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): 1117–1126.e12. doi:10.1016/j.cell.2018.07.020.

Structure and conformational dynamics of a COMPASS histone H3K4 methyltransferase complex

Qianhui Qu[#1,2], **Yohhei Takahashi**[#3], **Yidai Yang**[4], **Hongli Hu**[1,2], **Yan Zhang**[1,2], **Joseph S. Brunzelle**[5], **Jean-Francois Couture**[4], **Ali Shilatifard**[3,7], and **Georgios Skiniotis**[1,2,7,8]

[1]Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, 94305-5345, USA

[2]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, 94305-5345, USA

[3]Department of Biochemistry and Molecular Genetics, Feinberg School of Medicine, Northwestern University, Chicago, IL, 60611, USA

[4]Department of Biochemistry, Microbiology, and Immunology, Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON K1H 8M5, Canada

[5]Northwestern Synchrotron Research Center, Life Sciences Collaborative Access Team, Northwestern University, Argonne, IL, 60439, USA

[8]Lead Contact

[#] These authors contributed equally to this work.

Summary

The methylation of histone 3 Lysine 4 (H3K4) is carried out by an evolutionarily conserved family of methyltransferases referred to as COMplex of Proteins ASsociated with Set1 (COMPASS). The activity of the catalytic SET domain (***S***u(var)3–9, ***E***nhancer-of-zeste and ***T***rithorax) is endowed through forming a complex with a set of core proteins that are widely shared from yeast to humans. We obtained cryo-electron microscopy (cryo-EM) maps of the yeast Set1/COMPASS core complex at overall 4.0–4.4 Å resolution, providing insights into its structural organization and conformational dynamics. The Cps50 C-terminal tail weaves within the complex to provide a central scaffold for assembly. The SET domain, snugly positioned at the junction of the Y-shaped

[7] Correspondence: yiorgo@stanford.edu or ash@northwestern.edu.

Author Contributions

Q.Q. performed cryo-EM data collection, map calculation, model building and refinement; Y.T. expressed and purified the complex and performed yeast genetic analysis; Y.Y. conducted *in vitro* biochemical assays and the crystallography study of Cps50; H.H. and Y.Z. assisted with cryo-EM data collection; J.S.B assisted with structural analysis; J.F.C. directed the Cps50 crystallographic studies and assisted in the structural analysis of core COMPASS; A.S. and G.S. supervised the overall project. All authors contributed to manuscript preparation. All data generated or analyzed during this study are included in this published article and its supplementary information.

Declaration of Interests

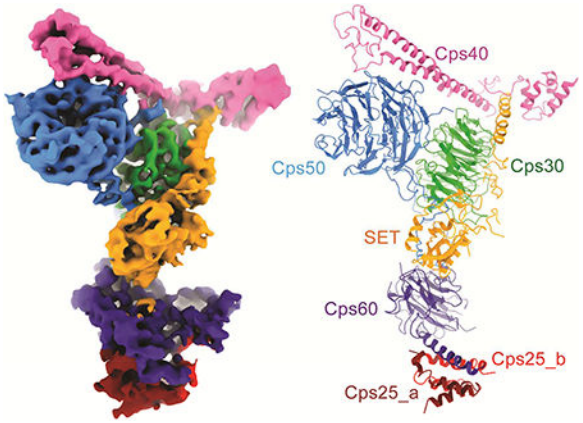
The authors declare no competing financial interests.

complex, is extensively contacted by Cps60 (Bre2), Cps50 (Swd1) and Cps30 (Swd3). The mobile SET-I motif of the SET domain is engaged by Cps30, explaining its key role in COMPASS catalytic activity towards higher H3K4 methylation states.

In Brief

The cryo-EM structure of the complete yeast COMPASS complex points to how it controls selective H3K4 mono-, di- and tri-methylation.

Graphical Abstract



Structure of a COMPASS histone H3K4 methyltransferase complex

Introduction

Members of the Set1/MLL (*M*ixed *L*ineage *L*eukemia) family of methyltransferases (KMT2) catalyze the methylation of Lysine 4 of histone H3 (H3K4) (Piunti and Shilatifard, 2016; Shilatifard, 2012). Conserved from yeast to human, the methyltransferase activity of these enzymes is associated with gene regulatory events that play a critical role in hematopoiesis, embryonic stem cell development and neurogenesis (Barski et al., 2007; Ruthenburg et al., 2007; Shilatifard, 2012; Sze and Shilatifard, 2016). Misregulation of the Set1/MLL family is highly correlated with diverse pathologies (Ng et al., 2010; Singh et al., 2016; Sze et al., 2017; Wang et al., 2018), such as pediatric acute lymphocytic leukemia and adult acute myeloid leukemia where recurrent chromosomal translocation of the *MLL1* gene is frequently identified (Andersson et al., 2015; Slany, 2016; Yang and Ernst, 2017). Humans possess at least six MLL homologues with non-redundant functions (Shilatifard, 2012). In budding yeast *S. cerevisiae,* the one-and-only Set1/COMPASS complex is responsible for the mono-, di- and tri-methylation of H3K4 (Bernstein et al., 2002; Briggs et al., 2001; Krogan et al., 2002; Miller et al., 2001; Nagy et al., 2002). Akin to its human homologue, the Set1 protein on its own displays weak mono-methyltransferase activity but is positively allosterically regulated upon association with a multi-subunit complex comprised of Cps30, Cps50, Cps60 and Cps25 (Takahashi et al., 2011). This specific set of core components is shared among species. For instance, the WRAD complex (**W**DR5/**R**bBP5/**A**SH2L/**D**PY30, corresponding to Cps30/50/60/25) is shared in all six MLLs methyltransferases in human

cells (Couture and Skiniotis, 2013) and absence of any core component compromises, to different extents, the function of MLL/COMPASS-like complexes (Avdic et al., 2011a; Dou et al., 2006; Patel et al., 2009; Steward et al., 2006; Zhang et al., 2012).

Our earlier EM studies showed the subunit organization and a similar Y-shaped architecture adopted by both core COMPASS and human MLL (Takahashi et al., 2011), but lacked the resolution to reveal structural details. On the other hand, high resolution structures of MLLs' fragment or subunits from us and others have provided crucial insights (Avdic et al., 2011a; Couture et al., 2006; Li et al., 2016; Odho et al., 2010; Zhang et al., 2015), but do not capture the entire spectrum of interactions contributing to the formation of MLL complexes and likely hinder our understanding of hitherto unknown but potentially important regions of partner proteins regulating KMT2 enzymes. Here we employed single-particle cryo-EM, complemented by X-ray crystallography and functional assays, to obtain a detailed structural model for the yeast *S. cerevisiae* COMPASS complex including Set1(726–1080aa), Cps30, Cps50, Cps60, Cps25 and Cps40. The results highlight the essential scaffolding role of Cps50 for COMPASS assembly, the conformationally dynamic character of the complex, and the essential role of Cps30 on inducing higher methylation states of H3K4.

Results

COMPASS core architecture

For structural studies, we co-expressed full-length *S. cerevisiae* Cps60, Cps50, Cps40, Cps30, Cps25 and a truncated active Set1 protein (762–1080aa, hereafter, SET762) in *Spodoptera frugiperda* (Sf9) cells, aiming to purify the intact six-subunit catalytic core complex with a total molecular weight of ~220kD (Figure S1A). A fine-tuned baculovirus infection and purification strategy enabled us to reconstitute and isolate highly homogeneous preparations of the complex with optimal methyltransferase activity (Thornton et al., 2014), and we subsequently applied cryo-EM to record a single-particle projection dataset (Figures S1B and S1C). After consecutive rounds of 3D classification, we isolated a particle subset that enabled 3D reconstruction of a global map with overall 4.0-Å resolution, and maps of sub-regions with further improved features after focused refinement (Figures 1A and S2 to S4). In addition, a second independent particle subset led to a 4.4-Å map of COMPASS adopting a more compact configuration (Figures 1B and S2).

To model the structure of the COMPASS subunits in the cryo-EM map we employed a combination of homology modeling, secondary and tertiary structure prediction, and *de novo* chain building (see Methods). Our modeling was also aided by the structure of the WD40 domain of Cps50 from *Myceliophthola thermophila,* which we obtained by X-ray crystallography (Figure S3A). The structural model was first built against the locally refined maps of the top half (SET/Cps30/Cps40/Cps50) and bottom half (SET/Cps60/Cps25) COMPASS sub-regions that yielded higher quality density features (Figures S3C and S4), and subsequently refined against the overall 4.0-Å resolution map (Figure S3D). Any protein regions we could not confidently assign due to their flexible disposition, e.g. the highly mobile N-terminal PHD domain of Cps40 or segments of the Cps60–25 module, were omitted from the final model. The structure, complemented by an array of mutagenesis and

functional assays (discussed below), highlights the intricate organization and functional modulation of the 6-subunit COMPASS core complex (Figure 1C).

