HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2019 August 23.

Published in final edited form as:

Cell. 2018 August 23; 174(5): 1106–1116.e9. doi:10.1016/j.cell.2018.06.038.

Crystal Structure of the COMPASS H3K4 Methyltransferase Catalytic Module

Peter L. Hsu[1], **Heng Li**[1], **Ho-Tak Lau**[1], **Calvin Leonen**[2], **Abhinav Dhall**[2], **Shao-En Ong**[1], **Champak Chatterjee**[2], and **Ning Zheng**[1,3]

[1]Department of Pharmacology, University of Washington, Seattle, WA 98195

[2]Department of Chemistry, University of Washington, Seattle, WA 98195

[3]Howard Hughes Medical Institute, Box 357280, University of Washington, Seattle, WA 98195

Abstract

The SET1/MLL family of histone methyltransferases are conserved in eukaryotes and regulate transcription by catalyzing histone H3K4 mono-, di-, and tri-methylation. These enzymes form a common five-subunit catalytic core, whose assembly is critical for their basal and regulated enzymatic activities through unknown mechanisms. Here we present the crystal structure of the intact yeast COMPASS histone methyltransferase catalytic module, consisting of Swd1, Swd3, Bre2, Sdc1, and Set1. The complex is organized by Swd1, whose conserved C-terminal tail not only nucleates Swd3 and a Bre2-Sdc1 subcomplex, but also joins Set1 to construct a regulatory pocket next to the catalytic site. This inter-subunit pocket is targeted by a previously unrecognized enzyme-modulating motif in Swd3 and features a doorstop-style mechanism dictating substrate selectivity among SET1/MLL family members. By spatially mapping the functional components of COMPASS, our results provide a structural framework for understanding the multifaceted functions and regulation of the H3K4 methyltransferase family.

Graphical Abstract

Lead contact: Ning Zheng, nzheng@uw.edu.

SUPPLEMENTAL INFORMATION

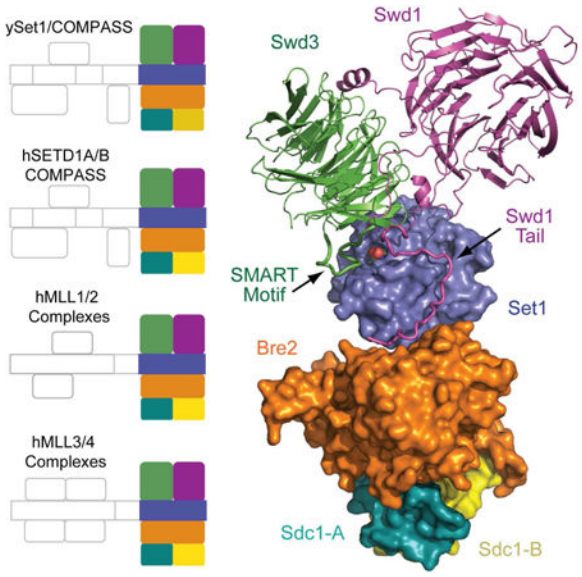
Supplemental information includes six figures and one table and can be found with this article online.

AUTHOR CONTRIBUTIONS

P.L.H. and N.Z. conceived the project. P.L.H. cloned and purified all proteins used in this study. Crystallization, X-ray data collection and processing were carried out by P.L.H. Both P.L.H. and H.L. performed model building and refinement, with input from N.Z. P.L.H. and N.Z. analyzed the structure to identify key structural elements for mutagenesis studies, and P.L.H. carried out all biochemical assays. C.J.L. and A.D. provided nucleosomes for methyltransferase assays. H-T.L. and S-E.O. provided mass spectrometry support. P.L.H. and N.Z. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.



In brief

The crystal structure of the yeast COMPASS histone methyltransferase catalytic module reveals a doorstop-style mechanism dictating substrate selectivity

Introduction

Methylation of histone H3 at K4 (H3K4) is an evolutionarily conserved post-translational modification that marks actively transcribed genes (Allis et al., 2015). The terminal amine of H3K4 can be modified to mono-, di-, and tri-methylated states, each with different locations throughout chromatin. These H3K4 methylation marks are recognized by specific readers, which relay these signals for downstream events, and are erased by specific demethylases to achieve dynamic regulation (Shi et al., 2004; Vermeulen et al., 2010; Yun et al., 2011). H3K4 methylation is catalyzed by the conserved SET1/MLL family of methyltransferases, which contain six functionally non-redundant members in humans, MLL1–4, SETD1A and SETD1B (Shilatifard, 2012). MLL1/2 and MLL3/4 are largely found at gene promoter and enhancer regions, respectively (Denissov et al., 2014; Herz et al., 2012; Hu et al., 2013; Lee et al., 2013; Wang et al., 2009a), while SETD1A and SETD1B are thought to be responsible for maintaining global tri-methylation in the genome (Ardehali et al., 2011; Wu et al., 2008). Due to their important roles in establishing the epigenetic landscape for actively transcribed genes, mutations in each of the SET1/MLL family members have been associated with human diseases, such as mixed lineage leukemia, schizophrenia, autism, and Kabuki syndrome (Hiraide et al., 2018; Muntean and Hess, 2012; Ng et al., 2010; Shinsky et al., 2014; Singh et al., 2016).

Studies of the SET1/MLL enzymes were pioneered by the identification of the *S. cerevisiae* Set1, which represents the sole H3K4 methyltransferase in yeast (Krogan et al., 2002; Miller et al., 2001; Nagy et al., 2002; Roguev et al., 2001). Isolation of the yeast Set1 complex known as COMPASS (COMplex of Proteins ASsociated with Set1) also provided a critical

template for mapping the enzymatic assemblies of the animal SET1/MLL orthologs. Members of the SET1/MLL family all utilize a C-terminal catalytic SET domain to methylate H3K4 in an Sadenosylmethionine (SAM)-dependent manner. In the yeast COMPASS complex, the integrity and enzymatic activity of Set1 is maintained by Swd3, Swd1, Bre2, and Sdc1, which form a stable complex with the C-terminal catalytic domain of the methyltransferase. Remarkably, their animal orthologs, WDR5, RBBP5, ASH2L, DPY-30, collectively known as WRAD, have also been demonstrated to bind and activate the SET domains of all animal SET1/MLL family members (Dehé et al., 2006; Dou et al., 2006; Halbach et al., 2009; Mersman et al., 2012). Interestingly, although these animal SET1/MLL paralogs share a highly similar catalytic domain sequence and a common mode of activation by WRAD, their biochemically isolated catalytic cores can exhibit strikingly different intrinsic substrate specificity (Shinsky et al., 2015).

In addition to the conserved C-terminal catalytic module, yeast Set1 and all six human SET1/MLL enzymes possess large and divergent N-terminal regions, which recruit different binding partners, localize the enzymes to distinct locations in the genome, and modulate the enzymatic activities of the C-terminal catalytic core (Rao and Dou, 2015). The best illustrated example came from genetic and biochemical studies of the yeast Set1/COMPASS complex and its closest mammalian counterparts, the SETD1A and SETD1B complexes. With a similar domain and subunit composition, these evolutionarily conserved SET1/MLL enzymes require histone H2B mono-ubiquitination (uH2B) as a prerequisite for H3K4 methylation (Briggs et al., 2002; Dover et al., 2002; Kim et al., 2009, 2013; Lee et al., 2007; Nakanishi et al., 2009; Racine et al., 2012; Sun and Allis, 2002). While the molecular basis of this histone code crosstalk remains controversial, it is evident that the enzymatic activity of their WRAD-activated catalytic cores are susceptible to further modulation by other internal components of the COMPASS complex as well as an external signal (Kim et al., 2013; Thornton et al., 2014; Vitaliano-Prunier et al., 2008).

Despite many structural studies of the SET1/MLL enzymes, a mechanistic understanding of their functions and regulation is still lacking. Recent structural analyses of a minimized 37 kDa MLL3-ASH2L-RBBP5 peptide complex have suggested the importance of restrained structural flexibility in the SET domain for enzyme activation. However, what properties of the fully-assembled catalytic core might confer differential enzyme regulation and substrate specificity to different SET1/MLL family members remain unclear. Here, we report the crystal structure of an intact yeast COMPASS catalytic core complex at 3.0 Å resolution. Our structure not only delineates the spatial organization and the functional roles of all five subunits, but also unveils a conserved inter-subunit regulatory pocket, which plays a role in enzyme tuning and determining substrate specificity of different SET1/MLL paralogs.

Results and discussion

Overall structure of the yeast COMPASS catalytic module

To reveal the complete architecture of the yeast COMPASS catalytic module, we co-expressed and purified the *K. lactis* Set1 catalytic domain together with the full-length or near full-length Swd3, Swd1, Bre2, and Sdc1 proteins from insect cells (Figure 1A). The yeast COMPASS catalytic module displayed strong distributive H3K4 methyltransferase

activity toward a reconstituted nucleosome substrate (Patel et al., 2009) (Figure 1B). In the presence of SAM and an H3 peptide bearing a K4M mutation, the intact complex was crystallized in the C2 space group with one molecular assembly in the asymmetric unit (Table S1). The catalytic module measures a total of ~145 Å in its longest dimension, and ~70 Å in its shortest. In agreement with previously published low-resolution structural results (Shinsky and Cosgrove, 2015; Takahashi et al., 2011), the catalytic module adopts a Y-shaped overall structure, with the Bre2-Sdc1 sub-complex serving as the "base" and the two β-propeller proteins Swd1 and Swd3 forming the two "arms" (Figure 1C). The SET domain of Set1 is situated in the center of the complex, making direct contacts with all other subunits except Sdc1. With both the SAM cofactor and the H3 mutant peptide bound, the Set1 catalytic domain is trapped in a pre-reaction state with a compact conformation.

