014).
 165. Mouradov D, Sloggett C, Jorissen RN, *et al.* Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res*. 74(12), 3238–3247 (2014).
 166. Song Y, Li L, Ou Y, *et al.* Identification of genomic alterations in oesophageal squamous cell cancer. *Nature*. 509(7498), 91–95 (2014).

167. Liu J, Lee W, Jiang Z, *et al.* Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. *Genome Res.* 22(12), 2315–2327 (2012).
168. Zeisig BB, Milne T, García-Cuéllar M-P, *et al.* Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization. *Mol. Cell. Biol.* 24(2), 617–628 (2004).
169. Kleckner IR, Foster MP. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta.* 1814(8), 942–968 (2011).
170. Bain AD. Chemical exchange in NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* 43(3-4), 63–103 (2003).
171. Shi J, Vakoc CR. The Mechanisms behind the Therapeutic Activity of BET Bromodomain Inhibition. *Mol. Cell.* 54(5), 728–736 (2014).
172. Ali M, Hom RA, Blakeslee W, Ikenouye L, Kutateladze TG. Diverse functions of PHD fingers of the MLL/KMT2 subfamily. *Biochim Biophys Acta.* 1843(2), 366–371 (2014).
173. Chambers AL, Pearl LH, Oliver AW, Downs J a. The BAH domain of Rsc2 is a histone H3 binding domain. *Nucleic Acids Res.* 41(19), 9168–9182 (2013).
174. Karplus PA, Diederichs K. Linking crystallographic model and data quality. *Science.* 336(6084), 1030–3 (2012).
175. Heerboth S, Lapinska K, Snyder N, Leary M, Rollinson S, Sarkar S. Use of Epigenetic Drugs in Disease: An Overview. *Genet. Epigenet.* 2014(6), 9–19 (2014).
176. Couture J-F, Collazo E, Brunzelle JS, Trievel RC. Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes Dev.* 19(12), 1455–1465 (2005).
177. Xiao B, Jing C, Wilson JR, *et al.* Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature.* 421(6923), 652–656 (2003).
178. Trievel RC, Flynn EM, Houtz RL, Hurley JH. Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSMT. *Nat. Struct. Biol.* 10(7), 545–552 (2003).
179. Zhang X, Yang Z, Khan SI, *et al.* Structural basis for the product specificity of histone lysine methyltransferases. *Mol. Cell.* 12(1), 177–185 (2003).
180. Xiao B, Wilson JR, Gamblin SJ. SET domains and histone methylation. *Curr. Opin. Struct. Biol.* 13(6), 699–705 (2003).
181. Armache K-J, Garlick JD, Canzio D, Narlikar GJ, Kingston RE. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science.* 334(6058), 977–982 (2011).

