



# Multistate structures of the MLL1-WRAD complex bound to H2B-ubiquitinated nucleosome

Sanim Rahman<sup>a,1</sup>, Niklas A. Hoffmann<sup>a,1</sup>, Evan J. Worden<sup>b</sup>, Marissa L. Smith<sup>c</sup>, Kevin E. W. Namitz<sup>c</sup>, Bruce A. Knutson<sup>c</sup>, Michael S. Cosgrove<sup>c</sup>, and Cynthia Wolberger<sup>a,2</sup>

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2019. Contributed by Cynthia Wolberger; received April 12, 2022; accepted August 9, 2022; reviewed by Andres Leschziner and Nicolas Thomä.

The human Mixed Lineage Leukemia-1 (MLL1) complex methylates histone H3K4 to promote transcription and is stimulated by monoubiquitination of histone H2B. Recent structures of the MLL1-WRAD core complex, which comprises the MLL1 methyltransferase, WDR5, RbBp5, Ash2L, and DPY-30, have revealed variability in the docking of MLL1-WRAD on nucleosomes. In addition, portions of the Ash2L structure and the position of DPY30 remain ambiguous. We used an integrated approach combining cryoelectron microscopy (cryo-EM) and mass spectrometry cross-linking to determine a structure of the MLL1-WRAD complex bound to ubiquitinated nucleosomes. The resulting model contains the Ash2L intrinsically disordered region (IDR), SPRY insertion region, Sdc1-DPY30 interacting region (SDI-motif), and the DPY30 dimer. We also resolved three additional states of MLL1-WRAD lacking one or more subunits, which may reflect different steps in the assembly of MLL1-WRAD. The docking of subunits in all four states differs from structures of MLL1-WRAD bound to unmodified nucleosomes, suggesting that H2B-ubiquitin favors assembly of the active complex. Our results provide a more complete picture of MLL1-WRAD and the role of ubiquitin in promoting formation of the active methyltransferase complex.

MLL1 | chromatin | ubiquitin | cryo-EM | methyltransferase

Enzymes that deposit posttranslational modifications on the core histones, H2A, H2B, H3, and H4, play a central role in regulating transcription in all eukaryotes (1). Modifications including acetylation, methylation, and ubiquitination impact chromatin structure and recruit enzymes that modulate transcription. Monomethylation, dimethylation, and trimethylation of Lys-4 of histone H3 (H3K4) are hallmarks of actively transcribed genes that stimulate hyperacetylation of histones by the SAGA complex (2, 3), initiate the assembly of the transcription preinitiation complex (4, 5), and recruit nucleosome remodeling enzymes such as CHD1 (6) and NURF (7). H3K4 methylation, in most cases, depends upon monoubiquitination of histone H2B K120 (H2B-K120Ub) (8–12), which is enriched in actively transcribed regions of the genome (8, 13, 14).

The Mixed Lineage Leukemia-1 (MLL1) methyltransferase is one of six human H3K4 methyltransferases that belong to the SET1 family (15). The MLL1 Set catalytic domain alone monomethylates H3K4, while the enzyme can only dimethylate and trimethylate H3K4 when MLL1 is incorporated into the WRAD subcomplex, which also contains WDR5 (tryptophan-aspartate repeat protein-5), RbBP5 (retinoblastoma-binding protein-5), Ash2L (Absent-small-homeotic-2-like), and two copies of DPY-30 (Dumpy-30) (16, 17). Chromosomal translocations that result in MLL1 fused to a variety of partner proteins result in MLL-rearranged leukemia (18, 19), which has a particularly poor prognosis (20). MLL1 is therefore an attractive therapeutic target, making a better understanding of the molecular basis of MLL1 function in the context of the WRAD complex important for drug development (21, 22).

The Ash2L subunit plays a key role in orchestrating MLL1 function. Ash2L stimulates MLL1 methylation activity *in vitro* (23), and knockdown of Ash2L abolishes H3K4 trimethylation in cells (24). Knockout of Ash2L in mice results in embryonic lethality (25). Studies have shown that transcription factors recruit the WRAD complex by interacting with Ash2L, often in a tissue-specific manner. This includes the reported interaction of Ash2L with the transcription factor, Ap2 $\delta$ , which recruits the MLL1 complex to the HoxC8 promoter (26). Ash2L also interacts with Tbx1, a transcription factor that is important for heart development and is deleted in human DiGeorge syndrome (25). Mef2d, a transcriptional regulator that targets muscle genes in myoblasts, recruits Ash2L to facilitate terminal differentiation (27). Ash2L has also been shown to shape neocortex formation by transcriptional regulation of Wnt- $\beta$ -catenin

## Significance

The Mixed Lineage Leukemia-1 (MLL1) complex plays a role in activating transcription by methylating lysine 4 in histone H3, a reaction that is stimulated by the presence of ubiquitin conjugated to histone H2B. Recent structures of the core MLL1 complex, termed MLL1-WRAD, have revealed the existence of multiple docking states and have also left ambiguous portions of the structure. Here we combine mass spectrometry cross-linking with cryo-EM to model additional regions of the MLL1-WRAD complex and identify a series of states that shed light on complex assembly and the role that ubiquitin may play in orienting MLL1-WRAD on nucleosomes.

Author affiliations: <sup>a</sup>Department of Biophysics and Biophysical Chemistry, The Johns Hopkins University School of Medicine, Baltimore, MD 21205; <sup>b</sup>Department of Structural Biology, Van Andel Research Institute, Grand Rapids, MI 49503; and <sup>c</sup>Department of Biochemistry and Molecular Biology, State University of New York Upstate Medical University, Syracuse, NY 13210

Author contributions: N.A.H., B.A.K., M.S.C., and C.W. designed research; S.R., N.A.H., M.L.S., K.E.W.N., B.A.K., and M.S.C. performed research; M.S.C. contributed new reagents/analytic tools; S.R., E.J.W., and C.W. analyzed data; and S.R., N.A.H., E.J.W., B.A.K., and C.W. wrote the paper.

Reviewers: A.L., University of California San Diego; and N.T., Friedrich Miescher Institute for Biomedical Research.

Competing interest statement: C.W. is a member of the ThermoFisher Scientific Advisory Board.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>S.R. and N.A.H. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: cwolberg@jhmi.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2205691119/-/DCSupplemental>.

Published September 12, 2022.

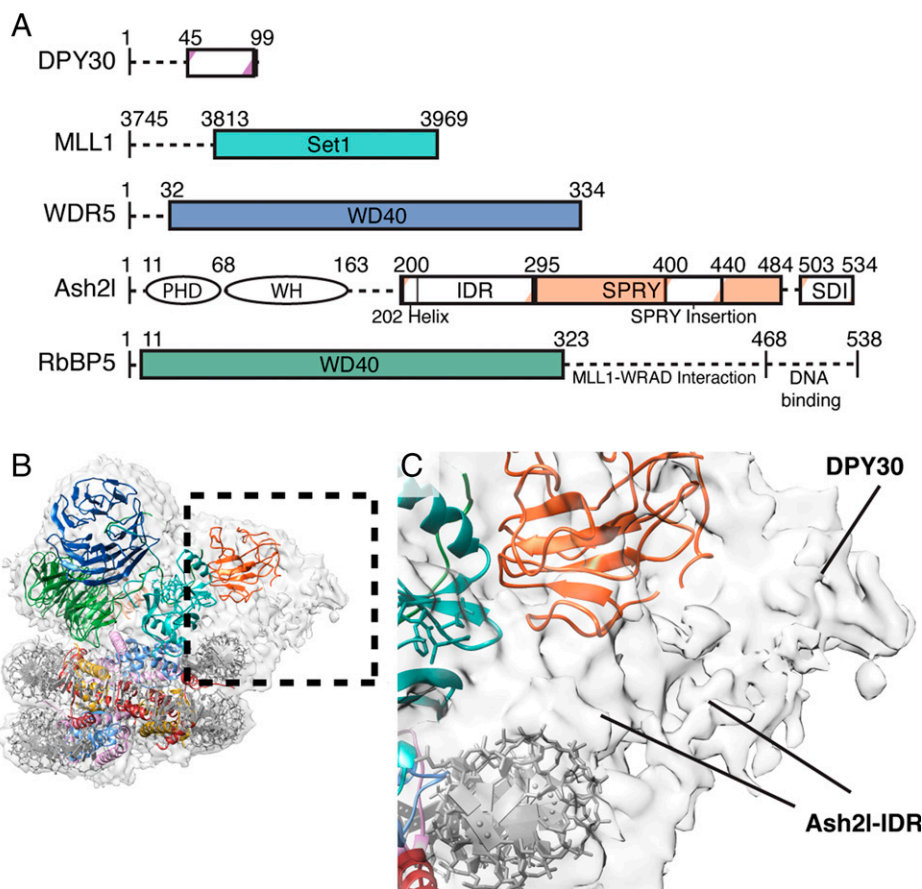
signaling (28) and plays a key role in maintaining open chromatin in embryonic stem cells (29).

Structural studies have provided insights into the role of Ash2L in MLL methyltransferase complexes. Ash2L is a 60-kDa protein harboring an N-terminal plant homeodomain (PHD) finger, winged helix (WH) DNA binding motif, a SPRY domain (residues 295 to 484) with a 40-residue insertion termed the SPRY insertion, and a helical Sdc1-DPY30 interacting region (SDI motif) (Fig. 1A). There is an intrinsically disordered region (IDR, residues 200 to 295) located between the WH and SPRY domains. Structural and biochemical studies have elucidated the interaction of the Ash2L WH domain with DNA (30, 31). Crystal structures of the Ash2L SPRY domain revealed its beta-sandwich topology (32–34) and interaction with WRAD partners, RbBP5 (34, 35) and MLL1 (23, 36). The C-terminal SDI motif of Ash2L protrudes from the SPRY domain and binds to DPY30 (33, 37). A predicted flexible region spanning residues 200 to 295 [referred to as the IDR (38)] and containing residues important for DNA binding (38), as well as the SPRY insertion (400 to 440), were not resolved in any of these previously determined structures.