The COMPASS catalytic module is organized by the WD40 repeat protein, Swd1, which sports a long and winding C-terminal tail wrapping around three of the four neighboring subunits, Set1, Swd3, and Bre2 (Figure 1C and1D). Swd1 alone buries a total of ~3430 $Å^2$ surface area, consistent with its critical role in nucleating COMPASS *in vivo*. Swd3 is anchored adjacent to Swd1, with the "top" surface of its β-propeller domain facing away from Swd1 (Figure 1C). A loop projecting from the Swd3 β-propeller contacts Set1 and an adjacent Swd1 tail loop through a small, albeit strategic, interface near the active site of the enzyme. Opposite to the Swd1-Swd3 "arms", Bre2 and Sdc1 packs against Set1 as a globular three polypeptide sub-complex with an asymmetric Sdc1 homo-dimer cuddling a long Bre2 C-terminal α-helix. The five subunits, Set1, Swd1, Swd3, Bre1 and Sdc1, therefore, assemble the COMPASS catalytic module subunits with a stoichiometry of 1:1:1:1:2.

Bre2 and ASH2L contain a non-canonical SPRY domain

Previous studies of an aggressively truncated human ASH2L central fragment revealed a canonical "SPRY-only" β-sandwich domain characterized by two tightly packed anti-parallel β-sheets (Chen et al., 2012; Zhang et al., 2015a). An RBBP5 C-terminal sequence-binding site has been subsequently mapped to a shallow surface pocket on one side of the β-sandwich fold (Cao et al., 2010; Zhang et al., 2015a) (Figure 2A). Although these properties of ASH2L are also found in Bre2 (Figure 2B), the structure of the intact yeast ortholog in the context of the entire COMPASS catalytic module unveils a far more structurally and functionally complex architecture that is conserved in the animal orthologs, but missing from previous studies.

The annotated "SPRY-only" domain of Bre2 is flanked by a ~90 amino acid N-terminal "pre-SPRY" region and a 20 amino acid C-terminal tail (Figure 1A). The central SPRY fold also features two long insertions loops, Ins-1 and Ins-2. Together, these sequences adopt a highly intertwined topology, which embraces and expands the SPRY-only β-sandwich into a much larger globular fold (Figure 2B). At the center of this SPRY domain extension is a twisted auxiliary β-sheet, which is constructed by two pairs of anti-parallel β-strands, one from the pre-SPRY region and the other from the Ins-1 loop (Figure 2B). Closer sequence analysis indicates that these key structural elements of the auxiliary β-sheet are conserved in the sequences of all yeast and animal Bre2/ASH2L orthologs (Figure S1). Intriguingly, next

to the SPRY-only β-sandwich, the auxiliary β-sheet provides an extended platform, whose surface is decorated by a cluster of highly, if not strictly, conserved residues, including P46, N48, K49, F52, Y54, D248, I251, Y253, K254, and E260 (Figure 2C and S1). Their mixed hydrophobic and polar side chains and their close proximity to the H3 peptide binding pocket of the Set1 catalytic domain hint at a functional role in either binding the nucleosomal substrate or a COMPASS regulatory factor. The potential functional relevance of the auxiliary β-sheet is further underscored by the α-helix formed by the Bre2 C-terminal tail, which buttresses the extended platform from the bottom (Figure 2D).

Sdcl forms an asymmetric dimer to stabilize Bre2

Sdc1 is the smallest subunit of COMPASS and interacts exclusively with Bre2 at the distal end of the catalytic module. Previous yeast genetics and biochemical analysis have shown that loss of Sdc1 downregulates H3K4 methylation by Set1 (Dehé et al., 2006; Kim et al., 2013; Krogan et al., 2002; South et al., 2010). The complete spatial separation between Sdc1 and Set1 strongly suggests that Sdc1 positively regulates the COMPASS catalytic module by stabilizing the non-canonical Bre2 SPRY domain.

Sdc1 and its animal ortholog, DPY-30, share a highly conserved central sequence (Figure S2). In DPY-30, this region forms three short a-helices and binds to the isolated C-terminal α-helix of ASH2L as a homo-dimer (Tremblay et al., 2014; Wang et al., 2009b). In complex with the full-length Bre2 protein, Sdc1 clasps the Bre2 helical tail in the same binding mode, but also makes additional interfaces with the expanded Bre2 non-canonical SPRY domain (Figure 2D and 3A). Superposition analysis of the two Sdc1 protomers (designated Sdc1-A/B) reveals substantial conformational variation within their N-terminal a-helices (Figure 3B). This structural asymmetry is attributable to the distinct interfaces made by the two copies of Sdc1 with Bre2. For example, in one protomer (Sdc1-A), the tip of its N-terminal α-helix uses the strictly conserved residue Arg88 to lock the C-terminal α-helix of Bre2 onto its non-canonical SPRY domain with a salt bridge and a hydrogen bond (Figure 2D, and 3C). By contrast, the N-terminal region of the other Sdc1 copy curves underneath the Bre2 auxiliary β-sheet with its R88 residue pointing away from Bre2 (Figure 2D, 3C and 3D). By clamping the helical tail of Bre2 onto its non-canonical SPRY domain, the asymmetric Sdc1 dimer appears to reinforce the entire Bre2 protein fold. To validate such a structural role, we show that a 17 amino acid Sdc1 N-terminal region (Asp74 to Arg89), which plays a minor role in tethering the isolated ASH2L C-terminal α-helix to DPY-30 (Tremblay et al., 2014), is critical for the production of soluble Bre2 and the formation of a stable Bre2-Sdc1 complex when co-expressed in insect cells (Figure 3E).

Swd1 organizes the catalytic module and creates an inter-subunit pocket next to active site

As the largest subunit of the COMPASS catalytic module, Swd1 and its orthologs contain a WD40 repeat β-propeller domain, which is preceded by a short N-terminal extension and followed by a long C-terminal tail that varies in length between different species (Figure 1A and S3A). The Swd1 C-terminal tail plays a central role in nucleating the entire complex and can be divided into two segments: a highly conserved WD40 repeat proximal (WDRP) region organizing Set1, Swd3 and Bre2, and a more variable distal portion responsible for

further securing Swd3 binding (Figure 4A). These two regions flank a short disordered loop invisible in the crystal.

The Swd1 WDRP segment emerges from the end of the β-propeller fold and spans roughly 25 amino acids (Figure 4A). This short sequence represents the most conserved region of the entire protein with 21 out of 25 residues either strictly identical or highly similar between yeast and animal orthologs (Figure S3A). Remarkably, the vast majority of the WDRP residues are directly involved in binding Set1, while about one third of these residues also participate in binding either Swd3 or Bre2 at the two ends. The N-terminal half of the WDRP sequence folds into a stable loop, which uses more than five invariant hydrophobic residues, including Trp356 and Phe363, to anchor at a hydrophobic surface covering the junction of the SET-N/C and SET-I subdomains, which sandwich substrate and cofactor in between (Figure 4B and 4C). The tip of this Swd1 WDRP loop joins the two Set1 subdomains and together encloses a well-defined pocket in close vicinity of the SAM-binding site (Figure 4C). Importantly, this Swd1-Set1 intersubunit pocket houses a Swd3 loop protruding from its WD40 repeat domain and is characterized by disease mutations as discussed in the following section. Beyond this Set1-Swd1-Swd3 interface, the Swd1 WDRP segment continues to meander its way around the SET domain and ends at the other tri-subunit interface, which has been previously revealed for the mammalian counterparts of Swd1, Set1, and Bre2 (Li et al., 2016) (Figure S3A and S4A). As a whole, the Swd1 WDRP sequence snakes around one side of the Set1 catalytic domain making multiple inter-subunit interactions. Despite its short length, the condensed interfaces made by the Swd1 WDRP region provides a structural explanation for its critical role in maintaining the integrity of Set1 and the entire COMPASS (Dehé et al., 2006; Halbach et al., 2009; Kim et al., 2013; Mersman et al., 2012).

Besides the WDRP region, Swd1 uses two additional “tentacle”-like structural elements to recruit Swd3 to the COMPASS catalytic module. The N-terminal extension of Swd1 folds into a short helix and latches onto the “bottom” surface of the Swd3 β-propeller domain next to the interface formed between Swd3 and the WDRP sequence (Figure 4A). The C-terminal distal tail of Swd1, meanwhile, utilizes a stretch of residues between Asp395 and Thr403 to dock to a hydrophobic cleft in between two blades of the Swd3 β-propeller, as previously revealed for the mammalian orthologs, WDR5 and RBBP5 (Figure S4B) (Avdic et al., 2011; Odho et al., 2010). These Swd1-Swd3 interfaces are further augmented by the extreme end of the Swd1 C-terminal distal tail, which funnels into a canal sandwiched between the WD40 domains of the two proteins and ends with a short α-helix that nestles in another inter-blade hydrophobic cleft on Swd3 (Figure 4D). In a co-purification experiment, removal of this last α-helix of Swd1, but not the N- terminal extension, resulted in a complete abrogation of Swd3 binding, indicating that the distal interface formed by this last helix is critical for the recruitment of Swd3 to the catalytic module (Figure 4E). Overall, by interacting with Swd3 through a multi-angled interface, Swd1 not only recruits Swd3, but also precisely positions the Swd3 loop protruding out of its β-propeller domain to the Swd1-Set1 inter-subunit pocket adjacent to the active site.

Swd3 regulates Set1 enzymatic activity via a SMART motif targeting the Kabuki pocket

The hallmark interactions of the Set1-centric COMPASS catalytic module lie at the tri-subunit interface, where the prominent Swd3 projecting loop inserts a tryptophan residue, Trp197, into the inter-subunit pocket formed between Set1 and Swd1 (Figure 4C and 5A). This Swd3 loop is found in all yeast orthologs (Figure S5) and is stabilized by a network of hydrogen bonds formed among backbone groups of the loop as well as hydrophobic side chain clustering (Figure 5A). Next to the Swd3 loop, the Swd1 WDRP loop provides continuous support by complementing the edge of the Swd3 propeller domain. Sitting snugly at the center of the Swd1-Set1 pocket, the indole ring of Swd3-Trp197 is sandwiched between Phe363 of Swd1 and Ile909 of Set1 (Figure 5B). It is further surrounded by a mixture of hydrophobic and polar residues highly conserved among all Swd1 and Set1 proteins from yeast to humans. Remarkably, the side chain of this Swd3 tryptophan residue is only about 5 Å away from the carboxyl end of the SAM cofactor. It also buries two Set1 residues, Tyr913 and Arg933, which are found in the majority of SET domain-containing enzymes with essential roles in catalysis (Zhang et al., 2002) (Figure 5B).