Consistent with its localization in our earlier studies, the catalytic SET domain is positioned at the junction of the Y-shaped complex (Figure 1) (Takahashi et al., 2011). The perpendicularly aligned Cps30 and Cps50 β-propeller domains form the upper tips of the "Y", wherein Cps30 directly contacts the SET domain. In contrast, Cps50 employs its C-terminal extension but not the WD40 domain to interact with the catalytic SET domain. The Cps40 subunit sits atop the "Y" to staple the two β-propeller domains via a pair of intertwined long a-helices. On the opposite end of the complex, the SPRY domain of Cps60 directly interacts with the SET domain of Set1 as well as a segment succeeding the Cps50 β-propeller domain. The last 20 residues of Cps60 fold into an α-helix that crosses orthogonally the two β sheets of its SPRY domain. This a-helical segment is embraced by a bundle of a-helices formed by a dimer of Cps25 (Figure S4). In this configuration, Cps60 and Cps25 form the lengthened tandem base of the COMPASS Y-shaped architecture.

The two COMPASS conformers allow for an indirect visualization of the structural dynamics underlying the complex in the absence of nucleosomal substrate. Comparisons of the two models reveal that there are no significant changes in the overall structure of each subunit and its interactions with other subunits (Figures 1B and S2E). Instead, the two conformers primarily differ in the positioning of two independent mobile components in the complex. One mobile element is the bottom base module (Cps60/Cps25), which maintains a main connection to the Cps50 C-terminal extension and is not stabilized by additional peripheral interactions with other COMPASS subunits; this characteristic allows for its relative variable disposition in respect to the larger upper module (SET/Cps30/Cps40/Cps50), as also evident by the lower resolution of its EM density. The second mobile element is the N-terminal region of Cps40, which is observed to reposition around a hinge region in the protein. As a result of both elements moving inwards towards each other, the second COMPASS conformer displays a more compact configuration on one side of the complex. Interestingly, this side forms an arc with dimensions that are compatible with those of a nucleosome. The center of the arc is occupied by the catalytic SET domain while its periphery presents structural elements that are involved in chromatin binding, such as the PHD domain of Cps40 which is disordered in our maps. PHD domains have been shown to recognize histone tails (Iwase et al., 2007; Lan et al., 2007; Li et al., 2006; Pena et al., 2006). It is thus tempting to speculate that this in-built conformational variability of the Cps60–25 module and the Cps40 N-terminal region may be important for capturing nucleosomes in order to methylate H3K4 (Figure 1D).

Cps50 orchestrates the assembly of the COMPASS core

The WD40 domain containing Cps50 protein or its mammalian homologue RbBP5 (Figure 2A) is essential for the methyltransferase activity of MLL/COMPASS-like complexes (Avdic et al., 2011a; Cao et al., 2010; Li et al., 2016; Odho et al., 2010; Takahashi et al., 2011). To start delineating the function of Cps50, we obtained a 1.8-Å resolution crystal structure of *Myceliophthola thermophila* Cps50 protein WD40 domain, which displays a similar structural fold to the canonical WD40 domain (Figure S3A). The cryo-EM map

shows that in addition to this WD40 β-propeller domain, Cps50 possesses a C-terminal wire-like density of approximately 80 amino acids (Figures 2B and S4). This density emanates from the Cps50 WD40 domain and knits along the Cps30 WD40 domain, the catalytic SET domain and Cps60 SPRY domain, and curls back towards the bottom cavity of Cps30 to finally exit through a slit between the two WD40 domains. The U-shaped wire establishes extensive contacts with most COMPASS subunits, including Cps50 itself (Figures 2 and 3). Although intensive attempts to recombinantly express individual *S. cerevisiae* proteins for *in vitro* assays were not successful, the homologous *Myceliophthola thermophila* Cps50 binds to and co-elutes with *Chaetomium thermophilum* Cps30, Cps40 or Cps60 (Figure S5), further suggesting the high conservation of these COMPASS core interactions among eukaryotes.

The short Cps50 N-terminus is entrapped by Cps30, the Cps50 WD40 domain, as well as the stretched linker region succeeding the Cps50 WD40 domain (Figure 3B). Of note, despite a relatively low sequence similarity, the amino acid length of this stretched linker is invariable among homologous Cps50/RbBP5 proteins (Figure S6). Three following adjacent short motifs on the C-terminal region of RbBP5 (Cps50 homolog) were found critical for MLL/COMPASS-like methyltransferase activity (Avdic et al., 2011a; Li et al., 2016; Zhang et al., 2015). The first functional motif (330–344aa) is referred to as the activation segment. In the COMPASS structure, the corresponding Cps50 amino acid sequence (348–366aa) fits well into an EM density that runs across the SET domain in an “S” fashion (Figures 2B and S4). Several residues at the beginning of the activation segment fold into a sharp U-turn and are tightly sandwiched between Cps30 and the SET domain (Figure 2C). More specifically, the activation segment is stabilized by surface residues located in the β17-strand of Cps30 and the SET-N as well as SET-C regions of the Set1 catalytic domain. Alanine substitution of the highly conserved “WASL” residues in Cps50 (Figure S6A) substantially lowers H3K4me3 levels in yeast, while H3K4me1 and H3K4me2 are marginally affected (Figure 2E).

Beyond the activation segment, the coiled portion (367–376aa) of Cps50/RbBP5 buttresses the interaction between Cps60 and the SET-I motif of the SET domain (Figure 2D). Mutation of residues on either Cps50 (residues 370–372, Figure 2E) or Cps60 (Arg183, Figure 2F) dramatically reduces COMPASS capacity to di- and tri-methylate H3K4 *in vivo.* Notably, swapping residue charges between Cps50 and Cps60 at this interface could restore the activities (Figure S6B). These observations further suggest the importance of this interaction surface and are consistent with our previous mutational analysis showing that a basic residue on ASH2L, which is encaged by several acidic residues of RbBP5, is important for the interaction of ASH2L with RbBP5 and subsequent stimulation of MLL methyltransferase activity (Zhang et al., 2015). Moreover, alanine substitution of the two adjacent acidic residues on Cps50 (D375A/E376A), which contacts the SET-I helix (Figure 2D), severely impairs H3K4 methylation in yeast cells (Figure 2E).

The Cps50 C-terminal tail further climbs onto Cps30 through the groove formed by the 5th and 6th β strands and brings the two WD40 domains into close proximity (Figure 3C). Intriguingly, mutation of Cps50 residues contacting Cps30 did not alter COMPASS activity *in vivo* (Figure 2E), suggesting that the integrity of COMPASS complex is well preserved by the extensive interaction network mediated by Cps50. The identified regions on Cps50 are

conserved among its homologous proteins across different species, including human RbBP5 (Figure S6A). We thus propose that Cps50 protein serves as an assembly and regulatory hub for COMPASS complex formation, a role which is likely applicable to the homologous RbBP5 proteins in other MLL/COMPASS-like methyltransferases.

Set1 is highly coordinated in COMPASS core complex

The Set1 protein and its MLL homologues display low basal mono-methyltransferase activity on their own. The enzyme gains optimal (mono-, di-, and tri-) methyltransferase activity when incorporated into the common set of partner proteins (Li et al., 2016; Takahashi et al., 2011; Wysocka et al., 2005). The Set1 construct used in this study (SET762), which covers the C-terminal catalytic SET domain, is fully active both *in vivo* and *in vitro* when reconstituted with its core subunits (Takahashi et al., 2011) (Figures 4 and 5). The cryo-EM map reveals a distinct density element corresponding to the n-SET region preceding the catalytic SET domain, which attaches to the Cps30 β-propeller domain, and climbs along its ridge all the way up to interact with Cps40 (Figure 4A). By combining bioinformatics analysis, the crystal structures of WDR5/MLLs complexes (Avdic et al., 2011a; Patel et al., 2008; Song and Kingston, 2008), and our mutagenesis functional analysis, we modeled the segment corresponding to residues 799–850 of the n-SET into the corresponding EM density. This region folds into a helix and a sharply curved loop (Figures 4B and 5A). Akin to a region known to interact with WDR5, the Cps30 homolog, which is referred to as WDR5 Interacting motif (Win-motif), this curved loop broadly attaches to Cps30 surface formed by residues located in blades B2, B3 and B4. Strikingly, the conserved basic residue K836 of the Win-motif (833–838aa) is inserted into the Cps30 cavity, in a similar way as the Arginine residue of MLLs Win-motif peptide interacting with the WDR5 protein (Figure 3B).

The N-terminal half of the n-SET helix is held by a bundle of four short stacked helices followed by long intertwined helices creating a “halberd” shape in Cps40 (Figures 4B and 5A). Truncation of this n-SET region reduces Set1 protein levels and diminishes global H3K4 methylation in yeast (Figure 5B). Analogously, knockout of Cps40 also substantially reduces Set1 protein levels and the methyltransferase activity *in vivo* (Dehe et al., 2006; Kim et al., 2013; Thornton et al., 2014) (Figure 5E). Cps40 also forms an extensive interface with the Cps50 WD40 domain (Figure 5C). Mutations of residues in these interfaces impair the interaction of Cps50 with Cps40 (Figure 5D). Similarly, internal deletion of the hinge region connecting the two Cps40 long a-helices decreases H3K4 tri-methylation by COMPASS in yeast cells (Figure 5E). These structural results provide the molecular basis for our previous observation that Cps40 is important for COMPASS integrity and function (Thornton et al., 2014).