182. Zheng Y, Mamdani F, Toptygin D, Brand L, Stivers JT, Cole PA. Fluorescence analysis of a dynamic loop in the PCAF/GCN5 histone acetyltransferase. *Biochemistry*. 44(31), 10501–10509 (2005).
183. Steiner S, Magno A, Huang D, Caflisch A. Does bromodomain flexibility influence histone recognition? *FEBS Lett*. 587(14), 2158–2163 (2013).
184. Hou Z, Danzer JR, Fox CA, Keck JL. Structure of the Sir3 protein bromo adjacent homology (BAH) domain from *S. cerevisiae* at 1.95 Å resolution. *Protein Sci*. 15(5), 1182–1186 (2006).
185. Connelly JJ, Yuan P, Hsu H-C, Li Z, Xu R-M, Sternglanz R. Structure and function of the *Saccharomyces cerevisiae* Sir3 BAH domain. *Mol. Cell. Biol*. 26(8), 3256–3265 (2006).
186. Albert JS, Blomberg N, Breeze AL, *et al*. An integrated approach to fragment-based lead generation: philosophy, strategy and case studies from AstraZeneca’s drug discovery programmes. *Curr. Top. Med. Chem*. 7(16), 1600–29 (2007).
187. Murray CW, Verdonk ML, Rees DC. Experiences in fragment-based drug discovery. *Trends Pharmacol. Sci*. 33(5), 224–32 (2012).
188. Fick RJ, Kroner GM, Nepal B, *et al*. Sulfur-Oxygen Chalcogen Bonding Mediates AdoMet Recognition in the Lysine Methyltransferase SET7/9. *ACS Chem. Biol*. (2016).
189. Chang CA, Chen W, Gilson MK. Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. U. S. A*. 104(5), 1534–9 (2007).
190. Pollock J, Borkin D, Lund G, *et al*. Rational Design of Orthogonal Multipolar Interactions with Fluorine in Protein-Ligand Complexes. *J. Med. Chem*. 58(18), 7465–74 (2015).
191. Fujimoto A, Furuta M, Totoki Y, *et al*. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet*. (2016).
192. Neve RM, Chin K, Fridlyand J, *et al*. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*. 10(6), 515–527 (2006).
193. Novak RL, Harper DP, Caudell D, Slape C, Beachy SH, Aplan PD. Gene expression profiling and candidate gene resequencing identifies pathways and mutations important for malignant transformation caused by leukemogenic fusion genes. *Exp. Hematol*. 40(12), 1016–27 (2012).
194. Caudell D, Aplan PD. The role of CALM-AF10 gene fusion in acute leukemia. *Leukemia*. 22(4), 678–85 (2008).

195. Polakis P. Wnt signaling in cancer. *Cold Spring Harb. Perspect. Biol.* 4(5) (2012).
196. Meyer C, Schneider B, Jakob S, *et al.* The MLL recombinome of acute leukemias. *Leukemia.* 20(5), 777–84 (2006).
197. Yagi H, Deguchi K, Aono A, Tani Y, Kishimoto T, Komori T. Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice. *Blood.* 92(1), 108–17 (1998).
198. Yu BD, Hess JL, Horning SE, Brown GA, Korsmeyer SJ. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature.* 378(6556), 505–8 (1995).
199. Tebar F, Bohlander SK, Sorkin A. Clathrin assembly lymphoid myeloid leukemia (CALM) protein: localization in endocytic-coated pits, interactions with clathrin, and the impact of overexpression on clathrin-mediated traffic. *Mol. Biol. Cell.* 10(8), 2687–702 (1999).
200. Klebig ML, Wall MD, Potter MD, Rowe EL, Carpenter DA, Rinchik EM. Mutations in the clathrin-assembly gene *Picalm* are responsible for the hematopoietic and iron metabolism abnormalities in *fit1* mice. *Proc. Natl. Acad. Sci. U. S. A.* 100(14), 8360–5 (2003).
201. Dik WA, Brahim W, Braun C, *et al.* CALM-AF10+ T-ALL expression profiles are characterized by overexpression of HOXA and BMI1 oncogenes. *Leukemia.* 19(11), 1948–57 (2005).
202. Frye S V. The art of the chemical probe. *Nat. Chem. Biol.* 6(3), 159–161 (2010).
203. Knutson SK, Wigle TJ, Warholc NM, *et al.* A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. *Nat. Chem. Biol.* 8(11), 890–6 (2012).
204. Cao F, Townsend EC, Karatas H, *et al.* Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol. Cell.* 53(2), 247–61 (2014).
205. Liu F, Barsyte-Lovejoy D, Li F, *et al.* Discovery of an in vivo chemical probe of the lysine methyltransferases G9a and GLP. *J. Med. Chem.* 56(21), 8931–42 (2013).
206. Knutson SK, Warholc NM, Wigle TJ, *et al.* Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proc. Natl. Acad. Sci. U. S. A.* 110(19), 7922–7 (2013).
207. Daigle SR, Olhava EJ, Therkelsen CA, *et al.* Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood.* 122(6), 1017–25 (2013).
208. Qiao Q, Li Y, Chen Z, Wang M, Reinberg D, Xu R-M. The structure of NSD1 reveals an