The structural configuration of this prominent tri-subunit interface strongly suggests that Swd3-Trp197 might play an important role in modulating the enzymatic activity of Set1 through the Set1-Swd1 pocket. In support of this notion, alanine mutation of Trp197 drastically compromised the catalytic activity of Set1 by slowing down the production of all methylation marks on nucleosomal substrates (Figure 5C). The functional importance of its targeting pocket is also manifested by mutations of two MLL4 /SETD1B residues found in both Kabuki syndrome and autism patients, which correspond to two yeast Set1 residues in direct contact with Swd3- Trp197: Arg933 and its adjacent Thr926 (Hiraide et al., 2018; Shinsky et al., 2014). To highlight their functional roles, we hereafter name the Swd3 loop sequence, DWIAE, as the Set1 Methyltransferase Activity RegulaTor (SMART) motif, and its targeting site as the “Kabuki” pocket.

The WD40 repeat domain of WDR5, the animal ortholog of Swd3, has been previously shown to possess a central H3 peptide-binding site, which is also utilized for engaging the WIN motif preceding the SET domains of all human SET1/MLL family members (Couture et al., 2006; Dharmarajan et al., 2012; Patel et al., 2008; Ruthenburg et al., 2006; Schuetz et al., 2006). In our structure, the equivalent site on Swd3 is more than 35 Å away from the active site of Set1 (Figure S4A), arguing against a cis-acting substrate adapter function of WDR5 if it adopts the same binding mode in the catalytic module. Interestingly, yeast Swd3 lacks a critical H3 peptide- binding phenylalanine residue (Ruthenburg et al., 2006) and does not display any detectable affinity toward the H3 peptide (Figure S5, data not shown), nor does yeast Set1 contains a WIN motif in front of its catalytic domain. Thus, the central peptide-binding site of the WDR5 β- propeller domain most likely has been co-evolved with the WIN motif in the expanded SET1/MLL family members in animals. Curiously, concurrent with the evolutionary gain of these functional elements, the animal WDR5 orthologs have lost the SMART motif found in Swd3 (Figure S5), which targets the Kabuki pocket and regulates the enzymatic activity of yeast Set1. Because the Swd1 and Set1 residues constructing the Kabuki pocket are highly conserved from yeast to humans, the inter-subunit pocket is expected to exist in all yeast and animal COMPASS and COMPASS-

like complexes and potentially serves as a common regulatory site. Minimally, the catalytic module of the animal SETD1A and SETD1B COMPASS complexes is anticipated to contain this regulatory pocket, which could be targeted by a structural element similar or analogous to the yeast Swd3 SMART motif. In fact, we were able to identify a putative SMART motif, DWLND, immediately upstream of the WIN motif in the vertebrate SETD1A and SETD1B protein themselves (Figure 5D). To our surprise, when we compared the recombinant WIN-SET domain of SETD1B, which was used for reconstituting the enzymatic activity of its catalytic core, with a longer construct that includes this upstream putative SMART motif, the activity of the WRAD-bound human enzyme was substantially inhibited, suggesting an auto- inhibitory mechanism in contrast to the stimulating effect of the Swd3 SMART loop (Figure 5C). Although the exact mode of action of this putative SETD1B SMART motif awaits future investigation, this result highlights the susceptibility of the mammalian COMPASS catalytic module to activity modulation.

Structural determinant of substrate specificity among SET1/MLL family members

The substrate specificity of a SET domain-containing enzyme is defined by its intrinsic ability to mono- or multi-methylate the lysine substrate. Although this property is usually dictated by a “Phe/Tyr” switch (Couture et al., 2008), the SET1/MLL enzymes share an identical Tyr residue at the relevant position, which fails to explain the differences in their enzymatic product profiles (Figure S3B). With the first structure of a robust prototypical tri-methyltransferase in the SET1/MLL family, we set to identify the structural determinant of substrate specificity among different family members (Li et al., 2016; Southall et al., 2009; Zhang et al., 2015b).

Previous studies have suggested that the catalytic domain of SET1/MLL enzymes can adopt open and closed conformations, which involve a hinge motion between the SET-N/C and SET-I subdomains (Li et al., 2016; Southall et al., 2009) (Figure 6A). Superposition analysis indicates that the Set1 catalytic domain is trapped in the closed conformation, which has also been reported for the mono-methyltransferase, MLL3 (Figure 6A). Structural comparison of the two enzymes revealed a strictly conserved substrate-binding pocket with few significant conformational variations. A closer examination with sequence alignment analysis, nevertheless, unveiled a variable “door” loop, which demarcates the Kabuki pocket and the cofactor binding site right above the substrate peptide (Figure 6B). Among yeast Set1 and the six human SET1/MLL family members, this 4–5 amino acids loop features a common central glycine residue flanked by variable amino acids on each side, [GI/NR]G[V/I/C/SS] (Figure 6C). Strikingly, sequence variations in this “door” loop seem to correlate with the substrate specificity of the enzymes (Shinsky et al., 2015) (Figure S6).

To investigate the possible role of this structural element in substrate specificity, we first probed the importance of the second half of the “door” loop, which directly contacts the SAM cofactor in yeast Set1 and forms a channel for the target lysine to tunnel through. Substitution of the tandem serine residues with single residues found in MLL1–4 gradually reduced the tri- methyltransferase activity with bulkier hydrophobic side chains (Figure 6D, lane 1–4). Similarly, replacing the single valine residue in MLL3 with either cysteine or tandem serine had discernable effects on the product pattern, either slightly enhancing multi-

methylation, or compromising mono-methylation (Figure 6D, lane 5–7). This portion of the loop, therefore, can affect, but not dramatically alter, substrate specificity of the enzymes. We next examined the role of the two N-terminal amino acids of the door loop, which belong to the Kabuki pocket. Mutation of Gly-Ile in Set1 to Asn-Arg rendered the yeast catalytic module completely insoluble. This result was not unexpected, because the hydrophobic isoleucine residue, Ile909, is a key component of the Set1 Kabuki pocket and directly interacts with Swd3-Trp197 of the SMART loop (Figure 5B). Strikingly, in our more challenging "up-engineering" attempt, the mono-methyltransferase MLL3 can be effectively converted into a robust tri-methyltransferase by simply replacing its Asn-Arg residues in the door loop with Gly-Ile found in MLL1/2 and SETD1A/B (Figure 6C and 6D, lane 5 vs. lane 8 & 9).

Based on our superposition analysis, we propose that these changes most likely eliminated a "doorstop" mechanism that keeps MLL3 from multi-methylating its substrate. When the complete RBBP5/Swd1 WDRP loop is modeled on MLL3 SET domain in its closed conformation, the arginine residue in the MLL3 door loop, Arg4822, is predicted to form a perfect salt bridge with the strictly conserved RBBP5/Swd1 aspartate residue, RBBP5-Asp335/Swd1-Asp362 (Figure 6E). Although an open conformation of MLL3 has recently been determined in complex with a partial RBBP5/Swd1 WDRP loop (Li et al., 2016), the two doorstop residues would sterically collide when the aspartate residue is modeled in with the rest of the WDRP loop (Figure 6F). The opening of the MLL3 SET-I sub-domain, therefore, is likely to be more limited in the presence of the complete RBBP5/Swd1 WDRP loop. In contrast, the isoleucine residue (Ile3880) in the equivalent position would not induce clashing in the widely open form of MLL1 (Figure 6F). Together, these analyses suggest that the signature Asn-Arg sequence most likely restricts the activity of MLL3/4 to mono-methylation by acting as a doorstop to effectively limit the large breathing motion of the SET-I sub-domain, which is necessary for up-taking and accommodating mono- and particularly di-methylated substrate lysine (Figure S6D). In congruence with this model, the RBBP5 Asp335Ala mutant, which is expected to partially release the doorstop mechanism, allowed the wild type MLL3 to catalyze higher order methylation (Figure 6G). While a minimized ASH2L-RBBP5 WDRP fragment has been previously shown to confer catalytic activity to the SET1/MLL catalytic domains by reducing their inherent flexibility, our results suggest that further restricted subdomain movement is involved in controlling substrate specificity.

Conclusion

By revealing the structural relationships of all subunits, the crystal structure of the intact yeast Set1-Swd1-Swd3-Bre2-Sdc1 complex helps establish the hetero-pentameric enzymatic assembly as a common catalytic module shared by all SET1/MLL family members. The inner workings of the catalytic module are highlighted by its complex and intricate architecture and a disease-mutated active site-neighboring regulatory pocket (Hiraide et al., 2018; Ng et al., 2010; Shinsky et al., 2014), which sensitizes different family members to activity modulation and confers them differential substrate specificity. The WRAD "activation" complex, therefore, not only stimulates the enzymatic activity of all SET1/MLL family members, but also creates a separate structural dial that enables paralog-specific

activity tuning. Future studies of the catalytic modules in context of the entire COMPASS and COMPASS-like super-complexes, and with additional binding partners, nucleosomal substrate, and regulatory signals, will shed light on the diverse mechanisms controlling the activity of the H3K4 methyltransferases in remodeling the chromatin landscape and regulating different cellular programs. Together, these studies hold the promise for identifying possible strategies for developing therapeutic agents targeting specific SET1/MLL family members.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Molecular biology and protein purification
  - Nucleosome reconstitution
  - COMPASS catalytic core crystallization
  - X-ray data collection and structure determination
  - Protein co-purification pulldown
  - Methyltransferase assays
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ning Zheng (nzheng@uw.edu.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

For DNA extraction, *E.coli* DH5α was used. For bacmid production, *E.coli* DH10Bac was used. For baculovirus production and amplification, Sf9 insect cells were used. For protein expression, both *E.coli* BL21(DE3) and HighFive insect cells were used.