The Win-motif that interacts with Cps30/WDR5 is conserved among Set1/MLL proteins. Similar to Cps40, knocking out Cps30 in yeast cells drastically decreases Set1 protein levels and H3K4 methylation. Surprisingly, mutations on either the Win-motif of Set1 (Figure 4D) or the corresponding interaction residues of Cps30 (Figure 6) do not appear to alter H3K4 methylation levels. We previously showed that a minimal Set1 construct which only contains the C-terminal catalytic SET domain (938–1080aa, SET938) maintained full capacity to

methylate H3K4 (mono-, di- and tri-) when reconstituted with Cps25/Cps30/Cps50/Cps60 *in vitro* (Takahashi et al., 2011). Moreover, a recent study also showed that the catalytic domain of all MLL homologues reconstituted with recombinant ASH2L/RbBP5/WDR5 could achieve methyltransferase activity *in vitro* (Li et al., 2016). Given that the truncation of n-SET region decreased the SET protein levels in yeast, these observations suggest that the n-SET region has a role in Set1 protein stability and activity *in vivo,* possibly through the interface with Cps40 and Cps30.

The C-terminal catalytic SET domain adopts the canonical topology that includes the SET-N, SET-I, SET-C and post-SET elements, similar to its homologous MLLs' SET domains. An extensive network of interactions is established among Cps30, Cps50, Cps60 and the SET domain (Figure 4C). Cps60 utilizes a concave surface located on the SPRY domain, which comprises two anti-parallel β sheets, to hold the bottom of the SET-I helix and the coiled region of Cps50. It has been hypothesized that an activated SET domain exhibits a smaller catalytic pocket than its inactive state, presumably due to the restraint upon the highly mobile SET-I motif. The comprehensive mutagenesis analysis of this interface indeed suggests that Cps50/Cps60 immobilize the SET-I motif by anchoring the catalytic domain at the base of the Y shaped architecture (Figure S6B). Notably, the cryo-EM map shows a stable density connecting the apical end of this SET-I helix to Cps30, implying that the Cps30 β-propeller may play an additional role on the methyltransferase activity of KMT2 enzymes (Figure 6A).

Cps30 establishes the tri-methylation activity of COMPASS

Cps30 is critical for the integrity and function of COMPASS in yeast cells (Figure 6B). Similarly, loss of WDR5 affects global H3K4 methylation levels in human cells (Wysocka et al., 2005). Besides the interface between the Win-motif of Set1 and Cps30, the COMPASS structure reveals an intersection between the SET-I motif and a loop connecting β strands 17 and 18 of Cps30 β-propeller (Figures 6A and S7A), a region we refer to as SIM (***S***ET-***I*** interacting ***M***otif). Despite variable lengths among Cps30/WDR5 proteins, the SIM region possesses several conserved negatively charged residues (Figure S7B), which are in close proximity to the SET-I motif and Set1 Win-motif (Figure 6A). Substitution of SIM region conserved acidic residues (D180/D182/E186) with basic amino acids results in nearly 50% loss of H3K4me3 level *in vivo,* compared to wild-type COMPASS complex. Of interest, exchanging the Cps30 SIM region with the corresponding fragment of wild-type mammalian WDR5 fully recapitulates COMPASS activity (Lane 7 in Figure 6B). Replacement of this region by alanine residues eliminates H3K4me3, while mono- and dimethylation levels are less affected (Lane 8 in Figure 6B).

To further validate the role of the Cps30 SIM region, we generated a chimeric Set1 protein by swapping a ten-residue *S. cerevisiae* SET-I loop region (983–992 aa), which contacts Cps30, with the corresponding segment (residues 2336–2344) of Drosophila *trithorax-related* protein (dTrr). dTrr is one of three Set1 homologues identified in *Drosophila melanogaster*, where it functions as a major H3K4 mono-methyltransferase on enhancers (Herz et al., 2012). The chimeric Set1 protein lost most of its H3K4 tri-methyltransferase activity in yeast (Figure 6C). Given that the Cps30-like WDR5 protein is shared by all three

Set1/dTrx/dTrr COMPASS-like complexes in Drosophila, we propose that their differences in methylation activity may be conferred by the WDR5 SIM motif recognizing subtle differences in the SET-I motifs of the three Set1, dTrx and dTrr proteins. Comparison of the COMPASS SET domain with MLL1 and MLL3 SET domains co-crystallized with fragments of ASH2L and RbBP5 shows that the peptide binding cleft and the methyltransferase pore are of similar sizes, suggesting that Cps30 does not lead to significant structural rearrangements in the catalytic domain. However, the same analysis revealed that the relative orientation of Cps50/RbBP5 and Cps60/ASH2L to the SET domain varies among the three complexes (Figures S7C and S7D), indicating that Cps30 may also serve as the anchor point to properly tether and space the other subunits.

Discussion

A key step in the regulation of gene expression is the methylation on histone H3 lysine 4, which is added by the multi-component MLL/COMPASS-like histone methyltransferases. Many studies have shown the requirement of shared core-subunits for optimal methylation activity of catalytic SET protein. Here we present a 4-Å resolution structure of the yeast COMPASS catalytic core complex. Together with systematic biochemical analyses *in vitro* and yeast genetics *in vivo*, the structure provides molecular insights into the assembly and function of COMPASS and COMPASS-like complexes.

Cps50 provides the molecular scaffold for assembling the COMPASS complex. Earlier studies showed the importance of few short motifs at the Cps50/RbBP5 C-terminal unstructured region for methylation (Avdic et al., 2011a; Li et al., 2016; Zhang et al., 2015). In COMPASS, the winding Cps50 C-terminal tail positions nearly all the core subunits in contact with the catalytic SET domain. Abrogating the interaction between the tail and its partners affected the activity of COMPASS to different extents. Here we also solved the crystal structure of Cps50 N-terminal WD40 domain, which shares the conserved seven β-strand blades with Cps30 and other WD40 domain proteins. The Cps50 WD40 domain sandwiches its C-terminal end together with the Cps30 WD40 domain, potentially to orient the Cps30 protein in a strict geometry related to the catalytic SET protein. Moreover, the extensive contact between Cps50 WD40 domain and Cps40 likely anchors this helical protein for its function.

The COMPASS structure reveals that Cps40 is a primarily a-helical protein, in agreement with secondary structure predictions. In addition, a chromatin binding PHD domain is predicted to be located at its N-terminus. Even though we could not model this region due to its flexibility, truncation of this N-terminal PHD domain affected the methylation activity of COMPASS in yeast (Figure 5E), which may reflect a reduced affinity with chromatin. Earlier studies have shown that the PHD domains can recognize methylated histone lysine residue (H3K4me3 (Li et al., 2006; Pena et al., 2006) and H3K9me3 (Iwase et al., 2007)) or unmethylated H3K4 (Lan et al., 2007). We note that the spacing between the Cps40 N-terminal region (PHD domain and helical-stack domain, connected by a hinge to the two-long helical rods) and Cps60/Cps25 base is about 10nm wide, which is compatible with the dimensions of a nucleosome (Figure S7E). The difference in the two COMPASS complex conformations we observed mainly resides in the disposition of the Cps40 N-terminal half

including PHD domain and the Cps60/Cps25 base. It is plausible that such movement between Cps40 PHD domain and the Cps60/Cps25 module enables the complex to “flex” and adapt to the nucleosomal substrate where H3K4 is located (Figure S7E), although detailed COMPASS studies in complex with the nucleosome are required to delineate these events.

The high sequence identity of the residues forming the methyltransferase pore of KMT2 enzymes combined with their divergent ability to either mono (MLL3/MLL4), di-(MLL1/MLL2) or tri-methylate (SET1A/SET1B) H3K4 (also referred to as product specificity) have for many years remained paradoxical. A prevailing model is that SET domain alone would experience a constant fluctuation between an open inactive state and a closed active state. With Cps50(RbBP5) and Cps60(ASH2L) to immobilize the bottom part of mobile SET-I region, the SET domain could adopt a closed conformation and achieve mono- and dimethylation activity. The presence of Cps30/WDR5 is critical for the tri-methylation activity of COMPASS/MLL complexes. Our COMPASS structural model reveals that Cps30 directly contacts the catalytic SET domain, which may provide further stabilization for the closed active state of SET domain. Interestingly, based on our structural and functional results, the interface between the Cps30 SIM region and SET-I motif plays a crucial role in the regulation of tri-methylation. Mutation of the conserved acidic residues (D180/D182/E186) of the Cps30/WDR5 SIM region specifically reduced the tri-methylation level of H3K4. Given that one WDR5 protein is shared by all MLL complexes, it will be important to investigate the determinants of product specificity on other SET/MLL complexes. Overall, our observations that Cps30 SIM region modulates the product specificity of KMT2 enzymes support a model wherein the interplay between the Cps30/WDR5 SIM region and the surface of the catalytic domain dictates the product specificity of KMT2 enzymes and thus epigenetic signaling.