Cryo-electron microscopy (cryo-EM) structures of the MLL1-WRAD complex bound to nucleosomes (38, 39) have revealed the overall topology of the complex and its association with the nucleosome. Interestingly, on nucleosomes that lack H2B-Ub, MLL1-WRAD adopts a tilted conformation facing the nucleosome dyad and shows no contacts between MLL1 and the nucleosome (38). In this state, low-resolution density of Ash2L allowed placement of the SPRY domain, while homology modeling based

on the structure of Bre2, the yeast homolog of Ash2L, was used to model the IDR and SPRY insertions. On ubiquitinated nucleosomes, the position of MLL1-WRAD is rotated over the nucleosome disk and adopts what is thought to be an “active state,” with MLL1 contacting the H2A–H2B acidic patch (39). While the Ash2L SPRY domain was placed in density in both structures, the IDR and SPRY insertions are poorly defined. The DPY30 dimer was not placed in the previous active state structure, due to poorly resolved density. In the putative active state structure (39), the authors interpreted weak density between DNA and the SPRY domain as corresponding to an antiparallel beta sheet of the IDR and the SPRY insertion based on homology to Bre2, the *Kluyveromyces lactis* (*K. lactis*) homolog of Ash2L (40). In the nonactive state structure (38), the authors modeled the IDR based on Bre2 complex.

We report here an integrative approach combining cryo-EM and mass spectrometry (MS) cross-linking data to model the Ash2L IDR region and SDI motif in the context of the MLL1 complex bound to nucleosomes containing monoubiquitinated H2B-K120. Our maps reveal an active state similar to that reported by Xue et al. (39), but with improved density for the region corresponding to the Ash2L IDR and SPRY insertions, as well as DPY30. Using MS cross-linking and molecular dynamics flexible fitting (MDFF) (41), we generated a model of the Ash2L IDR and SPRY insertion. We also captured three additional distinct states of the MLL1-WRAD complex bound to ubiquitinated nucleosomes. Together, our results provide insights into MLL1 structure and assembly, as well as the role of ubiquitin in promoting formation of the active state.



**Fig. 1.** Map of MLL1-WRAD complex shows additional density unexplained by previous structures. (A) Bar diagram depicting the MLL1-WRAD complex. Circled (not visible in map and PDB) and boxed areas represent structured domains, and white-striped boxes show newly added domains in this work. (B) Cryo-EM map of MLL1-WRAD calculated at 6-Å resolution showing fit of previously reported structure, PDB 6KIU, superimposed in density. Unaccounted-for density is shown in dashed box. (C) Zoom on unaccounted-for density indicating potential subunits that could account for it.

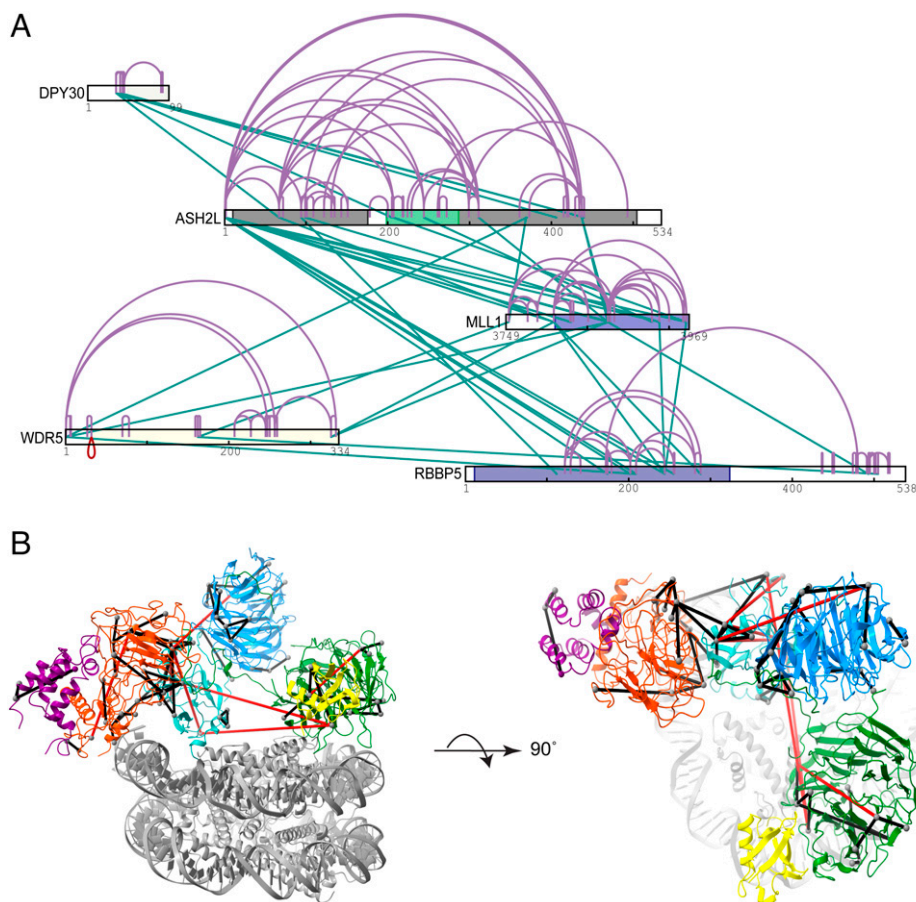
## Results

**Structure Determination of Multiple MLL1-WRAD States.** We determined the structure of the human MLL1-WRAD complex (17) bound to a ubiquitinated nucleosome using single-particle cryo-EM. The *Xenopus laevis* nucleosome contained ubiquitin linked to K120 of histone H2B via a nonhydrolyzable dichloroacetone (DCA) linkage (see *Materials and Methods* and ref. 42). In addition, K4 of histone H3 was substituted with norleucine, a nonnatural amino acid which has been shown to bind more tightly to SET methyltransferases than the native lysine (43). Replacing H3K4 with norleucine was previously shown to drive tighter binding of the yeast homolog of MLL1-WRAD, COMPASS, to nucleosomes (44). The minimal fully active human MLL1 fragment containing the catalytic SET domain and spanning residues 3745 to 3969 was combined with full-length RbBP5, Ash2L, DPY30, and WDR5 as previously reported (17) to reconstitute MLL1-WRAD. The complex of MLL1-WRAD bound to ubiquitinated nucleosomes was stabilized by cross-linking with glutaraldehyde (see *Materials and Methods*).

The MLL1-WRAD complex was imaged with a Titan Krios equipped with a Gatan K3 direct electron detector (*SI Appendix, Table S1*). Our approach (see *Materials and Methods*) enabled us to resolve four distinct MLL1-WRAD states bound to the nucleosome at an overall resolution (Fourier Shell Correlation [FSC] 0.143 criterion) of 3.4 Å for state 1, 3.6 Å for state 2, 4.7 Å for state 3, and 4.3 Å for state 4 (*SI Appendix, Fig. S1*). The local resolutions of the MLL complexes are lower than that of the nucleosome in all four states, with the

nucleosome at a  $\sim 3.5$ - to  $4.5$ -Å resolution, whereas MLL1-WRAD was resolved to a  $\sim 5$ - to  $9.5$ -Å resolution (*SI Appendix, Fig. S2*). Subunits from previously reported human MLL1-WRAD structures (38, 39) were placed in each map. State 4 fits well with a previously reported MLL1-WRAD structure (Protein Data Bank [PDB] ID 6KIU) (39), but also contained additional density, at a resolution of 7 Å to 9.5 Å, adjacent to the Ash2L SPRY domain that was not accounted for by the 6KIU model (Fig. 1 *B* and *C*). We hypothesized that this density might correspond to the DPY30 dimer and the more dynamic IDR and SPRY insertion regions of Ash2L (Fig. 1*A*), which were not resolved in previous reports (38, 39). The three additional MLL1-WRAD states identified in our study (*SI Appendix, Fig. S1*) and described in further detail below have not been previously reported.

**Cross-Linking MS on the MLL1-WRAD Complex.** To provide an experimental basis for molecular modeling of the extra density in state 4, we used cross-linking MS on the MLL1-WRAD complex to identify close contacts that could be used as restraints in model building. We cross-linked the MLL1-WRAD complex with BS3, a homobifunctional primary amine reactive cross-linker that has a theoretical maximum cross-linking distance constraint of 34 Å between C $\alpha$  atoms of cross-linked lysine residues (45). Each MLL1-WRAD subunit contains seven or more lysine residues well distributed throughout the polypeptide, with one or more lysines found in each domain. Cross-linked MLL1-WRAD was digested with trypsin and analyzed by MS, and the resulting spectra were searched with the Plink2 cross-link search algorithm (46).