METHOD DETAILS

Molecular biology and protein purification

All COMPASS subunits were cloned from *K. lactis* genomic DNA and subcloned into pFastBac vectors for protein expression in insect cells. Recombinant viruses were produced per manufacturer instructions (Life Technologies) and were amplified three times in Sf9

monolayer cells to produce high titer P4 viruses for use in infection of HighFive monolayer cells to produce protein.

Cells were harvested after 2–3 days post-infection and lysed by sonication in a buffer containing 40 mM HEPES pH 7.5, 350 mM NaCl, 5 mM βME, 10% glycerol. Lysates were clarified by ultracentrifugation. The cleared lysate was purified using a glutathione affinity column (GE Healthcare) pre-equilibrated in the lysis buffer. Protein was released from the column by overnight on-column cleavage of the GST-tag by TEV protease at 4ºC. Eluted material was further purified on a 5 mL HiTrap Q-HP column (GE Healthcare), followed by a Superdex 200 Increase gel filtration column (GE Healthcare) in a final buffer containing 10 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT. Proteins were concentrated to 10–20 mg/mL and flash frozen in liquid nitrogen for future use.

Mutants were generated using overlap extension PCR, and purified similarly as wild type Set1 complexes. SET1/MLL family members and WRAD subunits were all PCR amplified from a human cDNA library, and subcloned into pFastBac vectors for co-expression in insect cells. Wild-type and mutant complexes of SET1/MLL were purified similarly to the yeast Set1 complex, concentrated to 2–5 mg/mL and flash frozen in liquid nitrogen.

Nucleosome reconstitution

Histones were purified according to the one-pot refolding method described previously (Lee et al., 2015). The 147-bp 601 DNA was purified from a plasmid encoding 20 repeats of the sequence, following a previously established protocol (Dyer et al., 2004). The plasmid was a kind gift from R. McGinty (McGinty et al., 2016). Small scale reconstitutions were performed to find a 1: 1 match of histones:DNA. Large scale reconstitutions were subsequently performed by dialyzing the sample with a buffer containing high salt (2 M NaCl) to low salt (0 M NaCl) over a 36 hour gradient. Nucleosomes were concentrated to ~1 mg/mL and stored at 4ºC.

COMPASS catalytic core crystallization

Set1 complex samples were mixed with a synthetic H3K4M peptide (Genscript) and SAM (NEB) to final concentrations of 0.4 and 0.5 mM respectively, and incubated on ice for one hour. Trypsin (Sigma-Aldrich) was subsequently added to the sample at a w/w trypsin:protein ratio of 1:300. The samples were further incubated on ice for another 30 minutes prior to crystallization. Protein was mixed with precipitant (0.1 M Bicine pH 9.0, 7–10% PEG20000, 100 mM Potassium Sodium Tartrate) at a ratio of 2:1 in the hanging drop format at 4ºC. Crystals appeared in 2–3 days, and would complete growth in approximately a week. Crystals were then slowly transferred into mother liquor supplemented by 30% ethylene glycol by increasing the ethylene glycol concentrations in 10% intervals, with a minimum of 30 minutes per transfer step. Crystals were then frozen in liquid nitrogen for synchrotron data collection.

X-ray data collection and structure determination

X-ray diffraction data was collected at the Advanced Light Source at Berkeley on beam lines 8.2.1 and 8.2.2. Due to radiation damage and the relatively low symmetry of the crystal

(C2), several datasets with isomorphous crystals were collected and merged. Diffraction data was indexed, integrated, and scaled using the HKL2000 package (Otwinowski and Minor, 1997). Resolution cutoffs were determined using completeness (>80%) and I/σ>1 as primary criteria.

Initial phases were determined using a sequential molecular replacement search, first with the structure of WDR5 (PDB: 2H14) as a polyAla model, followed by a search using the structure of MLL3-ASH2LSPRY (PDB: 5F6K) also in polyAla chains. The RBBP5 peptide in this structure was removed for this search. With the successful identification of these three subunits, another search was performed using WDR5 to locate the WD40 repeat domain of Swd1 in the complex. Density modification was subsequently performed with RESOLVE (Terwilliger, 2000) to obtain an interpretable electron density map for model building using Coot (Emsley and Cowtan, 2004). After cycles of model building and refinement using Phenix (Adams et al., 2010), a final model was obtained with Rwork/Rfree of 22.8/27.1 (see Figure S4C for representative densities). All structure figures were drawn using PyMOL (Delano) and all structural modeling were based on structural superposition.

Protein co-purification pulldown

Three 150 mm plates of HighFive monolayer cells were co-infected with viruses encoding the proteins of interest. Cells were harvested after 2–3 days of infection and lysed in lysis buffer. Clarified lysates were loaded on to a 50 μL glutathione affinity column equilibrated in lysis buffer, and then washed with 20 column volumes of buffer. Material was eluted from the column by boiling the beads in SDS-PAGE loading buffer. After electrophoresis, the gel was stained by Coomassie for analysis. For inputs, lysates were loaded on a gel, and then transferred to PVDF membranes. The membranes were stained with Ponceau to assess even loading of lysates on the membranes, and then blocked by 5% milk in TBS-T for 30 minutes at room temperature with gentle rocking. The blots were probed by anti-His and anti-GST antibodies in 5% milk in TBS-T overnight at 4°C. In the following morning, blots were washed with TBS-T, and appropriate secondary antibody coupled with HRP was added to the blots and incubated for one hour at room temperature. Blots were washed with TBS-T to remove excess secondary. Finally, ECL reagent was added, and blots were exposed on an imager for analysis.

Methyltransferase assays

1 μM of enzyme and 0.5 μM nucleosomes were incubated in a buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT, 0.2 mM SAM for 30 minutes at 30°C. Reactions were quenched with SDS-PAGE loading buffer, resolved on a 15% gel, and transferred to PVDF membranes. Membranes were blocked in 5% milk for 30 minutes at room temperature, and then probed with antibodies against H3K4me1 (CST D1A9), H3K4me2 (Abcam ab7766), H3K4me3 (Abcam ab8580) and H3 (Abcam ab1791) overnight. Blots were washed with TBS-T and then incubated with secondary antibody for 1 hour at room temperature. Blots were subsequently washed with TBS-T to remove excess secondary. Finally, ECL reagent was then added and blots were exposed on an imager for analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Protein quantification was done using Bio-rad Protein Assay Dye, and comparing readings against a standard curve constructed using BSA. Nucleosome concentrations were determined using an A260 extinction coefficient of 2,784,500 $M^{-1}cm^{-1}$ on a Nanodrop spectrophotometer (Thermo-Fisher).

DATA AND SOFTWARE AVAILABILITY

The accession numbers for the data reported in this paper are PDB 6CHG.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOLEDGEMENTS

We thank the staff at the Advanced Light Source for support during crystallographic data collection. We thank Dr. Robert McGinty for the kind gift of the 20 repeats 147–601 DNA plasmid for nucleosome reconstitutions. N.Z. is an investigator of Howard Hughes Medical Institute. This work was supported by Howard Hughes Medical Institute (N.Z.), and the National Institutes of Health (1R01GM110430 to C.C. and R01AR065459 (S-E.O.)).

References

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr 66, 213–221. [PubMed: 20124702]

Allis CD, Caparros M-L, Jenuwein T, and Reinberg D (2015). Epigenetics (Cold Spring Harbor Laboratory Press).

Ardehali MB, Mei A, Zobeck KL, Caron M, Lis JT, and Kusch T (2011). Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. EMBO J 30, 2817–2828. [PubMed: 21694722]

Avdic V, Zhang P, Lanouette S, Groulx A, Tremblay V, Brunzelle J, and Couture J-F (2011). Structural and Biochemical Insights into MLL1 Core Complex Assembly. Structure 19, 101–108. [PubMed: 21220120]

Briggs SD, Xiao T, Sun Z-W, Caldwell JA, Shabanowitz J, Hunt DF, Allis CD, and Strahl BD (2002). Gene silencing: Trans-histone regulatory pathway in chromatin. Nature 418, 498. [PubMed: 12152067]

Cao F, Chen Y, Cierpicki T, Liu Y, Basrur V, Lei M, and Dou Y (2010). An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain. PloS One 5, e14102. [PubMed: 21124902]

Chen Y, Cao F, Wan B, Dou Y, and Lei M (2012). Structure of the SPRY domain of human Ash2L and its interactions with RbBP5 and DPY30. Cell Res 22, 598–602. [PubMed: 22231628]

Couture J-F, Collazo E, and Trievel RC (2006). Molecular recognition of histone H3 by the WD40 protein WDR5. Nat. Struct. Mol. Biol 13, 698–703. [PubMed: 16829960]

Couture J-F, Dirk LMA, Brunzelle JS, Houtz RL, and Trievel RC (2008). Structural origins for the product specificity of SET domain protein methyltransferases. Proc. Natl. Acad. Sci 105, 20659–20664. [PubMed: 19088188]

Dehé P-M, Dichtl B, Schaft D, Roguev A, Pamblanco M, Lebrun R, Rodríguez-Gil A, Mkandawire M, Landsberg K, Shevchenko A, et al. (2006). Protein Interactions within the Set1 Complex and Their Roles in the Regulation of Histone 3 Lysine 4 Methylation. J. Biol. Chem 281, 35404–35412. [PubMed: 16921172]

Denissov S, Hofemeister H, Marks H, Kranz A, Ciotta G, Singh S, Anastassiadis K, Stunnenberg HG, and Stewart AF (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. Dev. Camb. Engl 141, 526–537.