STAR Methods

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, Georgios Skiniotis (yiorgo@stanford.edu)

Experimental Model and Subject Details

For recombinant protein expression, we used *Spodoptera frugiperda* (Sf9) cells at 27 °C and *E. coli* strain Rosetta™ (DE3) pLysS at 18 °C, respectively. For yeast culture, all knock-out strains were haploid cells and purchased from the Yeast Knock-Out (YKO) collection (Dharmacon) except set1Δ and cps50Δ in the charge-swapping experiment, which were generated with a PCR-mediated gene disruption technique with HIS3 and LEU2 as selection marker, respectively. More details are described in the Method Details section.

Method Details

Expression, and purification of COMPASS

Saccharomyces cerevisiae Set1 starting at Met762 to the C-terminal end (SET762), and full-length Cps60, Cps50, Cps40, Cps30, and Cps25, the components of the extended core COMPASS, were cloned into pBacPAK8 vector with FLAG or His6-FLAG tag on the N-terminus. Each baculovirus was individually prepared and amplified by using BacPAK baculovirus expression system (Clontech), and mixed in a predetermined ratio to achieve stoichiometric expression in cells. Sf9 insect cells were co-transfected at 27 °C for 72–96 h before collection, and lysed in 10 mM Tris-HCl pH 8.0, 1.5 mM MgCl2, 10 mM KCl, supplemented with protease inhibitor cocktail (P8340, Sigma-Aldrich). The supernatant was cleared with centrifugation, supplemented with NaCl to be 500 mM, mixed with anti-FLAG M2 affinity agarose gel (Sigma-Aldrich), and gently stirred for 2 h at 4 °C. COMPASS-bound M2 resin was collected with centrifugation, washed with FLAG buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 μM $Zn(OAc)_2$ 10% glycerol), and loaded into a column. COMPASS was eluted with FLAG buffer supplemented with 0.4 mg/ml 3× FLAG peptide (ApexBio), concentrated with Amicon Ultra Centrifugal Filters (EMD Millipore), and loaded onto Superose 6 Increase 10/300 GL column pre-equilibrated with 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.5 mM TCEP with AKTA pure chromatography system (GE Healthcare).

Protein interaction assays

Complementary DNA corresponding to COMPASS subunits of thermophilic filamentous fungus *Chaetomium thermophilum* (Ct) for CtCps60, CtCps40, and CtCps30, and *Myceliophthola thermophila* (Mt) for MtCps50 were cloned into pET28a or pGEX4T3 vectors. Truncations and mutants of each thermophilic COMPASS subunit were generated using a site-directed mutagenesis kit (stratagene). Wild-type and mutant proteins were overexpressed as TEV cleavable His or GST tagged proteins in the *E. coli* strain Rosetta™ (DE3) pLysS using 0.2mM IPTG during 17h at 18 °C. Cells were harvested in 50mM sodium phosphate, 500mM NaCl and 5mM β-mercaptoethanol, 10% glycerol, 1% Triton X-100, lysed by sonication and clarified by centrifugation. His-tagged proteins were purified by Talon Co^{2+} affinity chromatography and following TEV cleavage, the proteins were further separated by size exclusion chromatography (Superdex 200). For size exclusion chromatography analysis, purified proteins were mixed at a 1:1.5 mole ratio of Cps50:Cps30, Cps40:Cps50 or Cps60:Cps50 the binding buffer (BB) composed of 50mM Tris pH8.0, 200mM NaCl, 5mM BME and 2% Glycerol for 2h at 4 °C. Complexes were separated by size exclusion chromatography using a Superdex 200 pre-equilibrated in the BB. For GST pull-down assays, cell lysates containing 2μg of GST or GST-tagged proteins were applied onto glutathione-sepharose beads during 1 hour and washed extensively using the BB supplemented with 0.05% Triton-X 100. These beads were then further incubated with 10μg of another COMPASS subunit in 1ml of BB-T buffer during 2h at 4 °C. Beads were collected by centrifugation (1000 rpm, 1min), and washed extensively with BB-T buffer, and subjected to elution using BB supplemented with 10mM reduced glutathione. The input and eluted samples were separated using SDS-PAGE and Coomassie staining.

Yeast strains

Yeast culture was performed using standard methods. All yeast stains used in this work were haploid cells. All knock-out strains were from the Yeast Knock-Out (YKO) collection (Dharmacon) except set1Δ and cps50Δ in the charge-swapping experiment, which were generated with a PCR-mediated gene disruption technique with HIS3 and LEU2 as selection marker, respectively. All truncation and point mutations of COMPASS subunits were generated with site-directed mutagenesis after cloning into pFA6a-kanMX6 (Set1, and Cps50 in the charge-swapping experiments) or pRS306 (Cps60, Cps50, and Cps30) vectors. Cps60, Cps50, and Cps30 were C-terminally FLAG-tagged. ADH1 terminator sequence was introduced after the stop codon of each cloned COMPASS subunit gene. Engineered COMPASS subunit genes were introduced into their original loci through homologous recombination with standard yeast transformation technique and G418 or uracil selection. The introduced gene sequences in the genome were verified with Sanger sequencing.

Western blotting analysis

Over-night grown yeast cultures were harvested, washed, suspended in nuclear isolation buffer (250 mM sucrose, 60 mM KCl, 12.5 mM NaCl, 5 mM MgCl2, 1mM CaCl2, 0.8% Triton X-100, supplemented with protease inhibitor cocktail (P8215, Sigma-Aldrich)), and lysed with glass-beads by using Mini-BeadBeater-1 (BioSpec) or multi-tube vortex mixer. Crude nuclei pellets were collected with centrifugation, mixed with SDS-Laemmli sample buffer, boiled, resolved by SDS-PAGE, transferred to nitrocellulose membrane, and probed with home-made anti-H3, anti-H3K4me1, anti-H3K4me2, anti-H3K4me3, and anti-Set1 antibodies.

Crystallization and Structure Determination

The WD repeat domain of Cps50 (mtCps50$^{1-347}$) was overexpressed in *E.coli* Rosetta cells (Novagen) as a TEV cleavable fusion protein with a hexa-histidine tag, induced by 0.2mM IPTG for 17 hours at 18°C. Cells were pooled and resuspended in 50mM sodium phosphate, 500mM NaCl, 5mM β-mercaptoethanol, 10% glycerol, 1% Triton X-100, lysed by sonication and clarified by centrifugation. Proteins were then purified by Talon Co^{2+} affinity chromatography and cleaved subsequently by TEV. Following gel filtration on Superdex 200 pre-equilibrated with a buffer composed of 50mM Tris pH8.0, 200mM NaCl, 5mM β-mercaptoethanol and 2% glycerol, the protein was concentrated and mixed with a mother liquor solution composed of 200 mM sodium citrate, 22% (w/v) polyethylene glycol 3350. Cps50 crystals were harvested, transferred into fomblin and flash-frozen in liquid nitrogen. A full dataset was collected at the 17-ID beamline at the Argonne National Laboratory's Advanced Photon Source and indexed using HKL-2000 (HKL Research) (Otwinowski and Minor, 1997). A search model generated by I-TASSER (Roy et al., 2010) was used to solve the structure of Cps50 by molecular replacement. Using PHASER, four molecules were placed in the asymmetric unit. Following several rounds of model building and refinement using Coot (Emsley et al., 2010) and PHENIX (Adams et al., 2010). The quality of the model was evaluated using Molprobity (Chen et al., 2010). Final refinement statistics are provided in Table S1.

Cryo-EM data collection

3μl protein aliquotes with a concentration of ~1.5mg/ml supplemented with 0.05% Octyl β-D-glucopyranoside were applied onto glow-discharged Lacey carbon grids, blotted with Vitrobot Mark IV (FEI Company) and flash frozen in liquid ethane. The grids were imaged at liquid nitrogen temperature on a FEI Titan Krios electron microscope operating at 300 kV. Cryo-EM images were recorded with a Gatan K2 Summit direct electron detector at a nominal magnification of 29,000X in counted mode, corresponding to a pixel size of 1.0 Å/pixel. A dose rate of ~9.0 electrons/$Å^2$ /s and defocus values ranging within −1.5~−3.5 μm were used. Total exposure of 10s per image were dose-fractionated into 50 movie frames, resulting in an accumulated dose of ~90 electrons per $Å^2$.