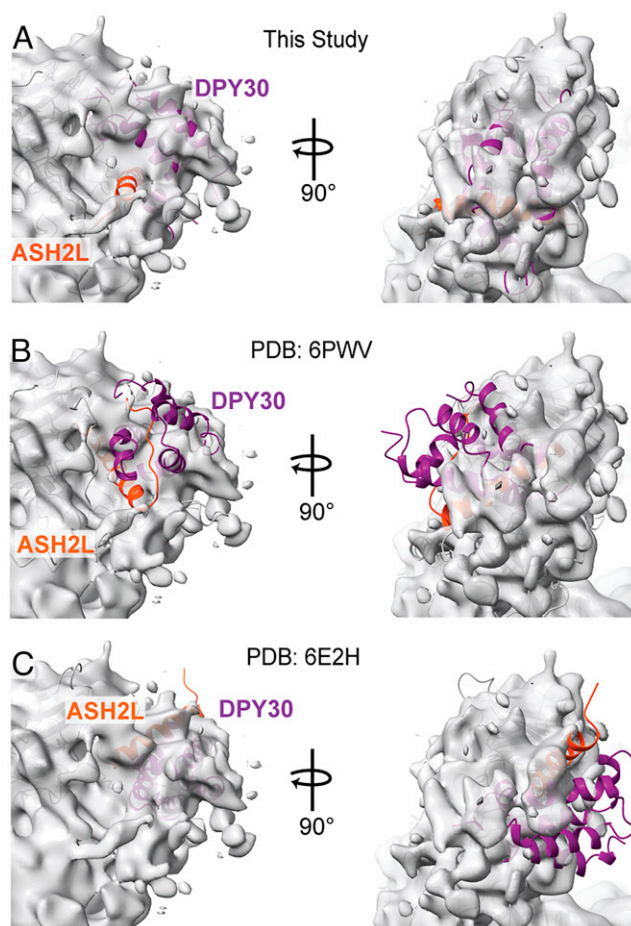
**Fig. 2.** Cross-links of MLL1-WRAD in solution. (A) Intraprotein cross-links (purple) and interprotein cross-links (green) are indicated by lines. (B) Visualization of cross-linked lysines on the MLL1-WRAD–nucleosome complex. Satisfied (black) and violated (red) cross-links were determined using BS3's theoretical C $\alpha$ –C $\alpha$  maximum crosslinking distance of 34 Å.

The final dataset comprised 167 lysine–lysine cross-links at a 2% false discovery rate (FDR) cutoff, with 40 interprotein cross-links and 127 intraprotein cross-links (Fig. 2). The diverse network of interprotein cross-links reflects the intertwined subunit architecture of the MLL1 complex. An exception is DPY30, which only has four cross-links to four residues in Ash2L, consistent with the position of DPY30 at the periphery of the MLL1-WRAD complex. Ash2L is the subunit with the largest number of cross-links, with 75 interprotein and intraprotein cross-links (Fig. 2 and *SI Appendix*, Fig. S3). There were 26 intersubunit cross-links for the MLL1-SET domain, reflecting this protein's central position in the MLL1-WRAD complex. By contrast, there were far fewer cross-links to WDR5 and RbBP5, which are larger proteins and contain a greater number of lysines compared to the MLL1-SET domain. The small number of cross-links to these subunits could be due to the relative inaccessibility of lysines in the WD40 folds of both proteins, which reduces cross-linking efficiency (47). Alternatively, it is possible that there is greater mobility in these regions, which could similarly reduce the number of observed cross-links.

### The DPY30-SDI Motif Occupies a Peripheral Density Stretch.

To fit missing portions of the model to the extra density in our maps, we first utilized two DPY30 crystal structures (PDB 4RIQ and 6E2H) (33, 37) and fit them to the density map using the cross-linking data as a guide (see *Materials and Methods*). Both crystal structures contain DPY30 bound to the Ash2L C-terminal SDI motif, which protrudes from the Ash2L SPRY domain (33). This portion of human WRAD was either absent or poorly defined in previous structures of the MLL1 complex bound to nucleosomes (38, 39). By placing the SPRY domain of crystal structure 6E2H (33) in our map, the Ash2L SDI motif (residues 503 to 523) and DPY30 from the crystal structure were near the unallocated density (Fig. 3C). A simple rotation of the SDI-DPY30 portion of PDB 6E2H relative to its SPRY domain positions these domains in the density (Fig. 3 and *SI Appendix*, Fig. S4B). The 15-residue linker (residues 485 to 500) connecting the Ash2L SDI-DPY30 region and the Ash2L SPRY domain was manually rebuilt and then refined using MDFF (see *Materials and Methods*). The resulting position of DPY30 and the Ash2L SDI motif fits well into the active state map and satisfies a cross-link (SPRY-K299 to SDI-Linker-K493) (Fig. 2B), consistent with the conformational rearrangement modeled here.

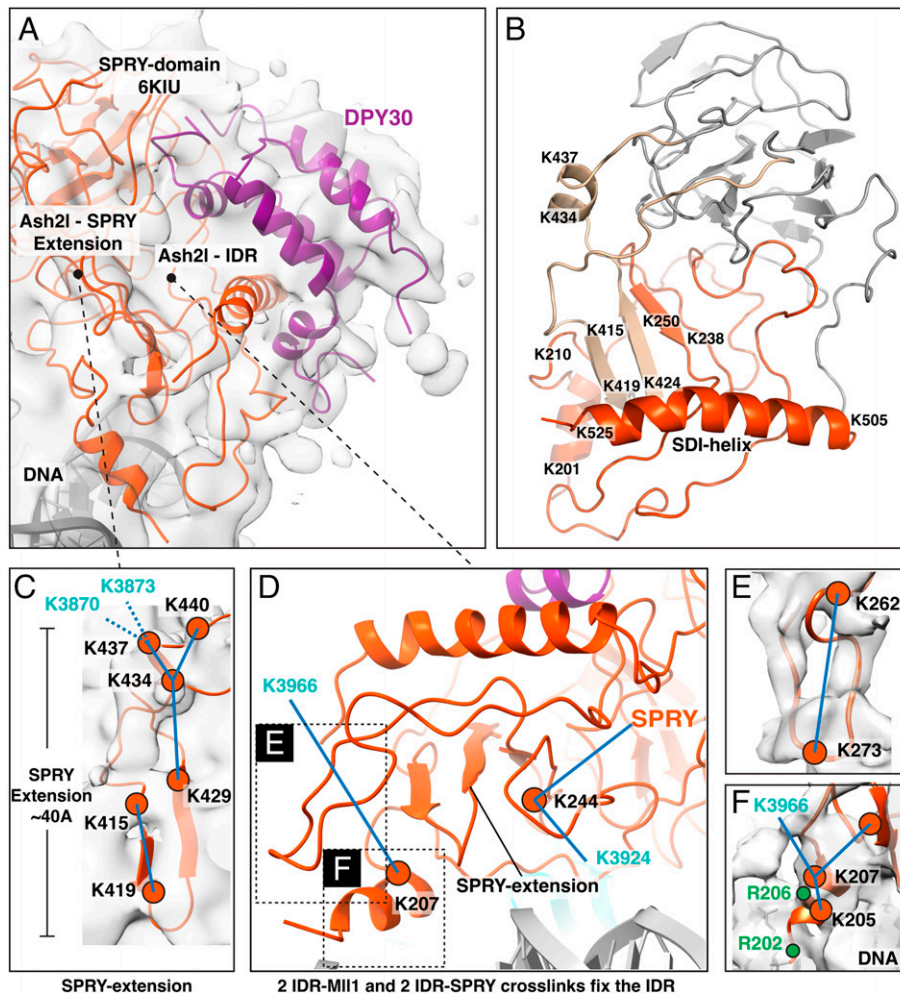
The ability of the Ash2L linker and the SDI motif protruding from the SPRY domain to adopt multiple conformations had been observed in previous structures (33, 37) (Fig. 3 and *SI Appendix*, Fig. S4B). In our model, the partially bent SDI motif most closely resembles the helix conformation observed in the crystal structure of DPY30 alone bound to the SDI motif (4RIQ) (37) but differs from that previously reported for the cryo-EM structure of MLL1 bound to a nucleosome, 6PWV (Fig. 3B and *SI Appendix*, Fig. S4B). We note that orientation of the SDI motif in our structure is dictated by the positioning of the DPY30 dimer, which is bound to the SDI motif. Whereas the reported position of the SDI motif in 6PWV could fit in the density of our active state map (Fig. 3B and *SI Appendix*, Fig. S4B), the resulting position of the bound DPY30 dimer does not fit the density in our map (Fig. 3B and *SI Appendix*, Fig. S4C). The position of DPY30 in our model is consistent with observed cross-links between DPY30 residues K35 and K40, which are located in a disordered region, and surface lysine K406 and K440 of the Ash2L SPRY domain (*Dataset S1*).



**Fig. 3.** Comparison of Ash2L-SDI motif (orange) and DPY30 dimer (purple) positioning on active state map. All structures were compared by aligning the Ash2L-SPRY domain. (A) Structure of the Ash2L-SDI motif and DPY30 dimer presented in this study. (B) Structure of the Ash2L-SDI motif and DPY30 dimer reported by Park et al. (38) (6PWV) positioned in our active state map. (C) Crystal structure of the Ash2L-SDI motif and DPY30 dimer reported by Haddad et al. (33) (6E2H) positioned in our active state map.

**Integrative Modeling of Ash2L.** Placement of DPY30 and the Ash2L SDI motif left an unassigned density of about 7- to 9.5-Å resolution adjacent to Ash2L that could potentially correspond to the IDR region (residues 200 to 300) and the SPRY insertion (residues 400 to 440) (Fig. 1A). There is no high-resolution structural information on the Ash2L IDR, and previously proposed models of this region based on the Ash2L yeast homolog, Bre2 (38), do not fit our EM map well, nor does a recently reported model generated by AlphaFold2 (48). The 40-residue insertion in the Ash2L SPRY domain has been shown to contact DNA (38, 39), and a homologous stretch of yeast Bre2 similarly extends toward nucleosomal DNA in COMPASS (49, 50). We used these observations, combined with cross-linking restraints and MDFF, to model the IDR and SPRY insertion domain in our map (Fig. 4B).