Dharmarajan V, Lee J-H, Patel A, Skalnik DG, and Cosgrove MS (2012). Structural basis for WDR5 interaction (Win) motif recognition in human SET1 family histone methyltransferases. J. Biol. Chem 287, 27275–27289. [PubMed: 22665483]

Dou Y, Milne TA, Ruthenburg AJ, Lee S, Lee JW, Verdine GL, Allis CD, and Roeder RG (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat. Struct. Mol. Biol 13, 713–719. [PubMed: 16878130]

Dover J, Schneider J, Tawiah-Boateng MA, Wood A, Dean K, Johnston M, and Shilatifard A (2002). Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. J. Biol. Chem 277, 28368–28371. [PubMed: 12070136]

Dyer PN, Edayathumangalam RS, White CL, Bao Y, Chakravarthy S, Muthurajan UM, and Luger K (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. Methods Enzymol 375, 23–44. [PubMed: 14870657]

Emsley P, and Cowtan K (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr 60, 2126–2132. [PubMed: 15572765]

Halbach A, Zhang H, Wengi A, Jablonska Z, Gruber IML, Halbeisen RE, Dehé P-M, Kemmeren P, Holstege F, Géli V, et al. (2009). Cotranslational assembly of the yeast SET1C histone methyltransferase complex. EMBO J 28, 2959–2970. [PubMed: 19713935]

Herz H-M, Mohan M, Garruss AS, Liang K, Takahashi Y-H, Mickey K, Voets O, Verrijzer CP, and Shilatifard A (2012). Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev 26, 2604–2620. [PubMed: 23166019]

Hiraide T, Nakashima M, Yamoto K, Fukuda T, Kato M, Ikeda H, Sugie Y, Aoto K, Kaname T, Nakabayashi K, et al. (2018). De novo variants in SETD1B are associated with intellectual disability, epilepsy and autism. Hum. Genet

Hu D, Gao X, Morgan MA, Herz H-M, Smith ER, and Shilatifard A (2013). The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol. Cell. Biol 33, 4745–4754. [PubMed: 24081332]

Kim J, Guermah M, McGinty RK, Lee J-S, Tang Z, Milne TA, Shilatifard A, Muir TW, and Roeder RG (2009). RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. Cell 137, 459–471. [PubMed: 19410543]

Kim J, Kim J-A, McGinty RK, Nguyen UTT, Muir TW, Allis CD, and Roeder RG (2013). The n-SET Domain of Set1 Regulates H2B Ubiquitylation-Dependent H3K4 Methylation. Mol. Cell 49, 1121–1133. [PubMed: 23453808]

Krogan NJ, Dover J, Khorrami S, Greenblatt JF, Schneider J, Johnston M, and Shilatifard A (2002). COMPASS, a Histone H3 (Lysine 4) Methyltransferase Required for Telomeric Silencing of Gene Expression. J. Biol. Chem 277, 10753–10755. [PubMed: 11805083]

Lee J-E, Wang C, Xu S, Cho Y-W, Wang L, Feng X, Baldridge A, Sartorelli V, Zhuang L, Peng W, et al. (2013). H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. ELife 2, e01503. [PubMed: 24368734]

Lee J-S, Shukla A, Schneider J, Swanson SK, Washburn MP, Florens L, Bhaumik SR, and Shilatifard A (2007). Histone Crosstalk between H2B Monoubiquitination and H3 Methylation Mediated by COMPASS. Cell 131, 1084–1096. [PubMed: 18083099]

Lee Y-T, Gibbons G, Lee SY, Nikolovska-Coleska Z, and Dou Y (2015). One-pot refolding of core histones from bacterial inclusion bodies allows rapid reconstitution of histone octamer. Protein Expr. Purif 110, 89–94. [PubMed: 25687285]

Li Y, Han J, Zhang Y, Cao F, Liu Z, Li S, Wu J, Hu C, Wang Y, Shuai J, et al.(2016). Structural basis for activity regulation of MLL family methyltransferases. Nature 530, 447–452. [PubMed: 26886794]

McGinty RK, Makde RD, and Tan S (2016). Preparation, crystallization and structure determination of chromatin enzyme/nucleosome complexes. Methods Enzymol 573, 43–65. [PubMed: 27372748]

Mersman DP, Du H-N, Fingerman IM, South PF, and Briggs SD (2012). Charge-based interaction conserved within histone H3 lysine 4 (H3K4) methyltransferase complexes is needed for protein

stability, histone methylation, and gene expression. J. Biol. Chem 287, 2652–2665. [PubMed: 22147691]

Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF, and Shilatifard A (2001). COMPASS: a complex of proteins associated with a trithorax- related SET domain protein. Proc. Natl. Acad. Sci. U. S. A 98, 12902–12907. [PubMed: 11687631]

Muntean AG, and Hess JL (2012). The pathogenesis of mixed-lineage leukemia. Annu. Rev. Pathol 7, 283–301. [PubMed: 22017583]

Nagy PL, Griesenbeck J, Kornberg RD, and Cleary ML (2002). A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. Proc. Natl. Acad. Sci 99, 90–94. [PubMed: 11752412]

Nakanishi S, Lee JS, Gardner KE, Gardner JM, Takahashi Y, Chandrasekharan MB, Sun Z-W, Osley MA, Strahl BD, Jaspersen SL, et al. (2009). Histone H2BK123 monoubiquitination is the critical determinant for H3K4 and H3K79 trimethylation by COMPASS and Dot1. J. Cell Biol 186, 371–377. [PubMed: 19667127]

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, et al. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat. Genet 42, 790–793. [PubMed: 20711175]

Odho Z, Southall SM, and Wilson JR (2010). Characterization of a Novel WDR5-binding Site That Recruits RbBP5 through a Conserved Motif to Enhance Methylation of Histone H3 Lysine 4 by Mixed Lineage Leukemia Protein-1. J. Biol. Chem 285, 32967–32976. [PubMed: 20716525]

Otwinowski Z, and Minor W (1997). [20] Processing of X-ray diffraction data collected in oscillation mode In Methods in Enzymology, (Elsevier), pp. 307–326.

Patel A, Vought VE, Dharmarajan V, and Cosgrove MS (2008). A Conserved Arginine- containing Motif Crucial for the Assembly and Enzymatic Activity of the Mixed Lineage Leukemia Protein-1 Core Complex. J. Biol. Chem 283, 32162–32175. [PubMed: 18829457]

Patel A, Dharmarajan V, Vought VE, and Cosgrove MS (2009). On the Mechanism of Multiple Lysine Methylation by the Human Mixed Lineage Leukemia Protein-1 (MLL1) Core Complex. J. Biol. Chem 284, 24242–24256. [PubMed: 19556245]

Racine A, Pagé V, Nagy S, Grabowski D, and Tanny JC (2012). Histone H2B Ubiquitylation Promotes Activity of the Intact Set1 Histone Methyltransferase Complex in Fission Yeast. J. Biol. Chem 287, 19040–19047. [PubMed: 22505722]

Rao RC, and Dou Y (2015). Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. Nat. Rev. Cancer 15, 334–346. [PubMed: 25998713]

Roguev A, Schaft D, Shevchenko A, Pijnappel WWMP, Wilm M, Aasland R, and Stewart AF (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. EMBO J 20, 7137–7148. [PubMed: 11742990]

Ruthenburg AJ, Wang W, Graybosch DM, Li H, Allis CD, Patel DJ, and Verdine GL (2006). Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. Nat. Struct. Mol. Biol 13, 704–712. [PubMed: 16829959]

Schuetz A, Allali-Hassani A, Martín F, Loppnau P, Vedadi M, Bochkarev A, Plotnikov AN, Arrowsmith CH, and Min J (2006). Structural basis for molecular recognition and presentation of histone H3 by WDR5. EMBO J 25, 4245–4252. [PubMed: 16946699]

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, and Shi Y (2004). Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. Cell 119, 941–953. [PubMed: 15620353]

Shilatifard A (2012). The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. Annu. Rev. Biochem 81, 65–95. [PubMed: 22663077]

Shinsky SA, and Cosgrove MS (2015). Unique Role of the WD-40 Repeat Protein 5 (WDR5) Subunit within the Mixed Lineage Leukemia 3 (MLL3) Histone Methyltransferase Complex. J. Biol. Chem 290, 25819–25833. [PubMed: 26324722]

Shinsky SA, Hu M, Vought VE, Ng SB, Bamshad MJ, Shendure J, and Cosgrove MS(2014). A Non-Active-Site SET Domain Surface Crucial for the Interaction of MLL1 and the RbBP5/Ash2L Heterodimer within MLL Family Core Complexes. J. Mol. Biol 426, 2283–2299. [PubMed: 24680668]

Shinsky SA, Monteith KE, Viggiano S, and Cosgrove MS (2015). Biochemical reconstitution and phylogenetic comparison of human SET1 family core complexes involved in histone methylation. J. Biol. Chem 290, 6361–6375. [PubMed: 25561738]

Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, Suvisaari J, Chheda H, Blackwood D, Breen G, et al. (2016). Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat. Neurosci 19, 571–577. [PubMed: 26974950]

South PF, Fingerman IM, Mersman DP, Du H-N, and Briggs SD (2010). A Conserved Interaction between the SDI Domain of Bre2 and the Dpy-30 Domain of Sdc1 Is Required for Histone Methylation and Gene Expression. J. Biol. Chem 285, 595–607. [PubMed: 19897479]

Southall SM, Wong P-S, Odho Z, Roe SM, and Wilson JR (2009). Structural Basis for the Requirement of Additional Factors for MLL1 SET Domain Activity and Recognition of Epigenetic Marks. Mol. Cell 33, 181–191. [PubMed: 19187761]

Sun Z-W, and Allis CD (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature 418, 104–108. [PubMed: 12077605]

Takahashi Y, Westfield GH, Oleskie AN, Trievel RC, Shilatifard A, and Skiniotis G (2011). Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. Proc. Natl. Acad. Sci 108, 20526–20531. [PubMed: 22158900]

Terwilliger TC (2000). Maximum-likelihood density modification. Acta Crystallogr. D Biol. Crystallogr 56, 965–972. [PubMed: 10944333]

Thornton JL, Westfield GH, Takahashi Y-H, Cook M, Gao X, Woodfin AR, Lee J-S, Morgan MA, Jackson J, Smith ER, et al. (2014). Context dependency of Set1/COMPASS- mediated histone H3 Lys4 trimethylation. Genes Dev 28, 115–120. [PubMed: 24402317]

Tremblay V, Zhang P, Chaturvedi C-P, Thornton J, Brunzelle JS, Skiniotis G, Shilatifard A, Brand M, and Couture J-F (2014). Molecular basis for DPY-30 association to COMPASS-like and NURF complexes. Struct. Lond. Engl 1993 22, 1821–1830.

Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, Lee KK, Olsen JV, Hyman AA, Stunnenberg HG, et al. (2010). Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. Cell 142, 967–980. [PubMed: 20850016]

Vitaliano-Prunier A, Menant A, Hobeika M, Géli V, Gwizdek C, and Dargemont C (2008). Ubiquitylation of the COMPASS component Swd2 links H2B ubiquitylation to H3K4 trimethylation. Nat. Cell Biol 10, 1365–1371. [PubMed: 18849979]

Wang P, Lin C, Smith ER, Guo H, Sanderson BW, Wu M, Gogol M, Alexander T, Seidel C, Wiedemann LM, et al. (2009a). Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. Mol. Cell. Biol 29, 6074–6085. [PubMed: 19703992]

Wang X, Lou Z, Dong X, Yang W, Peng Y, Yin B, Gong Y, Yuan J, Zhou W, Bartlam M, et al. (2009b). Crystal Structure of the C-Terminal Domain of Human DPY-30-Like Protein: A Component of the Histone Methyltransferase Complex. J. Mol. Biol 390, 530–537. [PubMed: 19481096]

Wu M, Wang PF, Lee JS, Martin-Brown S, Florens L, Washburn M, and Shilatifard A (2008). Molecular Regulation of H3K4 Trimethylation by Wdr82, a Component of Human Set1/COMPASS. Mol. Cell. Biol 28, 7337–7344. [PubMed: 18838538]

Yun M, Wu J, Workman JL, and Li B (2011). Readers of histone modifications. Cell Res 21, 564–578. [PubMed: 21423274]

Zhang P, Chaturvedi C-P, Tremblay V, Cramet M, Brunzelle JS, Skiniotis G, Brand M, Shilatifard A, and Couture J-F (2015a). A phosphorylation switch on RbBP5 regulates histone H3 Lys4 methylation. Genes Dev 29, 123–128. [PubMed: 25593305]

Zhang X, Tamaru H, Khan SI, Horton JR, Keefe LJ, Selker EU, and Xiaodong C (2002). Structure of the Neurospora SET Domain Protein DIM-5, a Histone H3 Lysine Methyltransferase. Cell 111, 117–127. [PubMed: 12372305]

Zhang Y, Mittal A, Reid J, Reich S, Gamblin SJ, and Wilson JR (2015b). Evolving Catalytic Properties of the MLL Family SET Domain. Structure 23, 1921–1933. [PubMed: 26320581]

Highlights

- Crystal structure of the conserved heteropentameric COMPASS catalytic core at 3.0Å
- Swd1 uses its C-terminal tail to form an inter-subunit regulatory pocket with Set1
- Swd3 stimulates Set1 activity through the inter-subunit pocket via its SMART motif
- A doorstop mechanism dictates substrate specificity of SET1/MLL methyltransferases

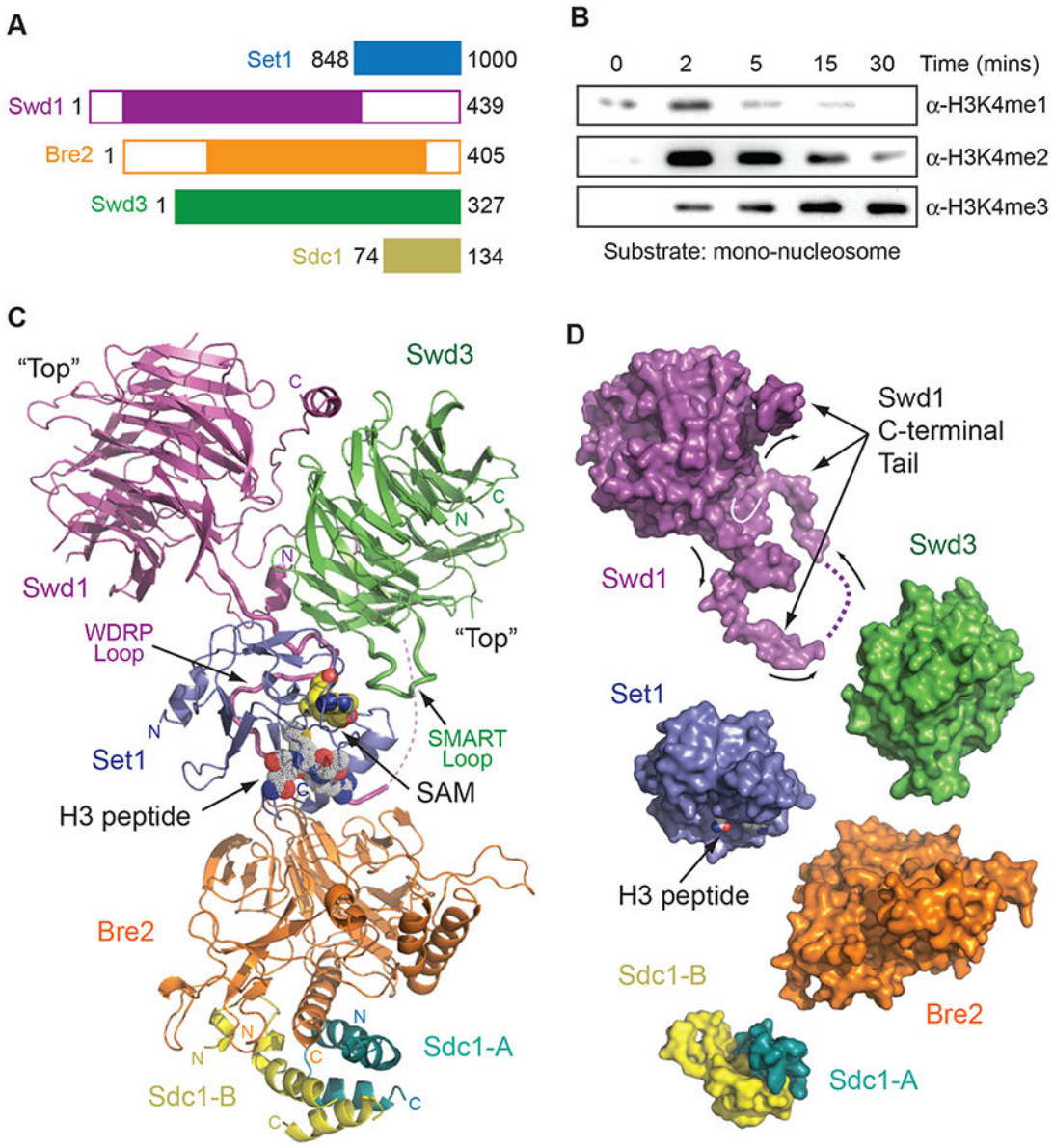


Figure 1. Overall architecture of the yeast COMPASS catalytic module

(A) Domain organization and construct design of yeast Set1, Swd1, Bre2, Swd3, and Sdc1. Only the SET domain of Set1 is shown.

(B) Time course of H3K4 methylation catalyzed by purified yeast COMPASS catalytic module with recombinant mono-nucleosome substrate. Reactions were quenched at specified time points, and analyzed by SDS-PAGE, followed by western blotting using methylation specific antibodies. Time point 0 min was taken seconds after addition of enzyme to the reaction.

(C) Overall structure of the COMPASS catalytic module. Set1 (blue), Swd1 (magenta), Swd3 (green), Bre2 (orange), and two copies of Sdc1 (yellow and turquoise) are shown in cartoon form. Both the H3 peptide (white) and SAM cofactor (yellow) are shown in space

filling model. The long C-terminal tail of Swd1 and unique SMART loop of Swd3 are shown in tube representation to highlight their positions in the complex.
(D) Individual subunits of the catalytic module are shown in surface representation to illustrate their relative spatial relationships. Orientation of the subunits is the same as shown in (C). For Swd1, from left to right, arrows show the winding path of the C-terminal tail. A dashed line in the tail illustrates a disordered region not observed in the structure.