Cryo-EM image processing

Micrograph movie stacks were first subjected to MotionCorr2 (Zheng et al., 2017) for whole-frame and local drift correction, with subsequent dose filtering applied to the integrated images. Visual inspection was conducted to discard micrographs with large carbon area, obvious ice contamination and visible astigmatism, resulting in a working data set of 8298 cryo-EM images. CTFFIND4 (Rohou and Grigorieff, 2015) was used for contrast transfer function (CTF) parameter determination on binned 2X2 images. About 10,000 particles interactively picked with EMAN2 (Tang et al., 2007) were subjected to ISAC (Yang et al., 2012) for reference-free 2D classification, and averages were employed to reconstruct an *ab initio* 3D model by VIPER (Penczek et al., 1994). Automatic particle picking, 2D classification and 3D classification were conducted in RELION1.4 (Scheres, 2012). A total of ~1.5 million particles selected through 2D classification were subjected to a first round of 3D classification into eight classes (Figure S2A). Particles from classes with well-defined structural features from the first 3D classification of the two subsets were combined (731,504 particles) and submitted to a second round 3D classification producing eight classes. A partition of 163,539 particles combined from two classes were refined to 4.0-Å using cisTEM (Grant et al., 2018). The angular distribution of this projections contributing to this reconstruction shown in Figure S2B. Another partition with 92,070 particles with a more compact conformation was refined to 4.4-Å resolution. Resolution estimations was obtained in cisTEM with gold standard Fourier shell correlation using the 0.143 cut-off criterion (Figure S2C) and local resolution of 4.0-Å map is estimated by Bsoft (Figure S2D). The two conformers and models were aligned in Chimera (Figures 1B and S2E).

Model building and refinement

The two clearly resolved doughnut-like density modules leaning almost perpendicularly against each other represent the classical seven-blade WD40 domains from Cps30 and Cps50, which have ~30% sequence similarity. To correctly assign the WD40 domains, which would serve as a starting point to build the complete model, we considered the following: First, we observed a long winding density element extending from one of the doughnut lobes of the EM map, suggesting that it corresponds to the unstructured ~80 amino acid C-terminus of Cps50, whereas the entire Cps30 sequence folds into one single WD40 propeller domain. Second, this well-defined density could perfectly accommodate the

reported structural fragments of RbBP5, the human homologue of yeast Cps50. Homology models generated by MODELLER (Eswar et al., 2007) for Cps30 and Cps50 WD40 domains derived from WDR5 (PDB 2H14) and our mtCps50 WD40 domain crystal structure were fit into the EM density. To complete the Cps50 C-terminal tail modeling, two fragments (352–376 and 386–391aa) sharing high similarity with homologous RbBP5 in structures of MLL3-RbBP5-ASH2L (PDB ID: 5F6K) and WDR5-MLL1-RbBP5 (PDB ID: 3P4F) were fit with very good agreement into the EM density, which anchored the tracing of the remainder main chain. Models for Set1 SET domain, Cps60 SPRY domain and its C-terminal helix in complex with Cps25 C-terminal a-helical domain were also generated by MODELLER, using crystal structures MLL3-RbBP5-ASH2L (PDB ID: 5F6K), ASH2L-RbBP5 (PDB ID: 4X8P) and ASH2L-DPY30 (PDB ID: 4RIQ) as templates, respectively. For the n-SET region modeling, the secondary structure prediction analysis suggested a α-helix which matches the EM density. The subsequent Win-motif was built into well-resolved EM density guided from the WDR5-MLL1-RbBP5 complex structure (PDB ID: 3P4F). A rod-like density on top of the two WD40 domains ridges corresponds to the Cps40 protein (Thornton et al., 2014). We built a model for scCps40 by combining secondary structure prediction in Psipred (Buchan et al., 2013), 3D *de novo* model building in Robetta (Kim et al., 2004) and I-TASSER (Roy et al., 2010), and biochemical mutagenesis assays (Figure 5). Given high flexibility, or local resolution not sufficient for *de novo* modeling, some regions of the subunits were not built in the final model: Cps40 residues 1–116 which containing a PHD domain; Cps60 regions 1–86, 115–141, 243–351, 372–432aa; Cps25 1–124aa; and Set1 segments 762–798 and 851–924aa. All models were docked into the cryo-EM density in Chimera (Pettersen et al., 2004), and iteratively adjusted manually using Coot (Emsley et al., 2010). The models were subjected to global refinement and minimization in real space, by using the module *phenix.real space refine* from PHENIX (Adams et al., 2010) package with secondary restraints. The quality of the model was evaluated using Molprobity (Chen et al., 2010). The final refinement statistics are provided in Table S2. For model deposition, the side chains were stubbed in areas of low resolution unless supported by crystal structures.

Quantification and Statistical Analysis—No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Data and Software Availability—The cryo-EM density map has been deposited in the Electron Microscopy Data Bank (EMDB; https://www.ebi.ac.uk/pdbe/emdb/): EMD-7303 and the coordinate for COMPASS has been deposited in the Protein Data Bank (PDB; https://www.rcsb.org/): 6BX3. The crystal structure of *Myceliophthola thermophila* Cps50 WD40 domain has been deposited in Protein Data Bank: 6E29.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by NIH grant DK090165 (G.S.), Outstanding Investigator Award R35CA197569 from National Cancer Institute (A.S.) and a CIHR grant (J.-F. C.).

References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, 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]

Andersson AK, Ma J, Wang J, Chen X, Gedman AL, Dang J, Nakitandwe J, Holmfeldt L, Parker M, Easton J, et al. (2015). The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. Nature genetics 47, 330–337. [PubMed: 25730765]

Avdic V, Zhang P, Lanouette S, Groulx A, Tremblay V, Brunzelle J, and Couture JF (2011a). Structural and biochemical insights into MLL1 core complex assembly. Structure 19, 101–108. [PubMed: 21220120]

Avdic V, Zhang P, Lanouette S, Voronova A, Skerjanc I, and Couture JF (2011b). Fine-tuning the stimulation of MLL1 methyltransferase activity by a histone H3-based peptide mimetic. FASEB J 25, 960–967. [PubMed: 21135039]

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, and Zhao K (2007). High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837. [PubMed: 17512414]

Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, Kouzarides T, and Schreiber SL (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. Proceedings of the National Academy of Sciences of the United States of America 99, 8695–8700. [PubMed: 12060701]

Briggs SD, Bryk M, Strahl BD, Cheung WL, Davie JK, Dent SY, Winston F, and Allis CD (2001). Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes & development 15, 3286–3295. [PubMed: 11751634]

Buchan DW, Minneci F, Nugent TC, Bryson K, and Jones DT (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res 41, W349–357. [PubMed: 23748958]

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 VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, and Richardson DC (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66, 12–21. [PubMed: 20057044]

Couture JF, Collazo E, and Trievel RC (2006). Molecular recognition of histone H3 by the WD40 protein WDR5. Nature structural & molecular biology 13, 698–703.

Couture JF, and Skiniotis G (2013). Assembling a COMPASS. Epigenetics 8, 349–354. [PubMed: 23470558]

Dehe PM, Dichtl B, Schaft D, Roguev A, Pamblanco M, Lebrun R, Rodriguez-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. The Journal of biological chemistry 281, 35404–35412. [PubMed: 16921172]

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. Nature structural & molecular biology 13, 713–719.

Emsley P, Lohkamp B, Scott WG, and Cowtan K (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66, 486–501. [PubMed: 20383002]

Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, and Sali A (2007). Comparative protein structure modeling using MODELLER Current protocols in protein science Chapter 2, Unit 2.9

Grant T, Rohou A, and Grigorieff N (2018). cisTEM, user-friendly software for single-particle image processing. eLife 7.

Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, 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 & development 26, 2604–2620. [PubMed: 23166019]

Iwase S, Lan F, Bayliss P, de la Torre-Ubieta L, Huarte M, Qi HH, Whetstine JR, Bonni A, Roberts TM, and Shi Y (2007). The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. Cell 128, 1077–1088. [PubMed: 17320160]

Kim DE, Chivian D, and Baker D (2004). Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32, W526–531. [PubMed: 15215442]

Kim J, Kim JA, McGinty RK, Nguyen UT, 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. The Journal of biological chemistry 277, 10753–10755. [PubMed: 11805083]

Lan F, Collins RE, De Cegli R, Alpatov R, Horton JR, Shi X, Gozani O, Cheng X, and Shi Y (2007). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. Nature 448, 718–722. [PubMed: 17687328]

Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, and Patel DJ (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature 442, 91–95. [PubMed: 16728978]

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]

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. Proceedings of the National Academy of Sciences of the United States of America 98, 12902–12907. [PubMed: 11687631]

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. Proceedings of the National Academy of Sciences of the United States of America 99, 90–94. [PubMed: 11752412]

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. Nature genetics 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. The Journal of biological chemistry 285, 32967–32976. [PubMed: 20716525]

Otwinowski Z, and Minor W (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276, 307–326.