The resulting model of Ash2L places the 40-residue SPRY insertion loop in the EM density at the periphery of the SPRY domain (Fig. 4C). The loop extends over 40 Å on the periphery of the SPRY domain to the nucleosomal DNA and contains an antiparallel beta sheet comprising residues 415 to 419 and 424 to 428 (Fig. 4C), whose homologous residues also form antiparallel beta strands in *K. lactis* Bre2 (51). This positions K419 and K421 near the DNA, consistent with the deleterious effect of substitution of these residues on DPY30-dependent methylation by MLL1 (52). The fit satisfies eight cross-links, including



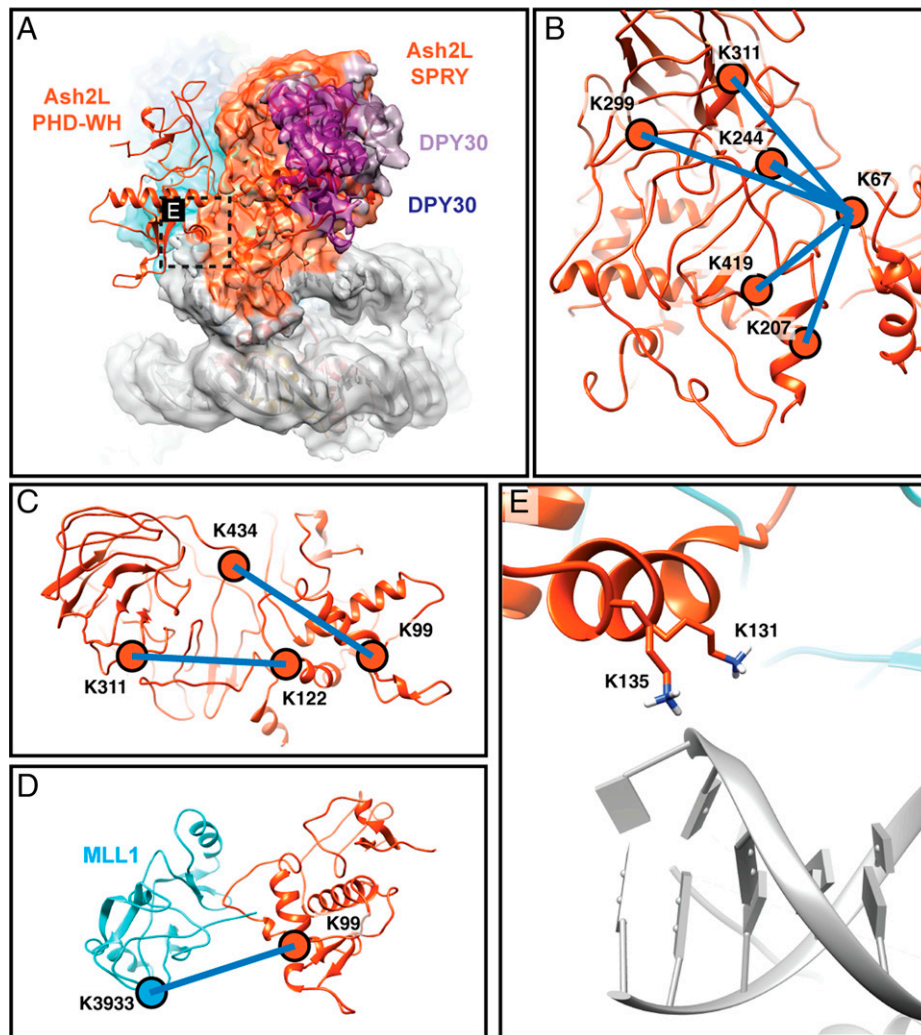
**Fig. 4.** Building and fitting of the Ash2L IDR-SPRY domains to experimental density. (A) Placement of Ash2L<sub>IDR-SPRY</sub> and DPY30 model into the active state map. (B) Overview of the Ash2L<sub>IDR-SPRY</sub> model built by integrative modeling and MDFF refinement. The IDR is colored orange, and the SPRY insertion region is colored gold. Regions colored gray correspond to the SPRY domain, which was built using 6KIU as a reference model. (C) SPRY insertion stretching over 40 Å fitted to EM density. Obtained cross-links are displayed (blue, satisfied). Residues labeled in cyan refer to satisfied cross-links to the MLL1 SET domain. (D) IDR region located in the vicinity of the SPRY insertion. Intra-SPRY cross-links from K207 and K244 are displayed and demonstrate the anchoring of the IDR region. (E) Zoom on indicated region in D showing IDR loop 254 to 270 and satisfied cross-link. (F) Zoom on indicated region in D showing IDR helix 200 to 210; all cross-links are shown, lysine residues are depicted in orange, and arginine residues are depicted in green.

two intersubunit cross-links between Ash2L residue K437 and MLL1 residues K3870 and 3873 (Fig. 4C). As a next step, we placed IDR residues K207 and K244 in the unassigned EM density based on cross-links between these residues with multiple lysine residues within Ash2L, as well as to MLL1 (Fig. 4D). Specifically, Ash2L K207 cross-links to two adjacent IDR lysines (K205 and K218) and to MLL1 K3966 (Fig. 4D and F), while Ash2L K244 cross-links to the SPRY-domain K311 and to MLL1 residue K3924 (Fig. 4D). Next, we included secondary structure based on Bre2 homology that predicts a beta strand comprising residues 247 to 250 that hydrogen bonds with the 424 to 428 strand in the SPRY insertion, forming a three-stranded beta sheet (Fig. 4B). Residues 241 to 246 were modeled as a loop and strand preceding the 247 to 250 beta strand based on sequence similarity to Bre2 (51). Based on secondary structure prediction (*SI Appendix, Fig. S5A*), we modeled Ash2L residues 200 to 210 as an  $\alpha$ -helix and placed it in corresponding density (Fig. 4F). Notably, the proximity of the helix to DNA (Fig. 4A) is consistent with the observation that alanine substitutions of K205/K206/K207 markedly reduce MLL1 methyltransferase activity on nucleosomes (38). Finally, we fit Ash2L residues 253 to 278 to a short stretch of unassigned density extending to the periphery of the

complex (Fig. 4D and E), and in a position that satisfies the observed cross-link between K262 and K273 (Fig. 4E and *SI Appendix, Fig. S4B*). Given the limited resolution of the maps in this region ( $\sim 7$ - to  $9.5$ -Å resolution), it is important to note that sidechains are included in the modeled regions solely for model completeness and cross-link validation, but are not experimentally determined based on map density.

The proximity of the Ash2L IDR and SPRY insertion region to nucleosomal DNA in our model (Fig. 4A) could account for the importance of DPY30 binding to MLL1-WRAD activity. Previous studies have shown that DPY30 greatly increases MLL1 methyltransferase activity on nucleosomes and that this increase in activity depends upon the Ash2L IDR (52). DPY30 forms multiple contacts with the IDR and the SDI helix, which help to stabilize folding of the IDR, as supported by NMR studies (52). This ordering of the IDR, in turn, would favor the predicted interactions with nucleosomal DNA, thus accounting for the importance of these residues to MLL1 activity on nucleosomes (38).

There was no apparent density corresponding to the Ash2L N-terminal PHD finger (residues 11 to 67), which has been shown to bind DNA (30, 31), or the atypical WH domain (residues 68 to 163). Previous cryo-EM studies of MLL complexes also did not locate these domains (38, 39). We therefore



**Fig. 5.** Modeling of Ash2L PHD-WH domain. (A) Model of Ash2L PhD-WH ensemble on the active state model predicted by cross-linked restrained docking. (B) Cross-links between Ash2L PhD-WH residue K67 and the Ash2L IDR and SPRY domain. (C) Additional cross-links between K99 and K122 in the Ash2L PhD-WH with K311 and K434 in the SPRY domain. (D) Single cross-link connecting Ash2L PhD-WH residue K99 to MLL1 SET domain residue K3933. (E) Zoom on indicated region in A showing the positioning of the K131 and K135 in close proximity to the backbone of the nucleosomal DNA.

utilized the cross-links we identified involving the Ash2L N-terminal WH and PHD domains as restraints to guide docking of the crystal structure of these domains (PDB ID 3RSN) (30) to our active state model (Fig. 5A). Of the 22 cross-links we obtained (Dataset S1), the 12 cross-links between lysines within the WH and PHD finger domains were satisfied by the crystal structure. The additional 10 cross-links were used as restraints to dock the crystal structure on our MLL1-WRAD active state model. Eight of the cross-links connected to other regions of Ash2L (Fig. 5B and C), and two cross-links connected to MLL1 (Fig. 5D) and RbBP5 (SI Appendix, Fig. S3B). In the highest-scoring model, all cross-links are satisfied except for one connecting K94 of the Ash2L WH and PHD finger domains to K112 of RbBP5 (SI Appendix, Fig. S3B), and the top four hits position the WH and PHD finger comparably, with a backbone RMSD of  $0.336 \pm 0.045$  Å. The modeled position of the Ash2L WH and PHD finger domains (Fig. 5A) lies in a groove between MLL1 and the Ash2L SPRY domain. Importantly, the model places two lysine residues, K131 and K135, in contact with the sugar-phosphate backbone of the nucleosomal DNA (Fig. 5E), consistent with their importance to the binding of Ash2L to DNA (30). The location of Ash2L at the periphery of the nucleosome could facilitate interactions

with transcription factors known to be recruited by this subunit (25). Given the positioning of the WH at the end of the DNA in the nucleosome core particle, it is possible that this domain could make additional contacts with DNA that extends beyond the minimal fragment associated with the nucleosome.