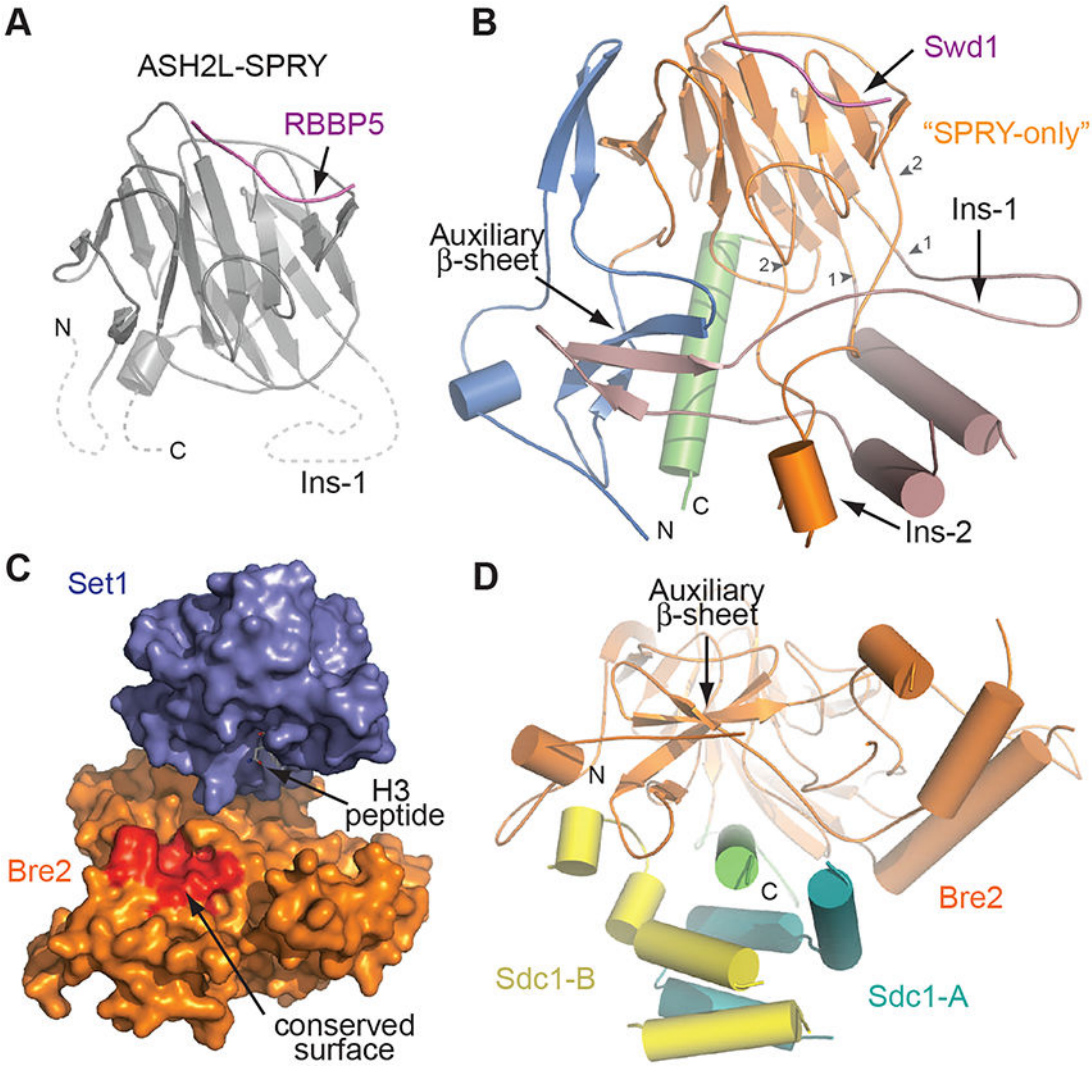


Figure 2. Bre2/ASH2L fold into an extended non-canonical SPRY domain

(A) Crystal structure of human ASH2L (grey) bound to a peptide of RBBP5 (magenta) (PDB: 4X8P). Dashed lines indicate the absent N-, insertion, and C-terminal sequences in the structure.

(B) Structure of yeast Bre2. The conserved SPRY-only domain and Ins-2 is shown in orange, while the N-terminal pre-SPRY and Ins-1 are shown in blue and pale pink, respectively. The C-terminal helix is colored in green. Small arrowheads indicate the terminal ends of the insertion elements. The Swd1 peptide equivalent to that of RBBP5 in (A) is shown in purple. Structure is depicted in smooth lines representation for clarity.

(C) Surface representation of Set1 (blue) and Bre2 (orange). A highly conserved surface formed by the auxiliary β-sheet in Bre2 is highlighted in red.

(D) View of Bre2 looking down the C-terminal helix (green), illustrating the position of the auxiliary β-sheet right above the helix. Sdc1-A (yellow) and Sdc1-B (turquoise) prop up this sheet.

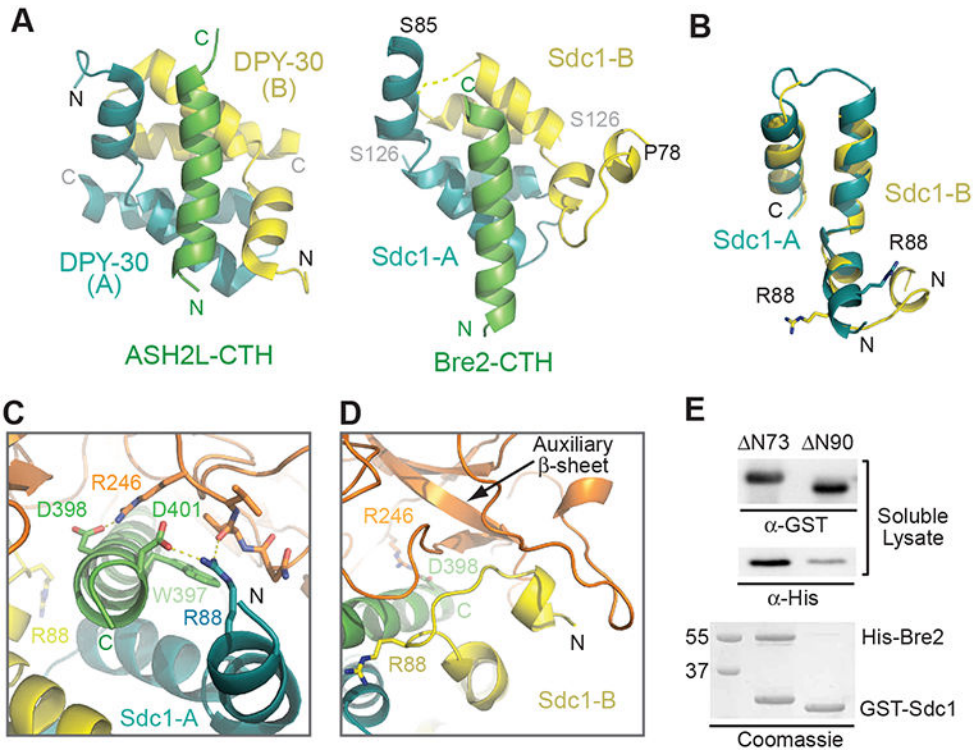


Figure 3. Sdc1 assembles as an asymmetric dimer to bind Bre2

(A) Left panel: structure of DPY-30 (turquoise and yellow) bound to the C-terminal helix of ASH2L (green) (PDB: 4RIQ). Right panel: equivalent view of Sdc1 (yellow and turquoise) bound to the Bre2 C-terminal helix (green). A dashed line in Sdc1-B denotes a poorly structured loop that was not clearly resolved in the crystal.

(B) Superposition of the two protomers of Sdc1 highlighting the two different positions of the same R88 side chain in the individual copies.

(C) Close-up view of the interface between Sdc1-A (turquoise) and Bre2 (orange and green). R88 of Sdc1 forms a salt bridge with Bre2-D401 on the C-terminal helix and a backbone hydrogen bonding interaction with the body of Bre2.

(D) Close-up view of the interface between Sdc1-B (yellow) and Bre2 (orange and green). The same R88 seen in (C) is facing towards Sdc1-A with the rest of the N-terminus packing against Bre2 near the auxiliary β-sheet.

(E) Glutathione affinity co-purification of GST-Sdc1 N-terminal truncation mutants and His-Bre2 co-expressed in insect cell. Protein interactions were assessed by SDS-PAGE followed by Coomassie staining. Inputs from soluble lysates were detected using antibodies against the respective tags.

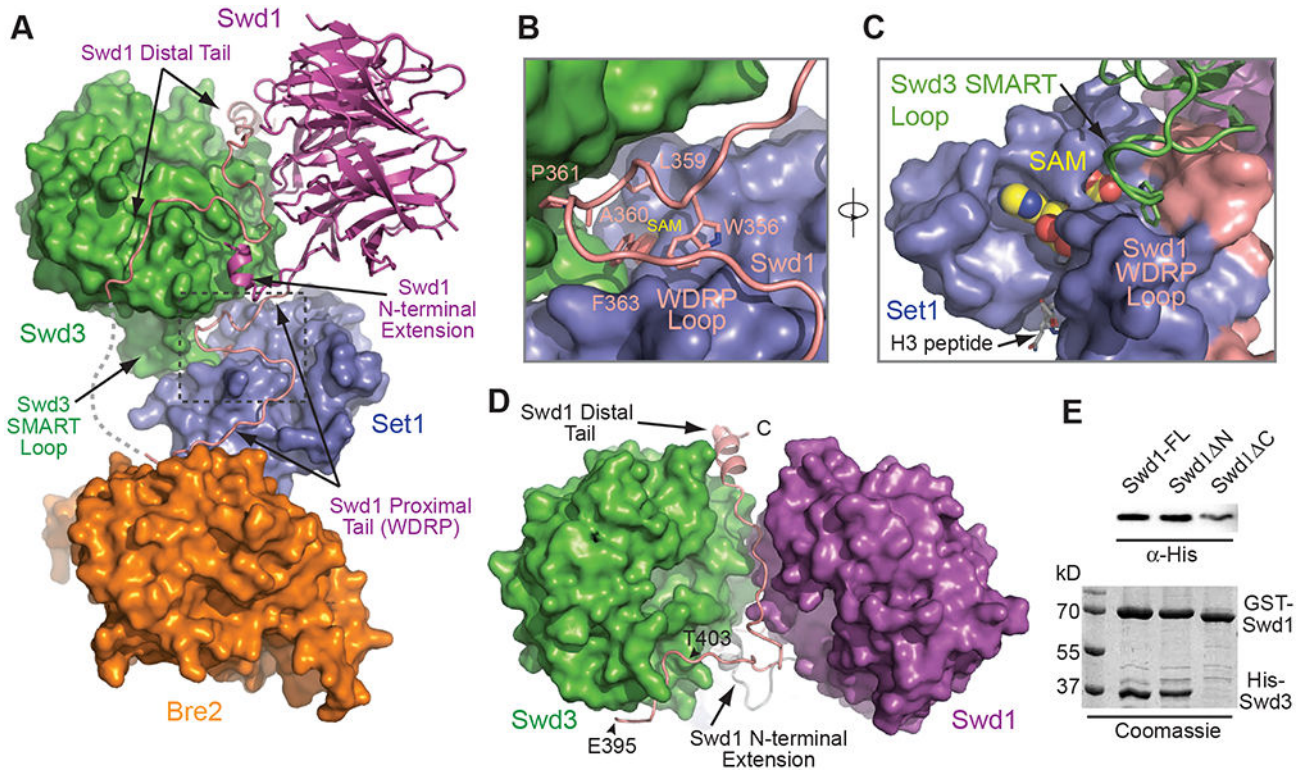


Figure 4. The C-terminal tail of Swd1 organizes the catalytic module and forms a key pocket with Set1

(A) A global view of the catalytic module highlighting the path of the Swd1 C-terminal tail. Set1 (blue), Swd3 (green), and Bre2 (orange) are shown in surface representation, while Swd1 (magenta) and its C-terminal tail (pink) are shown in cartoon form. A dotted line connecting the two halves of the C-terminal tail represents a disordered region not seen in the structure.