Patel A, Dharmarajan V, and Cosgrove MS (2008). Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. The Journal of biological chemistry 283, 32158–32161. [PubMed: 18829459]

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. The Journal of biological chemistry 284, 24242–24256. [PubMed: 19556245]

Pena PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R, and Kutateladze TG (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature 442, 100–103. [PubMed: 16728977]

Penczek PA, Grassucci RA, and Frank J (1994). The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy 53, 251–270. [PubMed: 8160308]

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry 25, 1605–1612. [PubMed: 15264254]

Piunti A, and Shilatifard A (2016). Epigenetic balance of gene expression by Polycomb and COMPASS families. Science 352, aad9780. [PubMed: 27257261]

Rohou A, and Grigorieff N (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol 192, 216–221. [PubMed: 26278980]

Roy A, Kucukural A, and Zhang Y (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5, 725–738. [PubMed: 20360767]

Ruthenburg AJ, Allis CD, and Wysocka J (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol Cell 25, 15–30. [PubMed: 17218268]

Scheres SH (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 180, 519–530. [PubMed: 23000701]

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]

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. Nature neuroscience 19, 571–577. [PubMed: 26974950]

Slany RK (2016). The molecular mechanics of mixed lineage leukemia. Oncogene 35, 5215–5223. [PubMed: 26923329]

Song JJ, and Kingston RE (2008). WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. The Journal of biological chemistry 283, 35258–35264. [PubMed: 18840606]

Steward MM, Lee JS, O'Donovan A, Wyatt M, Bernstein BE, and Shilatifard A (2006). Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. Nature structural & molecular biology 13, 852–854.

Sze CC, Cao K, Collings CK, Marshall SA, Rendleman EJ, Ozark PA, Chen FX, Morgan MA, Wang L, and Shilatifard A (2017). Histone H3K4 methylation-dependent and -independent functions of Set1A/COMPASS in embryonic stem cell self-renewal and differentiation. Genes & development.

Sze CC, and Shilatifard A (2016). MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer. Cold Spring Harb Perspect Med 6.

Takahashi YH, 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. Proceedings of the National Academy of Sciences of the United States of America 108, 20526–20531. [PubMed: 22158900]

Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, and Ludtke SJ (2007). EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol 157, 38–46. [PubMed: 16859925]

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

Wang L, Zhao Z, Ozark PA, Fantini D, Marshall SA, Rendleman EJ, Cozzolino KA, Louis N, He X, Morgan MA, et al. (2018). Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer therapy. Nature medicine 24, 758–769.

Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, Roeder RG, Brivanlou AH, and Allis CD (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell 121, 859–872. [PubMed: 15960974]

Yang W, and Ernst P (2017). SET/MLL family proteins in hematopoiesis and leukemia. International journal of hematology 105, 7–16. [PubMed: 27796741]

Yang Z, Fang J, Chittuluru J, Asturias FJ, and Penczek PA (2012). Iterative stable alignment and clustering of 2D transmission electron microscope images. Structure 20, 237–247. [PubMed: 22325773]

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

Zhang P, Lee H, Brunzelle JS, and Couture JF (2012). The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. Nucleic Acids Res 40, 4237–4246. [PubMed: 22266653]

Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, and Agard DA (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14, 331–332. [PubMed: 28250466]

Highlights

1. Cryo-EM structure of a fully active H3K4 methyltransferase COMPASS core
2. The assembly of COMPASS is facilitated by Cps50
3. Catalytic SET protein is structurally coordinated by Cps60, Cps50 and Cps30
4. The methylation product specificity of COMPASS is modulated by Cps30

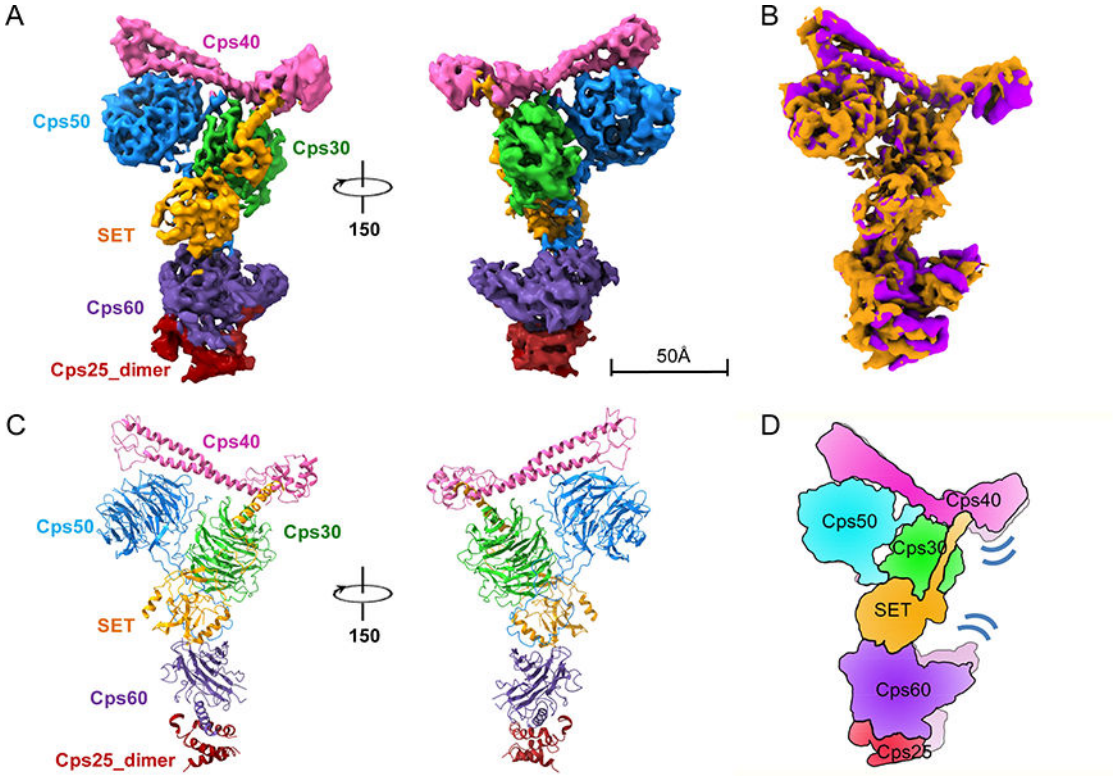


Figure 1. Architecture of core COMPASS.

(A) Views of the overall 4.0-Å cryo-EM density map of COMPASS with subunits rendered in different colors (Cps40 in pink, Cps50 in blue, Cps30 in green, SET762 in orange, Cps60 in purple, Cps25 in dark red). (B) Superimposition of 4.0-Å and 4.4-Å cryo-EM maps shown in orange and purple, respectively. (C) Structural model of the COMPASS in the same orientations and colors as shown in (A). (D) A schematic model of COMPASS structural dynamics adapts to the chromatin environment. See also Figures S1–S4 and S7.

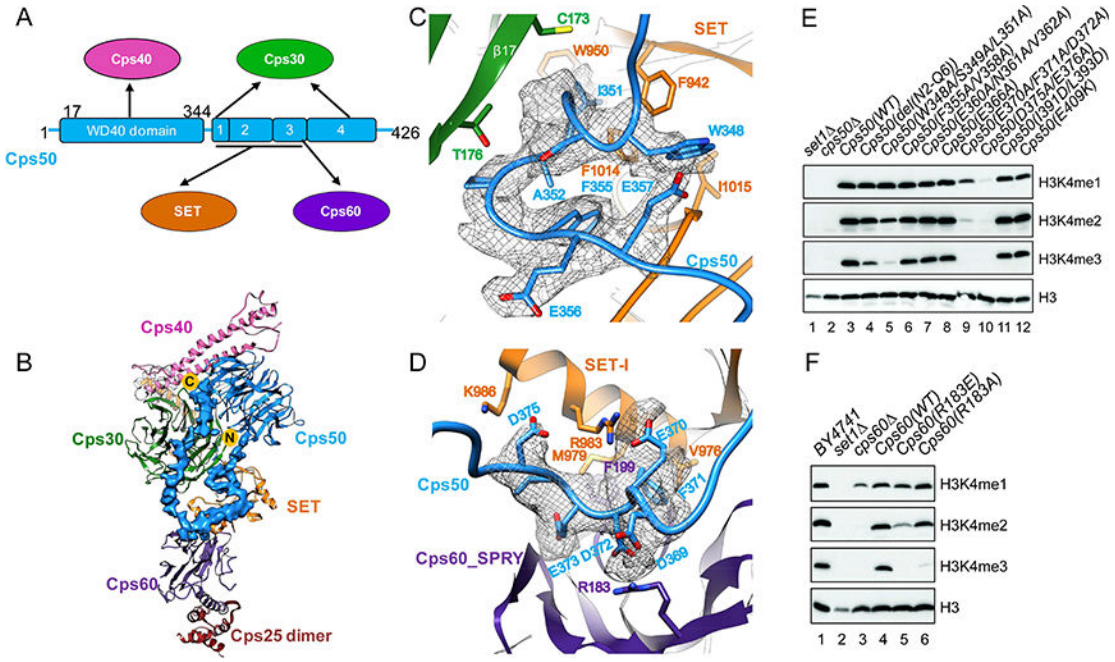


Figure 2. Cps50 serves as the assembly and regulatory hub for COMPASS.