#### Four MLL1-WRAD States Represent Putative Assembly Intermediates.

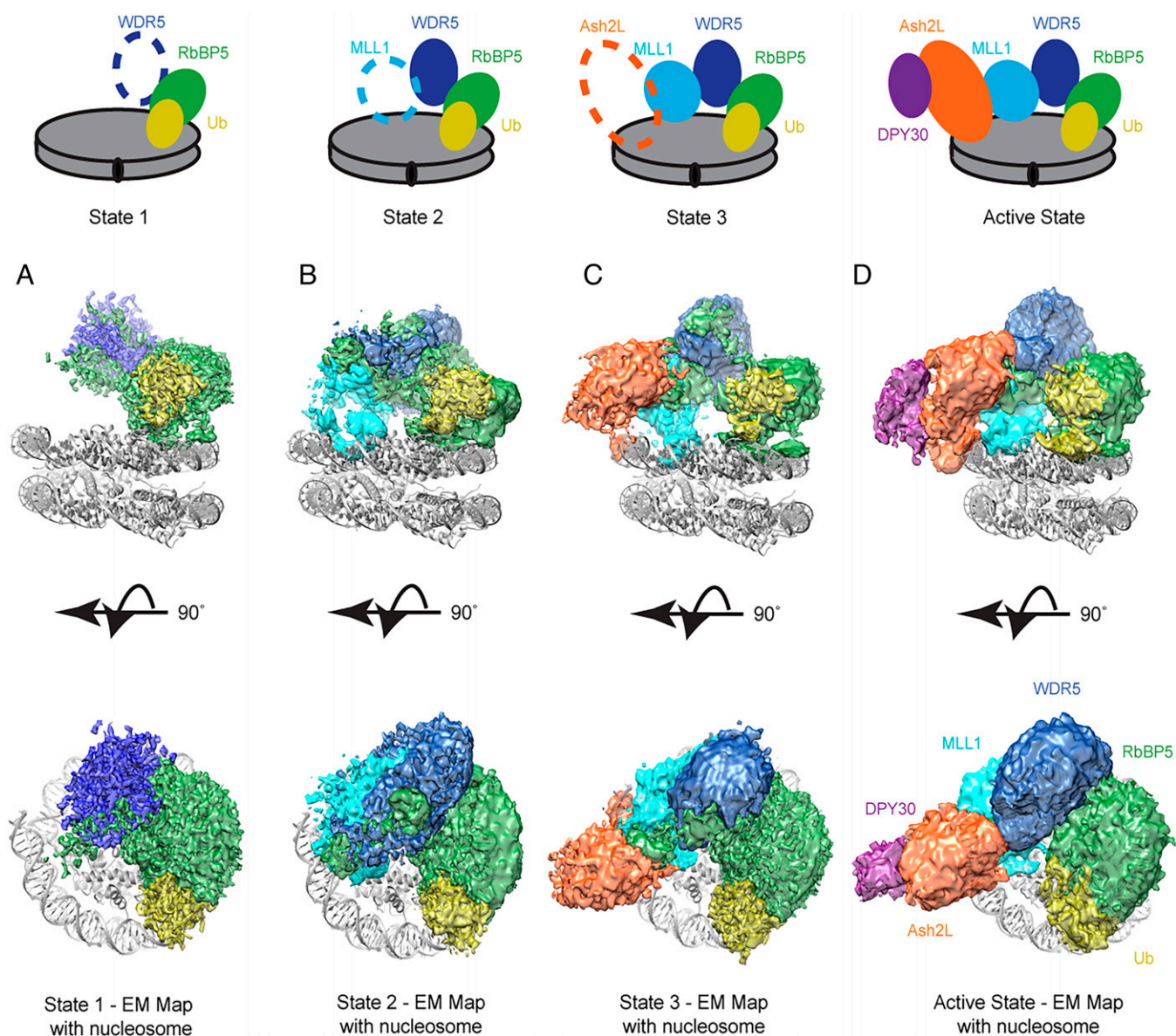
In addition to the active state MLL1-WRAD complex described above, we were able to resolve three additional states in our dataset (Fig. 6 and SI Appendix, Fig. S1 and Table S3). All four states have well-resolved density for RbBP5-WD40 (~5- to 6-Å resolution), whereas other MLL1-WRAD subunits are less well resolved but seem to be in similar positions as compared to the active state model. In state 1 (Fig. 6A), only the RbBP5-WD40 domain shows strong density on the nucleosome. There is no density corresponding to the remainder of RbBP5 or to the methyltransferase subunit, MLL1. In state 2 (Fig. 6B), WDR5 is also well resolved (~6- to 7-Å resolution), in addition to the strong RbBP5-WD40 density, while the remaining MLL1-WRAD density is poorly resolved. State 2 contains additional strong density contiguous with the DNA at SHL -6 to SHL -7 (Fig. 6B) that may correspond to a slight peeling away of the DNA from the histone core, although the density is not

sufficiently well resolved to rule out alternative models that place additional protein in this location. State 3 contains equally well resolved density for RbBP5-WD40 and WDR5 as well as better-resolved MLL1-SET domain density ( $\sim 7$ - to  $8$ -Å resolution) and low-resolution Ash2L density ( $\sim 7$ - to  $9.5$ -Å resolution) (Fig. 6C). Interestingly, in this state, the DNA density at SHL  $-6$  to SHL  $-7$  is very weak, as is the MLL1 density as it approaches the nucleosome disk.

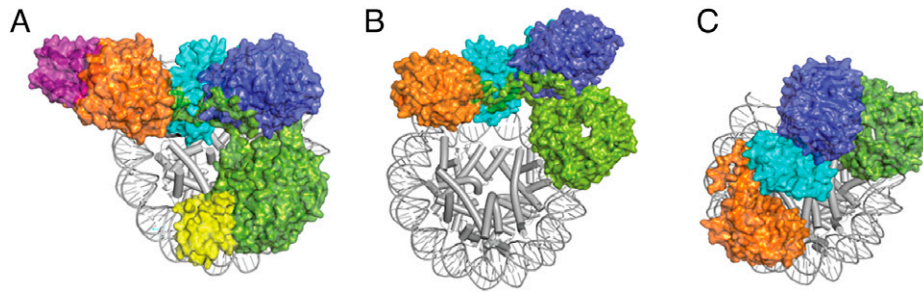
A comparison of the four states reveals little change in the docking of the MLL1-WRAD complex on the nucleosome, except for state 2, where the density indicates that WDR5 is somewhat tilted toward RbBP5 as compared to states 3 and 4 (SI Appendix, Fig. S6). This altered position could be a result of the absence or high mobility of the MLL1-SET domain, although further experiments would be needed to validate this hypothesis. Overall, we speculate that the four resolved states may represent an apparent progressive assembly of subunits from state 1, which contains RbBP5 and partial density for WDR5, through state 4 (the active state), which contains all

five subunits (Fig. 6D). While it is not possible to distinguish ordered assembly from simple heterogeneity of states bound to the nucleosome, one possibility is that RbBP5 binding to the ubiquitin conjugated to histone H2B-K120 facilitates stepwise assembly of the full MLL1-WRAD complex on nucleosomes. While the RbBP5-WD40 domain contacts ubiquitin, the flexible C-terminal loops of RbBP5 (residues 330 to 475) contact all MLL1-WRAD subunits except DPY30. We speculate that the RbBP5 C-terminal residues could play a role in the assembly of MLL1's active state on the nucleosome. Consistent with the model for stepwise assembly of the complex, a hydrodynamic analysis of the MLL1-WRAD complex identified multiple assembly states in the absence of nucleosomes, which included several different substates in addition to the full five-component complex (53).

**Role of Ubiquitin and Ash2L/DPY30 in Positioning the MLL1 Complex on Nucleosomes.** A striking feature of reported structures of MLL1 complexes (38, 39, 52, 54) is their highly variable



**Fig. 6.** Four distinct states of MLL1-WRAD on the ubiquitinated nucleosome. (Top) The cartoon depictions indicate the four states discussed in the text. (Middle and Bottom) Two views of the corresponding density maps are shown beneath each cartoon, with the structure of the active state model superimposed on each map. (A) State 1, (B) state 2, (C) state 3, and (D) active state map calculated at 6 Å, shown for comparison. Densities in all panels are colored according to the cartoon diagrams.



**Fig. 7.** Position of MLL1 complexes on the nucleosomes. The solvent-accessible surface is shown for each resolved subunit. Color scheme for all panels: ubiquitin, yellow; RbBP5, green; WDR5, blue; MLL1, cyan; Ash2L, orange; and DPY30, purple. (A) MLL1-WRAD complex bound to H2B-ubiquitinated nucleosome, present study (7UD5). (B) MLL1-WRAD bound to unmodified nucleosomes (6KIZ). (C) MLL1-WRA bound to unmodified nucleosome (6W5M).

positioning on the nucleosome, in contrast with the conserved spatial relationship among the five subunits of the MLL1-WRAD complex. These orientations of the complexes on the nucleosome can be categorized according to the superhelical positions of RbBP5 and Ash2L/DPY30, which define the two ends of the complex. The positions of these subunits in the present and in previous studies are listed in *SI Appendix, Table S4*, and a subset are depicted in Fig. 7.

A comparison of structures suggests a role for the ubiquitin conjugated to H2B-K120 in positioning the MLL1-WRAD complex on the nucleosome. Monoubiquitination of H2B has been shown to increase the catalytic efficiency of MLL1-WRAD on nucleosomes by twofold *in vitro* (39). The active state MLL1-WRAD complex reported here (Fig. 1*B* and 6*D*) is docked on the H2B-ubiquitinated nucleosome in essentially the same orientation as a previously reported structure of MLL1-WRAD bound to H2B-ubiquitinated nucleosome [6KIV (39)], with RbBP5 at SHL  $-2.5$  and Ash2L at SHL  $-7$  (Fig. 7*A*). The MLL3-WRAD complex (6KIW), which contains a related methyltransferase subunit and the same four adapter proteins, is oriented in a similar manner on H2B-ubiquitinated nucleosome (39). By contrast, structures of MLL1 complexes bound to unmodified nucleosomes (38, 39, 54) show highly variable positioning of the MLL1 complex. Two structures of MLL1-WRAD bound to unmodified nucleosomes, 6PWV (38) and 6KIZ (39), have the Ash2L subunit positioned around SHL  $-7$ , as in the active complex, but with MLL1-WRAD markedly pivoted by about  $35^\circ$  (Fig. 7*A* and *B*). This rotation places RbBP5 in contact with nucleosomal DNA at SHL  $-1.5$ , in addition to reducing the contact area with the nucleosome core. A subset of these particles adopted the active orientation (6KIW), which suggested that MLL1-WRAD could sample both positions in the absence of ubiquitin (39). A similar heterogeneity in positions resembling the active and peripheral complexes was also observed in a recent study of MLL1-WRAD bound to unmodified nucleosomes [7MBM and 7MBN (54)]. Taken together, these observations suggest that the interaction with ubiquitin favors the docking arrangement of the active complex (Fig. 7*A*).