(B) Close-up view of the interface between Swd1-WDRP loop (pink) with Swd3 (green) and Set1 (blue). Swd3 and Set1 are shown in surface representation, while the WDRP region is shown in cartoon. Highly conserved Swd1 residues involved in interacting with both Set1 and the Swd3 loop are shown in sticks. The two sub-domains of Set1 are labeled as SET-N/C and SET-I.

(C) Alternate view of the Set1-Swd1-Swd3 interface highlighting the crevasse formed between Set1 (blue) and Swd1-WDRP loop (pink), both of which are shown in surface representation. The residues from Swd1 that line this pocket are labeled. The Swd3 SMART loop (green) shown in cartoon inserts a tryptophan residue (sticks) into the Set1-Swd1 pocket. The SAM cofactor and the H3 peptide are shown in CPK form and sticks, respectively.

(D) Path of the Swd1 distal tail that recruits Swd3 to the catalytic module. The tail is shown in cartoon form in pink, while the WD40 domains of Swd3 and Swd1 are shown in surface representation in green and magenta, respectively.

(E) Glutathione affinity co-purification of full-length (FL) GST-Swd1 and N-terminal (ΔN) and C-terminal truncation (ΔC) mutants with His-Swd3 co-expressed in insect cells. Protein

interactions were assessed by SDS-PAGE followed by Coomassie staining. His-Swd3 input from soluble lysates were detected using antibodies against the His-tag.

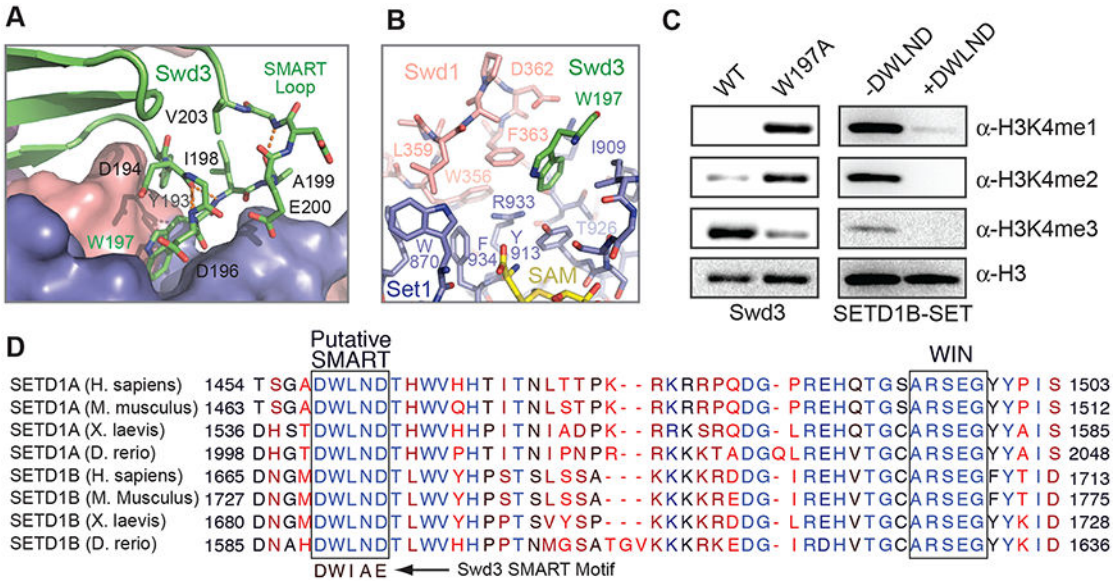


Figure 5. The SMART loop of Swd3 regulates Set1 activity through the Kabuki pocket

(A) Close up view of the Set1-Swd1-Swd3 interface, centered around the Swd3 SMART loop. Set1 and Swd1 (blue and pink respectively) are shown in surface, while the Swd3 loop is in cartoon form. The residues that form the SMART loop are shown in sticks. Dashed lines indicate hydrogen bonds.

(B) Detailed view of the residues that form the Kabuki pocket. On the Set1 side of the pocket are the amino acids W870, I909, Y913, T926, R933 and F934, while the cofactor SAM sits at the periphery. Swd1 contributes W356, L359, D362, and F363 to the formation of the pocket. The Swd3 tryptophan residue, W197, inserted into the pocket is shown in sticks without the rest of the Swd3 SMART loop.

(C) H3K4 methyltransferase activities of yeast and mammalian catalytic module containing wild type and mutant Swd3 and two different constructs of SETD1B against nucleosomes. Left panel shows the yeast catalytic module with Swd3 bearing a W197A mutation. Right panel shows the activity of the assembled SETD1B catalytic module containing the SETD1B C-terminal WIN-SET domain excluding (-DWLND) or including (+DWLND) the putative SMART motif shown in (D).

(D) Sequence alignment of mammalian SETD1A and SETD1B. The putative SMART motif of mammalian SETD1A/SETD1B is upstream of the conserved WDR5-interacting motif (WIN). For clarity, the SET domain of each SETD1A/B vertebrate ortholog C-terminal to the WIN motif is not shown. The SMART motif from *K. lactis* Swd3 is shown below the mammalian sequence as a comparison. A gradient of red to blue colors are used to indicate low to high degree of conservation.

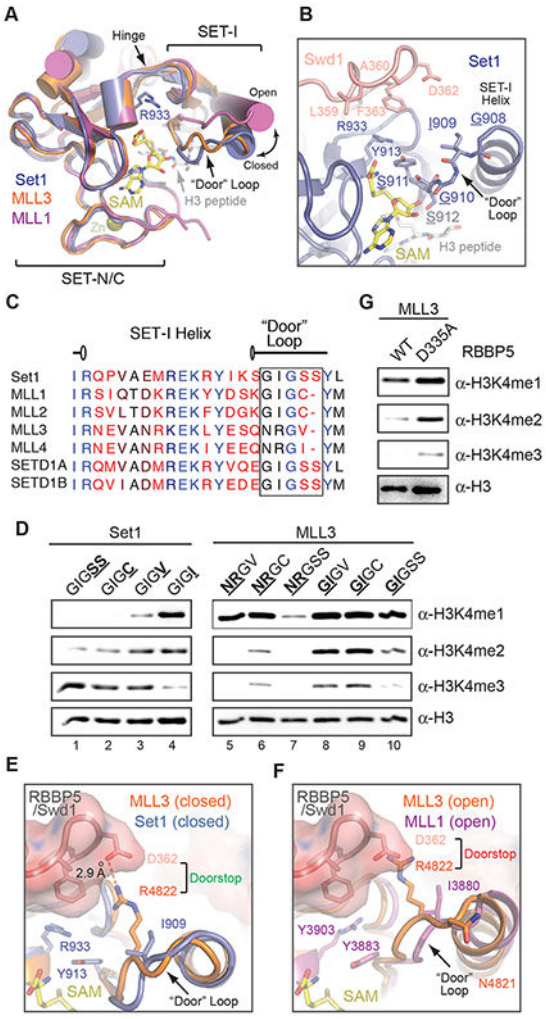


Figure 6. A "doorstop" mechanism determines substrate specificity in SET1/MLL enzymes

(A) Superposition of yeast Set1 (blue), MLL3 (orange, PDB: 5F6K chain C), and MLL1 (purple, PDB: 5F6I). Helices are shown as cylinders for clarity. Note the variability in the SET-I helix between the three structures superposed, shifting between an open and closed state. The side chain of the Arg933 residue at the bottom of the Kabuki pocket is shown in sticks as a reference point for orientation.

(B) View looking down the Kabuki pocket, highlighting the door loop of yeast Set1 (blue). The side chains of the door loop are shown as sticks. Residues that line the Kabuki pocket from the Swd1 side are shown in pink sticks. SAM and the H3 peptide are shown in yellow and grey sticks to illustrate the spatial relationship of the door loop to both the cofactor and substrate.

(C) Sequence alignment of yeast Set1 and human SET1/MLL family members surrounding the SET-I helix and the following door loop.

(D) H3K4 methyltransferase activity of yeast and mammalian catalytic module containing Set1 and MLL3 door loop mutants against nucleosome substrates. Left panel are Set1 mutants with mutations in the second half of the door loop motif showing gradual reduction of tri-methylation upon introduction of bulkier side chains. Left panel are mutants of MLL3-

WRAD with mutations in the doorstop position of the loop. The simple replacement of NR with GI results in robust gain of function for MLL3.

(E) Superposition of MLL3 in its closed conformation (orange, PDB: 5F6K chain C) with Set1 (blue) bound to Swd1 (pink). Arg4822 on MLL3 forms a perfect salt bridge with D362 of Swd1, whose corresponding residue in RBBP5 is D335. Doorstop in green indicates a geometrically allowed structural configuration.

(F) Superposition modeling of RBBP5/Swd1-bound MLL3 (orange, PDB: 5F6K chain E) in the experimentally determined open state and MLL1 (purple, PDB: 5F6I) in the open state based on the yeast catalytic module structure. Arg4822 of MLL3 clashes with Asp362 on Swd1, while Ile3880 of MLL1 can fully open even in the presence of Swd1/RBBP5. Doorstop in red indicates a geometrically disallowed structural configuration.

(G) H3K4 methyltransferase activity of the mammalian MLL3-centered catalytic module containing wild type (WT) RBBP5 and a RBBP5-D335A mutant against nucleosome substrates. The catalytic module with the RBBP5 mutant produced an increased amount of di-and tri-methylated H3K4 products relative to wild-type protein.