(A) Interaction map of Cps50 with its partners. Regions 1–4 span sequences as follows: 1. 348–351aa; 2. 352–366aa; 3. 367–376aa; 4. 387–409aa. (B) Cryo-EM density map (colored in blue) of the Cps50 C-terminal region, which contacts most of the COMPASS subunits. (C and D) Close-up views of Cps50 interaction sites with Cps30 and SET domain (C), and Cps60 and SET domain (D). The residues involved in protein-protein interactions modeled according to EM map features, homologous structures and biochemical data are shown as stick in (C) and (D). The Cps50 residues involved in interactions are shown in their corresponding EM density (wire frame). (E and F), Representative mutagenesis analysis of H3K4 methylation profiles from yeast cells with mutated Cps50 (E) or Cps60 (F), respectively. See also Figures S4–S6.

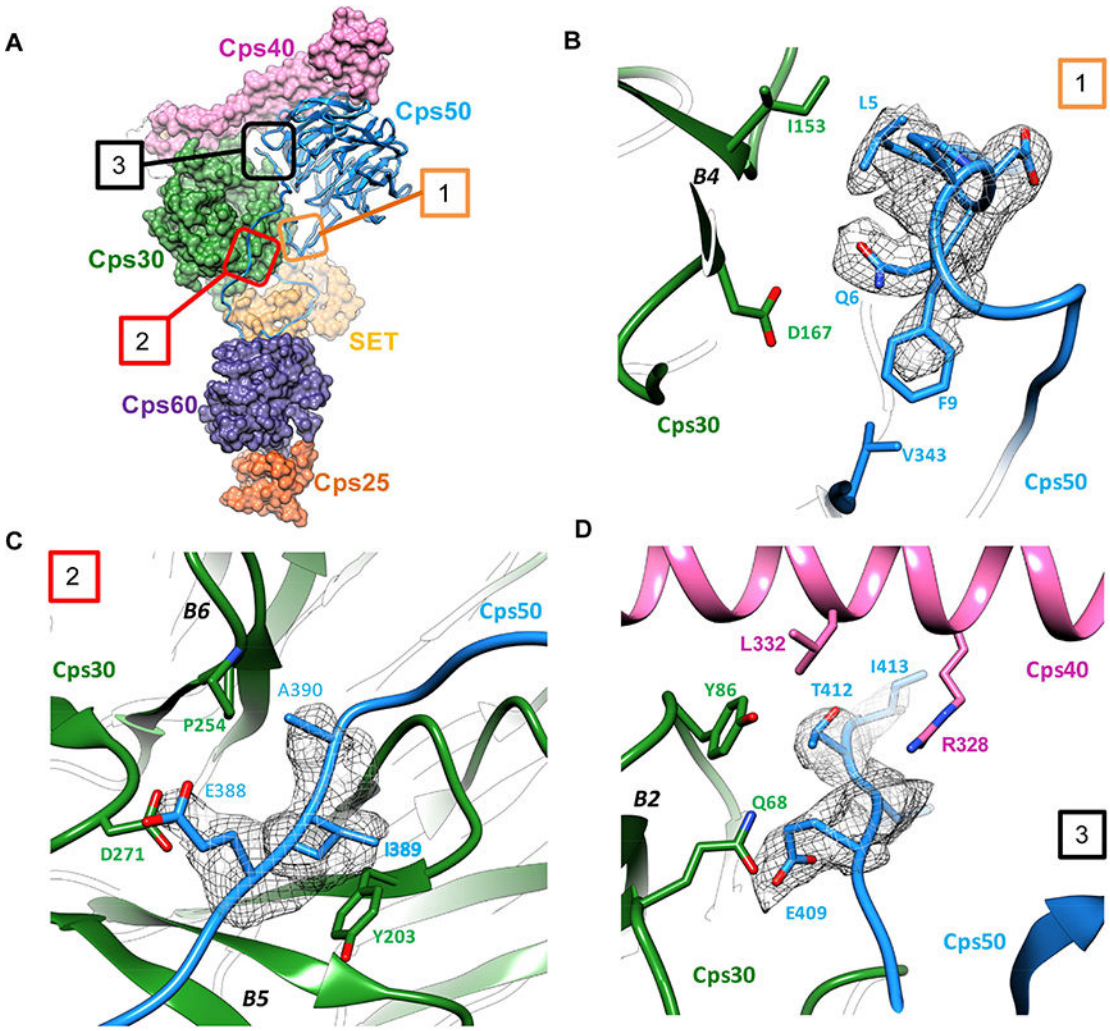


Figure 3. Interface details between Cps50 and its neighboring subunits.

(A) The Cps50 protein (shown as blue ribbon) connects COMPASS subunits (surface rendering). (B to D) Zoom-in views of boxed areas 1–3. (B) The conserved Cps50 N-terminus interacts with Cps30 mainly through the outer sides of 4th and 5th β sheets, and also associates with the stretched region (342–347aa) emanating from the Cps50 WD40 domain. (C) A short conserved motif in the Cps50 C-terminus is engulfed in the groove formed between the Cps30 5th and 6th β-sheets. (D) The extended Cps50 C-terminus is embraced by the Cps30 and Cps50 WD40 domains, and the Cps40 elongated helix. See also Figures S4–S6.

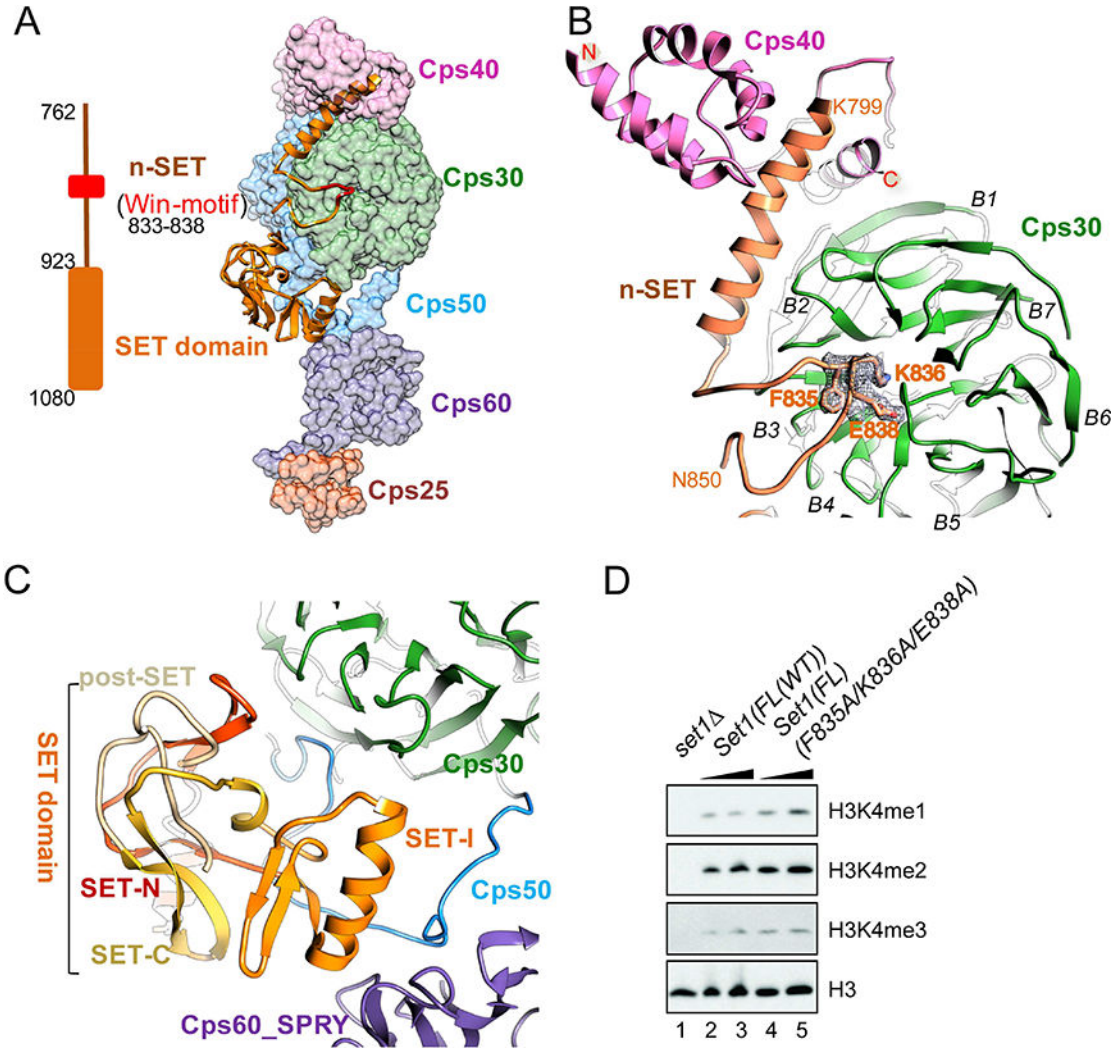


Figure 4. Set1 is highly coordinated in COMPASS.