The DPY30 subunit may also play a role in positioning MLL1 complexes on nucleosomes lacking H2B-K120Ub. Structural studies of a four-protein MLL1-WRA complex lacking both DPY30 and the Ash2L SDI helix (6W5I, 6W5M, and 6W5N) (52) showed the complex to bind unmodified nucleosomes in multiple orientations (*SI Appendix, Table S4*). While one class (6W5N) resembles that of the five-protein MLL1-WRAD complex bound to unmodified nucleosomes (6PWV), with Ash2L at SHL  $-7$ , the remaining classes bind in another orientation, with Ash2L at SHL  $-4.5$  (6W5I and 6W5M) (Fig. 7*C*). This observation is consistent with a role for DPY30 in favoring binding between SHL  $-6$  and SHL  $-7$ , as

observed in structures of the complete MLL1-WRAD complex containing DPY30 (6KIV, 6KIX, 6KIZ, 6PWV, and this work). DPY30 likely exerts its effect through its interactions with the Ash2L IDR, which undergoes conformational changes in the presence of DPY30, as reflected in changes in NMR spectra (52). The proximity of the Ash2L IDR and SPRY insertion region to nucleosomal DNA in our model (Fig. 5), as well as in a previously proposed model (38), could account for the coupling of DPY30 binding to ordering of this region of Ash2L.

## Discussion

Structural studies of the MLL1-WRAD complex have revealed surprising heterogeneity in the positioning of the complex on nucleosomes. Since the substrate lysine, K4, lies in the flexible tail of histone H3 and can therefore access the active site of MLL1 in all reported structures, there has been some question as to which one represents the active conformation (38, 54). The structure of the MLL1-WRAD complex bound to an H2B-ubiquitinated nucleosome reported by Xue et al. (39) most closely resembles the docking of yeast homolog, COMPASS, on nucleosomes (49, 51) and was, in part, for this reason, referred to as the active complex. However, in contrast with the intrinsic variability of MLL1-WRAD positioning, structures of COMPASS revealed a stable nucleosome-bound complex with no differences in docking to ubiquitinated versus unmodified nucleosomes (49, 50). Unlike COMPASS, structures of MLL1-WRAD bound to unmodified nucleosomes have revealed multiple docking arrangements (38, 39, 52, 54) (*SI Appendix, Table S4*). The structure of the MLL1-WRAD complex we report in this work, which contains an H2B-ubiquitinated nucleosome, also contains a docking arrangement that is very similar to that observed for the only previous MLL1 complex structure that also contains a ubiquitinated nucleosome (39). The same docking was also observed for the MLL3-WRAD complex bound to ubiquitinated nucleosome (39). Taken together, these observations strongly suggest that the presence of ubiquitinated H2B-K120 orients the MLL1-WRAD complex on the nucleosome and that the resulting docking represents the active conformation. This effect is mediated through contacts between ubiquitin and RbBP5, which positions this subunit at SHL  $-2.5$  (*SI Appendix, Table S4*). In the absence of ubiquitin, the full MLL1-WRAD complex binds in a markedly different orientation that places RbBP5 at SHL  $-1.5$  and had reduced contacts with the face of the histone octamer (38). Since H2B ubiquitination increases methyltransferase activity *in vitro* by about twofold (39), it is likely that the docking observed in the absence of ubiquitination simply represents a lower activity state, rather than an inactive state.



H2B ubiquitination could enhance activity by favoring docking of MLL1 in the more active conformation.

The role of H2B-K120Ub in positioning MLL1 in an active docking orientation most closely resembles the mechanism by which H2B-K120Ub activates the histone H3K79 methyltransferase, Dot1L. Whereas Dot1L also forms higher-order complexes with partner proteins such as AF4 and AF9 (55), the catalytic domain alone is strongly stimulated by the presence of monoubiquitin conjugated to histone H2B-K120 (56). Like MLL1-WRAD, H2B ubiquitin orients Dot1L on the nucleosome via contacts with ubiquitin (44, 57, 58), but, in the absence of ubiquitin, Dot1L can adopt other, inactive docking arrangements (58, 59). Since the H3K79 sidechain is in the globular core of the nucleosome and not, like H3K4, in a flexible tail, correct positioning of Dot1L is even more critical to its activity. In the case of Dot1L, however, ubiquitin positions the C-terminal portion of the catalytic domain but still allows Dot1L to pivot between a poised orientation, in which the active site is ~20 Å from the substrate lysine, and an active position, in which Dot1L induces a conformational change in histone H3 that enables the substrate lysine to enter the active site (44). It is not known what role, if any, Dot1L partner proteins may play in further favoring the full active conformation on ubiquitinated nucleosomes.

Our study also points to a role for Ash2L in defining a second docking point for the MLL1-WRAD complex on the nucleosome. By combining cryo-EM data with MS cross-linking and MDFF refinement, we have been able to propose a model for the Ash2L IDR (Fig. 4) that differs from a previous model in which the MLL1 complex is positioned on the nucleosome in the low-activity conformation (38). We also propose a model for docking of the Ash2L WH-PHD domain based on cross-linking data (Fig. 5). Since our model of Ash2L predicts contacts with the DNA only, the similar relative position of Ash2L in both the active and inactive state suggests that there might be additional contacts with the histone core that further orient Ash2L.

We have identified three additional substates of the MLL1-WRAD complex that share the same overall positioning as the full active state but lack one or more subunits (Fig. 6). An interesting question is whether the four states represent assembly intermediates of the active, methylating complex. Biochemical evidence suggests that H3K4 methylation requires association of the MLL1-SET domain with the nucleosome octamer (60, 61), even though H3K4 lies in the flexible N-terminal tail of histone H3. The position of the MLL1-SET domain on the histone octamer surface in our active state structure agrees well with the histone residues that were identified as important to H3K4 methylation by the yeast MLL1 complex homolog, COMPASS (60). We speculate that the three less complete MLL1-WRAD states on the nucleosome, states 1, 2, and 3, may represent intermediates in assembly of MLL1-WRAD on the nucleosome. The RbBP5-WD40 domain present in all four states could be an anchor for the complex that then allows further stabilization of the complex on the nucleosome. An alternative possibility is that the remaining subunits of the MLL-1 complex are flexibly tethered, and that the different states observed represent stepwise tethering of the full complex on the nucleosome surface. We note that we cannot rule out the possibility that the intermediate states observed are due to the stripping of the complex at the air-water interface, leading to the presence of partially assembled complexes. Although partial MLL1 complexes have also been observed by Park (38) on recombinant nucleosomes, the significance and roles of the observed alternative states will require further study.

## Materials and Methods

**Expression and Purification of the MLL1 Complex.** The MLL1-WRAD complex was reconstituted from proteins expressed in *Escherichia coli* (17, 62).

A human MLL1 construct comprising residues 3745 to 3969, as well as full-length human WDR5, RbBP5, and ASH2L proteins, were individually expressed in *E. coli* (Rosetta II, Novagen) and purified as described previously (62). DPY30 was expressed and reconstituted with the other four proteins as described in ref. 17.

**Preparation of Ubiquitinated Nucleosomes Containing Norleucine.** Nucleosome core particles containing histone H3 with norleucine in place of lysine 4 as well as H2B ubiquitinated at K120 via a nonhydrolyzable DCA linkage were prepared as described (49). Briefly, histones H2A, H4, and the Widom 601 DNA were expressed and purified as described in ref. 63. Histone H3 containing norleucine in place of lysine 4 was expressed and purified as described in ref. 44. Ubiquitinated H2B was generated by expressing and purifying H2B K120C and ubiquitin K79C, which were cross-linked with DCA as described in ref. 42. A complete protocol for generating ubiquitinated nucleosomes can be found in ref. 64.

**Cross-Linking MS.** MLL1-WRAD complex was buffer exchanged into a cross-linking compatible buffer (50 mM Hepes pH 8.0, 200 mM KCl, 5 mM MgCl<sub>2</sub>, 0.1 mM (ethylenedinitrilo)tetraacetic acid, 5% glycerol, 1 mM tris(2-carboxyethyl)phosphine [TCEP]). Twenty micrograms of the buffer-exchanged complex was cross-linked with 2 mM and 4 mM BS3 for 4 h at 4 °C and frozen at -20 °C until further processing for liquid chromatography (LC)-MS/MS as described in refs. 65 and 66 with minor modifications. Briefly, cross-linked complexes were thawed at room temperature, and an equal volume of trifluoroethanol was added and incubated at 60 °C for 30 min to denature. The proteins were then reduced by the addition of 5 mM TCEP for 30 min at 37 °C followed by alkylation with iodoacetamide at a 10-mM final concentration for 30 min in the dark at room temperature. The sample was diluted 10-fold with 20 mM triethanolamine, and then digested with 2 µg of trypsin (Promega) overnight at 37 °C. The peptides were further purified on Sep-Pak C18 cartridge (Waters), dried, and resuspended in 5% acetonitrile/0.1% trifluoroacetic acid solution and then analyzed by LC-MS/MS. BS3-cross-linked peptides were analyzed on a Thermo Scientific Orbitrap Elite and Lumos with higher energy collision dissociation (HCD) fragmentation and serial MS events that included one FTMS1 event at 30,000 resolution followed by 10 FTMS2 events at 15,000 resolution. Other instrument settings included MS mass range greater than 1,800; *m/z* value as masses enabled; charge-state rejection: +1, +2, and unassigned charges; monoisotopic precursor selection enabled; dynamic exclusion enabled: repeat count 1, exclusion list size 500, exclusion duration 30 s; HCD normalized collision energy 35%, isolation width 3 Da, minimum signal count 5,000; and FTMS MSn AGC target 50,000. The RAW files were converted to mgf files and analyzed by the cross-link database-searching algorithm pLink2 under default settings: 1) up to three missed cleavages, 2) differential oxidation modification on methionine (+15.9949 Da), 3) differential modification on the peptide N-terminal glutamate residues (-18.0106 Da) or N-terminal glutamine residues (-17.0265 Da), 4) static modification on cysteines (+57.0215 Da), and 5) 2% FDR. All possible tryptic peptide pairs within 20 ppm of the precursor mass are used for cross-linked peptide searches. The cross-linked peptides were considered confidently identified if at least four consecutive b or y ions for each peptide were observed. Cross-links used for this study are in [Dataset S1](#).