- autoregulatory mechanism underlying histone H3K36 methylation. *J. Biol. Chem.* 286(10), 8361–8 (2011).
209. Makley LN, McMenimen KA, DeVree BT, *et al.* Pharmacological chaperone for α -crystallin partially restores transparency in cataract models. *Science.* 350(6261), 674–7 (2015).
210. Zhao L, Chen X-J, Zhu J, *et al.* Lanosterol reverses protein aggregation in cataracts. *Nature.* 523(7562), 607–11 (2015).
211. Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today. Technol.* 1(4), 337–41 (2004).
212. Shearn A. The ash-1, ash-2 and trithorax genes of *Drosophila melanogaster* are functionally related. *Genetics.* 121(3), 517–25 (1989).
213. Rozovskaia T, Tillib S, Smith S, *et al.* Trithorax and ASH1 interact directly and associate with the trithorax group-responsive bxd region of the Ultrabithorax promoter. *Mol. Cell. Biol.* 19(9), 6441–7 (1999).
214. Jude CD, Climer L, Xu D, Artinger E, Fisher JK, Ernst P. Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. *Cell Stem Cell.* 1(3), 324–37 (2007).
215. Artinger EL, Mishra BP, Zaffuto KM, *et al.* An MLL-dependent network sustains hematopoiesis. *Proc. Natl. Acad. Sci. U. S. A.* 110(29), 12000–5 (2013).
216. Muntean AG, Hess JL. The pathogenesis of mixed-lineage leukemia. *Annu. Rev. Pathol.* 7, 283–301 (2012).
217. Borkin D, He S, Miao H, *et al.* Pharmacologic inhibition of the Menin-MLL interaction blocks progression of MLL leukemia in vivo. *Cancer Cell.* 27(4), 589–602 (2015).
218. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6(3), 277–293 (1995).
219. Goddard TD, Kneller DG. SPARKY 3. Univ. California, San Fr. .
220. Hardman CH, Broadhurst RW, Raine ARC, Grasser KD, Thomas JO, Laue ED. Structure of the A-Domain of HMG1 and Its Interaction with DNA as Studied by Heteronuclear Three- and Four-Dimensional NMR Spectroscopy. *Biochemistry.* 34(51), 16596–16607 (1995).

221. Shim Y, Duan M-R, Chen X, Smerdon MJ, Min J-H. Polycistronic coexpression and nondenaturing purification of histone octamers. *Anal. Biochem.* 427(2), 190–192 (2012).
222. Otwinowski Z, Minor W. Processing of X-ray Diffraction Data Collected in Oscillation Mode [Internet]. In: *Methods in enzymology*. Carter Jr. CW, Sweet RM (Eds.). . Academic Press (New York), 307–326 (1997) [cited 2014 Dec 4].
223. Vagin A, Teplyakov A. MOLREP : an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* 30(6), 1022–1025 (1997).
224. Murshudov GN, Skubák P, Lebedev AA, *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.* 67(Pt 4), 355–367 (2011).
225. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* 66(Pt 4), 486–501 (2010).
226. Winn MD, Ballard CC, Cowtan KD, *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* 67(Pt 4), 235–242 (2011).
227. Chen VB, Arendall WB, Headd JJ, *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* 66(Pt 1), 12–21 (2010).
228. Schneider B, Gelly JC, de Brevern AG, Černý J. Local dynamics of proteins and DNA evaluated from crystallographic B factors. *Acta Crystallogr. D. Biol. Crystallogr.* 70(Pt 9), 2413–2419 (2014).
229. Muntean AG, Tan J, Sitwala K, *et al.* The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. *Cancer Cell.* 17(6), 609–21 (2010).