(A) Schematic (left) of SET762 with Win-motif (red) and SET domain (orange), and structure model (right) of SET762 with directly associated subunits in surface representation (right). The regions 762–798 and 851–924aa are not assigned in the structure due to intrinsic flexibility. (B and C) Zoom-in views at the interfaces between SET762 and its partners. (B) The n-SET helix region and Win-motif interact with Cps40 and Cps30, respectively. The Win-motif residues involved in interactions are shown in their corresponding EM density. (C) The catalytic SET domain is engaged in a quadripartite interface with Cps30, Cps50 and Cps60. (D) Mutation of the Win-motif region does not have an observable effect on COMPASS activity *in vivo*.

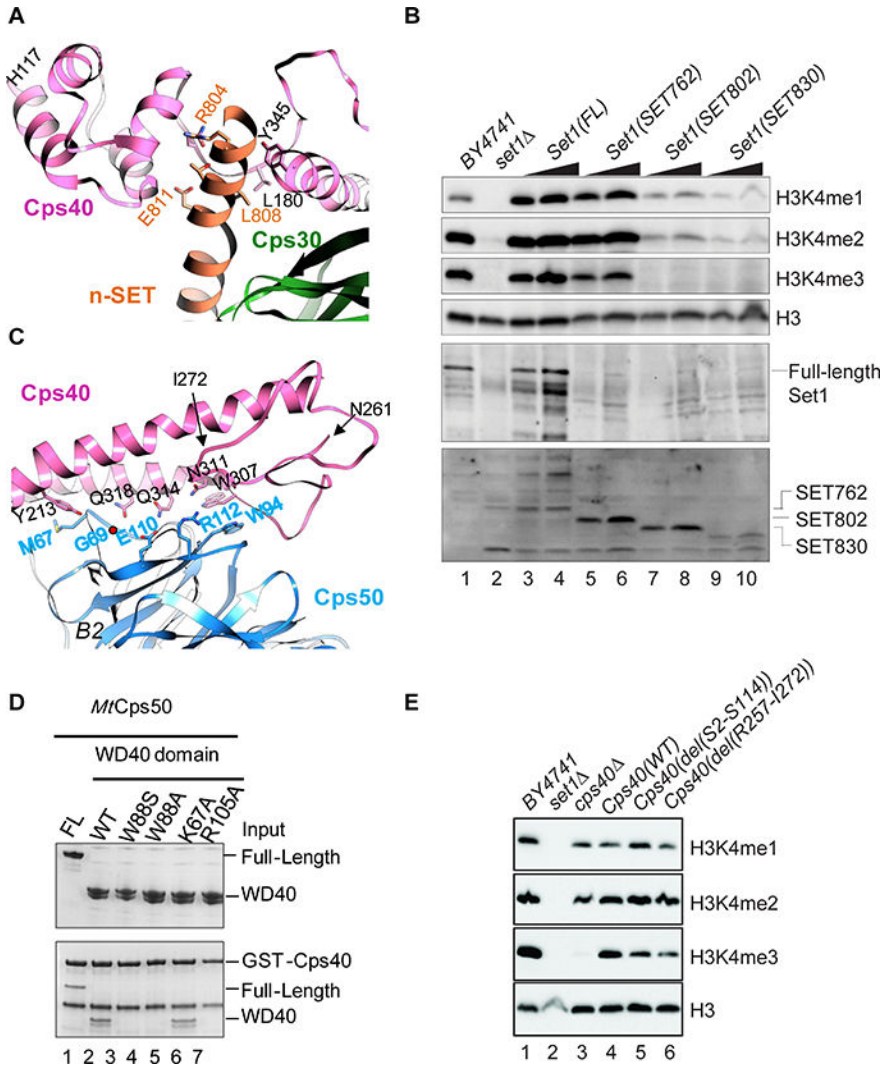


Figure 5. Cps40 is important for COMPASS integrity and function.

(A) Interface among n-SET, Cps30 and Cps40. The helical segment of n-SET leans against the Cps30 WD40 domain and stabilized by the Cps40 cleft formed between its N-terminal helical stack and the C-terminal elongated helices. (B) *In vivo* analysis of the n-SET region important for COMPASS activity. Western blotting detection of H3K4 methylation in cells expressing full-length Set1, N-terminally truncated variants of Set1 protein (SET762, SET802 and SET830). (C) Zoom-in view of the interface between Cps40 and Cps50. (D) Recombinant *Mt*Cps50 proteins with mutations corresponding to interfacial residues on *Sc*Cps50 shown in (C) were pulled down by GST-*Ct*Cps40 protein. *Mt*Cps50: K67,W88, R105 are relative to *Sc*Cps50: R73, W94, R112. (E) Effects of *Sc*Cps40 mutants on COMPASS activity.

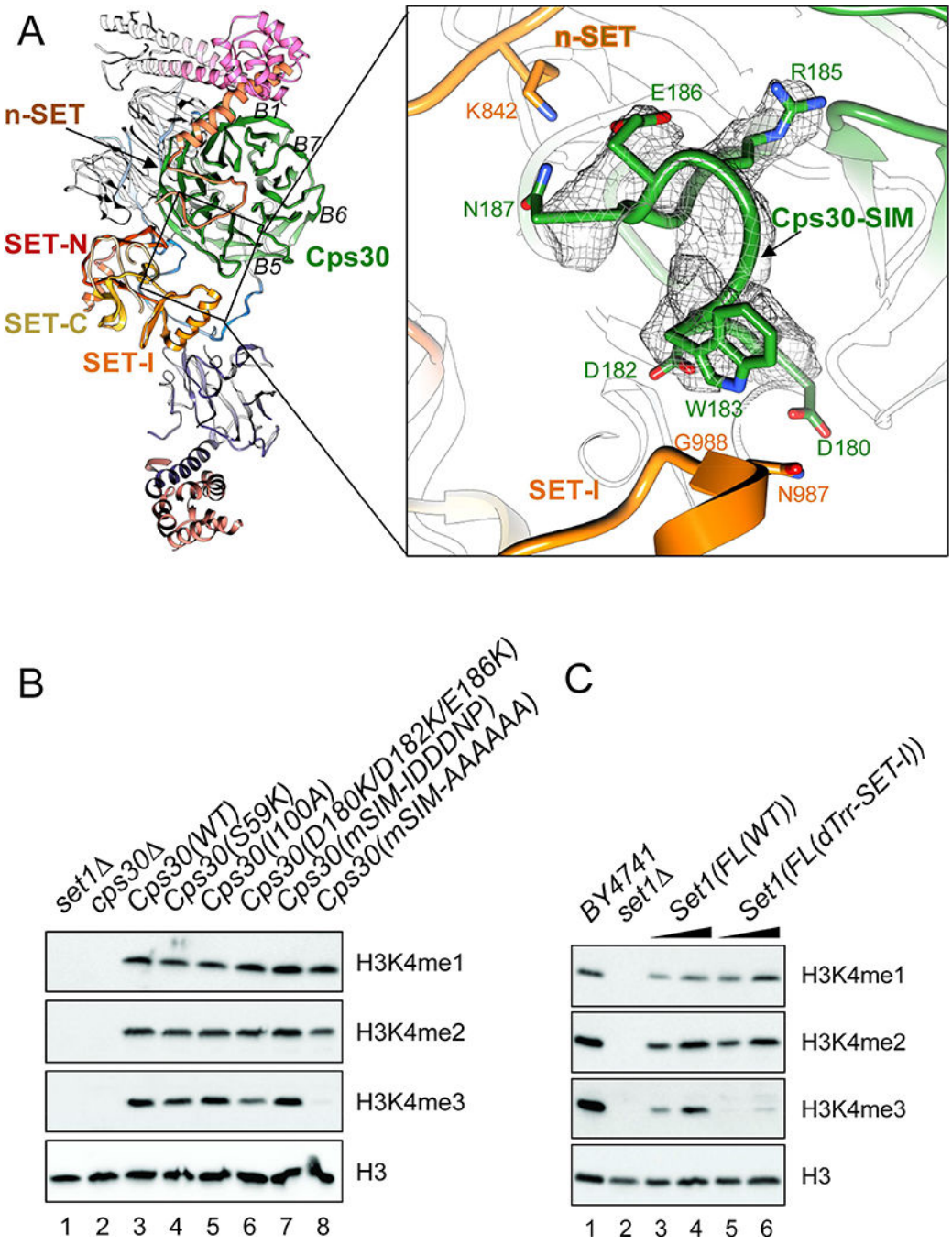


Figure 6. Cps30 establishes COMPASS tri-methylation activity.

(A) View of the COMPASS structure focusing on Cps30 interactions with SET762. SET762 is divided into sub-domains and colored as in Figure 4C. Zoom-in view of Cps30-SIM loop contacts with SET-I motif (right). Conserved residues in Cps30-SIM in their corresponding EM density (wire frame), with the most proximal interacting residues from SET-I motif and Win-motif shown in stick. (B) Western blot analysis of *In vivo* H3K4 methylation profiles of Cps30 mutant cells targeting its Win-motif binding site and SIM integrity. (C) The SET-I region (residues 983–992) of *Sc*Set1 is replaced with residues 2336–2344 of *Drosophila*

melanogaster Trr (dTrr), and the resulting chimeric Set1(FL(dTrr^{SET-I})) mutant cells are subjected to western blot analysis. See also Figure S7.