**Sample Preparation for Cryo-EM.** To form stable MLL1-WRAD-nucleosome complexes, 100 nM of modified nucleosome was mixed with 5× WRAD in EM buffer (20 mM Hepes, 300 mM NaCl, 2 mM TCEP, 0.1 µM Zinc chloride) and incubated for 1 h on ice. The sample mix was then cross-linked using fresh 0.05% (wt/vol) glutaraldehyde for 1 h on ice and quenched with 100 mM Tris for 2 h. The sample was concentrated to 0.5 mg/mL and directly applied to freshly glow-discharged 2/2 Cu Quantifoil grids in a Vitrobot Mark IV (Thermo Fisher Scientific) at 4 °C and 100% humidity. The sample was immediately blotted (3-s blot time) and flash frozen in liquid ethane.

**Cryo-EM Data Acquisition and Image Processing.** All data were acquired at the Johns Hopkins Beckman Cryo EM Center on a Thermo Fisher Titan Krios G3 equipped with a Gatan K3 direct electron detector at a magnification of 21,500

in superresolution counting mode, corresponding to a pixel size of 0.529 Å/pix. A total of 5,091 movies were recorded using Serial-EM (67) using a varying negative defocus of 1.0 μm to 2.8 μm and recording 40 frames at 1.5 e<sup>4</sup>/Å<sup>2</sup> per frame at 60 e<sup>4</sup>/Å<sup>2</sup> total dose. Data acquisition was monitored and frequently evaluated using cisTEM (68). Movie stacks were aligned and down-scaled to a pixel size of 1.058 Å/pix (bin 1) using MotionCor2 (69), and contrast transfer function (CTF) correction was performed using Ctfind4 (70). The full dataset was then manually inspected, and 4,804 movie stacks were selected for further processing in Relion 3.0 (71).

Initial particle picking yielded 2,511,854 particles that were subjected to two-dimensional (2D) and 3D classification using threefold downsampled particle images (bin3) for faster processing speed and stronger signal-to-noise ratio, which removed junk particles and yielded 608,972 “good” particles that displayed MLL1 complex bound to the nucleosome. All “good” particles were reextracted at 1.058 Å/pix (bin1), consensus refined, and subjected to masked 3D classification using an Ash2L DNA interface mask, which yielded six unique classes displaying differences in Ash2L density and DNA topology (SI Appendix, Fig. S1). The 103,849 particles were subjected to one round of CTF refinement followed by Bayesian polishing as implemented in Relion-3.0. This procedure yielded a reconstruction after refinement of 3.33-Å resolution based on the FSC 0.143 (72) criterion termed state 1 (SI Appendix, Figs. S1 and S2), which was sharpened using an automatically calculated B factor of -70 Å<sup>2</sup>. Particles grouped in classes 2 and 5 from the initial masked classification were combined after confirming similar classes by visual inspection and reclassified into four classes (SI Appendix, Fig. S1). The resulting best class containing 134,528 particles showed flexible Ash2L and MLL1 density and density stretches associated with the DNA between SHL -6 to SHL -7 and was termed state 2. Particles were subjected to one round of CTF refinement followed by Bayesian polishing, yielding a 3.55-Å resolution reconstruction (according to the FSC 0.143 criterion) that was sharpened using an automatically calculated B factor of -86 Å<sup>2</sup>. The 56,526 particles grouped in class 3 from the initial masked classification were CTF corrected, particle polished, and subsequently refined to a final resolution of 4.65 Å (0.143 criterion). This reconstruction was sharpened using an automatically calculated B factor of -117 Å<sup>2</sup> (SI Appendix, Fig. S2) and termed state 3. The 66,449 particles grouped in class 4 from the initial masked classification were CTF refined, particle polished, and further refined to a final resolution of 3.95 Å (based on the 0.143 criterion). This class contained highly resolved nucleosome density but more poorly resolved MLL1 complex and Ash2L density. To further deblur the Ash2L components in this state, an additional 3D classification using only local searches was performed (SI Appendix, Fig. S1). This process yielded an overall better-resolved class (25,525 particles) that refined to 4.25 Å after postprocessing and was sharpened using a B factor of -75 Å<sup>2</sup>. This map (state 4) was used to guide Ash2L modeling as described below. We note that efforts to identify additional states using cryoDRGN (73) did not yield useful results.

**Model Building and Refinement.** Crystal structures of human ubiquitin (1UBQ), human WDR5<sup>31-334</sup> (3EG6), human MLL1<sup>3814-3969</sup>-Ash2L<sup>285-504</sup>, RbBP5<sup>330-375</sup> (5F6L), DPY30<sup>50-96</sup> (6E2H), human RbBP5<sup>11-329</sup>, and *X. laevis* nucleosome (6KIU) were subjected to rigid-body fitting in the state 4 map using Chimera (74). These combined fitted structures allowed us to assign most of the density of the state 4 map, except for a region of continuous density (Fig. 1C) which was judged to correspond to Ash2L and DPY30 based on the similarity of the complex to solved structures of yeast homologs (49–51). The Ash2L model was computed through iterative rounds of cross-linking-based homology modeling utilizing available structures and manual model building, followed by MDFF using our state 4 map. Specifically, we use the previously resolved SPRY domain (287 to 504) from the MLL1<sup>SET</sup>-ASH2L<sup>SPRY</sup>-RbBP5<sup>330-375</sup> complex (PDB ID 5F6L) (36) and added residues 398 to 440 from a recently reported Ash2L homology

model (38). All MDFF simulations were done in NAMD2.13 (75). The CHARMM36 force field (76, 77) was used for both protein and DNA. All simulations were carried out at 300 K in vacuum with a scaling factor of 1.0. To avoid potential structural artifacts that can arise from MDFF, secondary structure, chirality, and cis-peptide restraints were applied. For targeted refinement of Ash2L and DPY30, a 500 kcal/mol/Å harmonic restraint was applied on all backbone atoms except for Ash2L and DPY30. The cross-correlation coefficient was computed using the MDFF package implemented in VMD 1.9.3 (78). We then manually optimized the models fit to our experimental density, which generated the initial Ash2L model. We assigned adjacent helical tube density to DPY30 and performed automated fitting of the Ash2L-SDI/DPY30 dimer crystal structure (PDB 6E2H). In addition, our density allowed us to fully build the C-terminal helix of Ash2L (487 to 528). Ash2L residues 199 to 286 were built using MODELER (79) using cross-link restraints and underwent iterative rounds of manual model building guided by the EM density. The model was further refined through a 3.0-ns MDFF simulation (41). To improve the overall fit of Ash2L and DPY30, the state 4 map was filtered to 6.0 Å and was used as the template map. After refinement with MDFF, the cross-correlation coefficient between the overall structure and template map improved from 0.79 to 0.83. To correct any rotamer outliers that may occur from MDFF refinement, the model of the full complex underwent 1,000 iterations of minimization with secondary structure restraints using the Phenix geometry minimization module. Following, the model was iteratively refined with Phenix real-space refinement. Final statistics for the state 4 model are shown in SI Appendix, Table S1.

For the state 2 model, the positioning of WDR5, RbBP5<sup>11-323</sup>, ubiquitin, and the nucleosome in state 4 were used as starting initial positions. WDR5 was then subjected to rigid-body fitting into the state 2 map using the Fit in Map function in ChimeraX. The entire structure was then subject to real-space refinement in Phenix. Final statistics for the state 2 model are shown in SI Appendix, Table S2.

**Ash2L N-Terminal Docking.** Docking of the human Ash2L WH-PHD domain (PDB 3RSN) guided by cross-linking data was carried out with HADDOCK 2.4 (80). Cross-links were used as unambiguous restraints between the Ash2L N-terminal residues and the remainder of the MLL1-WRAD complex. In HADDOCK, 5,000 rigid-body calculations were first carried out, and the top 1,000 structures underwent semiflexible refinement, resulting in 200 water-refined structures. Clusters were defined using fraction of common contacts with a cutoff of 0.60. All 200 structures were identified to a single cluster with a final HADDOCK score of -213.7 ± 1.4.

**Data, Materials, and Software Availability.** Coordinates of the full active complex, state 4, and the state 2 complex have been deposited in PDB with accession codes 7UD5 (81) and 8DU4 (82), respectively. Maps for states 1 to 4 have been deposited in the Electron Microscopy Data Bank with accession codes EMD-27802 (state 1) (83), EMD-27715 (state 2) (84), EMD-27803 (state 3) (85), and EMD-26454 (state 4) (86). MS data were deposited in JPOST (87), accession number JPST001677 (88).

**ACKNOWLEDGMENTS.** We thank Duncan Sousa for assistance with data collection. We thank Dr. Phil Gafkin at the Fred Hutchinson Cancer Research Center proteomic facility and Dr. Ebbing De Jong at the State University of New York-Upstate proteomics facility for their help with MS. Computational resources were provided by the Maryland Advanced Research Computing Center. This work was supported by a European Molecular Biology Organization Long-Term Fellowship (N.A.H.), a Damon Runyon Cancer Research Fund Postdoctoral Fellowship (E.J.W.), National Institute of General Medical Sciences Grants GM130393 (C.W.) and T32 GM008403 (S.R.), and National Cancer Institute Grants CA184235 (B.A.K.) and CA140522 (M.S.C.).

1. T. Jenuwein, C. D. Allis, Translating the histone code. *Science* **293**, 1074–1080 (2001).
2. A. E. Ringel, A. M. Cieniewicz, S. D. Taverna, C. Wolberger, Nucleosome competition reveals processive acetylation by the SAGA HAT module. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5461–E5470 (2015).
3. C. Bian *et al.*, Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO J.* **30**, 2829–2842 (2011).
4. M. Vermeulen *et al.*, Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).

5. S. M. Lauberth *et al.*, H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152**, 1021–1036 (2013).
6. R. J. Sims III *et al.*, Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J. Biol. Chem.* **280**, 41789–41792 (2005).
7. J. Wysocka *et al.*, A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86–90 (2006).
8. Z. W. Sun, C. D. Allis, Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* **418**, 104–108 (2002).

9. K. Hyun, J. Jeon, K. Park, J. Kim, Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* **49**, e324 (2017).
10. A. Shilatifard, The COMPASS family of histone H3K4 methylases: Mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* **81**, 65–95 (2012).
11. H. Santos-Rosa *et al.*, Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).
12. M. Kwon *et al.*, H2B ubiquitylation enhances H3K4 methylation activities of human KMT2 family complexes. *Nucleic Acids Res.* **48**, 5442–5456 (2020).
13. N. Minsky *et al.*, Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nat. Cell Biol.* **10**, 483–488 (2008).
14. M. B. Chandrasekharan, F. Huang, Z. W. Sun, Histone H2B ubiquitination and beyond: Regulation of nucleosome stability, chromatin dynamics and the trans-histone H3 methylation. *Epigenetics* **5**, 460–468 (2010).
15. A. Patel *et al.*, Automethylation activities within the mixed lineage leukemia-1 (MLL1) core complex reveal evidence supporting a “two-active site” model for multiple histone H3 lysine 4 methylation. *J. Biol. Chem.* **289**, 868–884 (2014).
16. Y. Dou *et al.*, Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat. Struct. Mol. Biol.* **13**, 713–719 (2006).
17. A. Patel, V. Dharmarajan, V. E. Vought, M. S. Cosgrove, On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. *J. Biol. Chem.* **284**, 24242–24256 (2009).
18. S. Takahashi, A. Yokoyama, The molecular functions of common and atypical MLL fusion protein complexes. *Biochim. Biophys. Acta. Gene Regul. Mech.* **1863**, 194548 (2020).
19. C. Meyer *et al.*, The MLL recombinome of acute leukemias in 2017. *Leukemia* **32**, 273–284 (2018).
20. A. V. Krivtsov, S. A. Armstrong, MLL translocations, histone modifications and leukaemia stem-cell development. *Nat. Rev. Cancer* **7**, 823–833 (2007).
21. V. Dharmarajan, J. H. Lee, A. Patel, D. G. Skalnik, M. S. Cosgrove, Structural basis for WDR5 interaction (Win) motif recognition in human SET1 family histone methyltransferases. *J. Biol. Chem.* **287**, 27275–27289 (2012).
22. F. Cao *et al.*, Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol. Cell* **53**, 247–261 (2014).
23. S. A. Shinsky, K. E. Monteith, S. Viggiano, M. S. Cosgrove, Biochemical reconstitution and phylogenetic comparison of human SET1 family core complexes involved in histone methylation. *J. Biol. Chem.* **290**, 6361–6375 (2015).
24. M. M. Steward *et al.*, Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat. Struct. Mol. Biol.* **13**, 852–854 (2006).
25. J. Z. Stoller *et al.*, Ash2l interacts with Tbx1 and is required during early embryogenesis. *Exp. Biol. Med. (Maywood)* **235**, 569–576 (2010).
26. C. C. Tan *et al.*, Transcription factor Ap2delta associates with Ash2l and ALR, a trithorax family histone methyltransferase, to activate Hoxc8 transcription. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7472–7477 (2008).
27. S. Rampalli *et al.*, p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation. *Nat. Struct. Mol. Biol.* **14**, 1150–1156 (2007).
28. L. Li *et al.*, The COMPASS family protein ASH2L mediates corticogenesis via transcriptional regulation of Wnt signaling. *Cell Rep.* **28**, 698–711.e5 (2019).
29. M. Wan *et al.*, The trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells. *J. Biol. Chem.* **288**, 5039–5048 (2013).
30. Y. Chen *et al.*, Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding. *EMBO Rep.* **12**, 797–803 (2011).
31. S. Sarvan *et al.*, Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain. *Nat. Struct. Mol. Biol.* **18**, 857–859 (2011).
32. Y. Chen, F. Cao, B. Wan, Y. Dou, M. Lei, Structure of the SPRY domain of human Ash2L and its interactions with RbBP5 and DPY30. *Cell Res.* **22**, 598–602 (2012).
33. J. F. Haddad *et al.*, Structural analysis of the Ash2L/Dpy-30 complex reveals a heterogeneity in H3K4 methylation. *Structure* **26**, 1594–1603.e4 (2018).
34. P. Zhang *et al.*, A phosphorylation switch on RbBP5 regulates histone H3 Lys4 methylation. *Genes Dev.* **29**, 123–128 (2015).
35. S. A. Shinsky *et al.*, 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 (2014).
36. Y. Li *et al.*, Structural basis for activity regulation of MLL family methyltransferases. *Nature* **530**, 447–452 (2016).
37. V. Tremblay *et al.*, Molecular basis for DPY-30 association to COMPASS-like and NURF complexes. *Structure* **22**, 1821–1830 (2014).
38. S. H. Park *et al.*, Cryo-EM structure of the human MLL1 core complex bound to the nucleosome. *Nat. Commun.* **10**, 5540 (2019).
39. H. Xue *et al.*, Structural basis of nucleosome recognition and modification by MLL methyltransferases. *Nature* **573**, 445–449 (2019).
40. P. L. Hsu *et al.*, Crystal structure of the COMPASS H3K4 methyltransferase catalytic module. *Cell* **174**, 1106–1116.e9 (2018).
41. L. G. Trabuco, E. Villa, K. Mitra, J. Frank, K. Schulten, Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).
42. M. T. Morgan *et al.*, Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science* **351**, 725–728 (2016).
43. P. W. Lewis *et al.*, Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* **340**, 857–861 (2013).
44. E. J. Worden, N. A. Hoffmann, C. W. Hicks, C. Wolberger, Mechanism of cross-talk between H2B ubiquitination and H3 methylation by Dot1L. *Cell* **176**, 1490–1501.e12 (2019).
45. E. D. Merkley *et al.*, Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci.* **23**, 747–759 (2014).
46. Z. L. Chen *et al.*, A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun.* **10**, 3404 (2019).
47. J. Kosinski *et al.*, Xlink Analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* **189**, 177–183 (2015).
48. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
49. E. J. Worden, X. Zhang, C. Wolberger, Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome. *eLife* **9**, e53199 (2020).
50. P. L. Hsu *et al.*, Structural basis of H2B ubiquitination-dependent H3K4 methylation by COMPASS. *Mol. Cell* **76**, 712–723.e4 (2019).
51. Q. Qu *et al.*, Structure and conformational dynamics of a COMPASS histone H3K4 methyltransferase complex. *Cell* **174**, 1117–1126.e12 (2018).
52. Y. T. Lee *et al.*, Mechanism for DPY30 and ASH2L intrinsically disordered regions to modulate the MLL/SET1 activity on chromatin. *Nat. Commun.* **12**, 2953 (2021).
53. K. E. W. Namitz, S. Tan, M. S. Cosgrove, Hierarchical assembly of the MLL1 core complex within a biomolecular condensate regulates H3K4 methylation. *bioRxiv [Preprint]* (2019). <https://doi.org/10.1101/870667>. Accessed 12 April 2022.
54. A. Ayoub, S. H. Park, Y. T. Lee, U. S. Cho, Y. Dou, Regulation of MLL1 methyltransferase activity in two distinct nucleosome binding modes. *Biochemistry* **61**, 1–9 (2022).
55. H. Vlaming, F. van Leeuwen, The upstreams and downstreams of H3K79 methylation by DOT1L. *Chromosoma* **125**, 593–605 (2016).
56. R. K. McGinty, J. Kim, C. Chatterjee, R. G. Roeder, T. W. Muir, Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **453**, 812–816 (2008).
57. C. J. Anderson *et al.*, Structural basis for recognition of ubiquitylated nucleosome by Dot1L methyltransferase. *Cell Rep.* **26**, 1681–1690.e5 (2019).
58. M. I. Valencia-Sánchez *et al.*, Structural basis of Dot1L stimulation by histone H2B lysine 120 ubiquitination. *Mol. Cell* **74**, 1010–1019.e6 (2019).
59. S. Jang *et al.*, Structural basis of recognition and destabilization of the histone H2B ubiquitinated nucleosome by the DOT1L histone H3 Lys79 methyltransferase. *Genes Dev.* **33**, 620–625 (2019).
60. S. Nakanishi *et al.*, A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. *Nat. Struct. Mol. Biol.* **15**, 881–888 (2008).
61. A. Patel, V. E. Vought, V. Dharmarajan, M. S. Cosgrove, A novel non-SET domain multi-subunit methyltransferase required for sequential nucleosomal histone H3 methylation by the mixed lineage leukemia protein-1 (MLL1) core complex. *J. Biol. Chem.* **286**, 3359–3369 (2011).
62. A. Patel, V. E. Vought, V. Dharmarajan, M. S. Cosgrove, 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 (2008).
63. P. N. Dyer *et al.*, Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375**, 23–44 (2004).
64. M. Morgan, M. Jbara, A. Brik, C. Wolberger, Semisynthesis of ubiquitinated histone H2B with a native or nonhydrolyzable linkage. *Methods Enzymol.* **618**, 1–27 (2019).
65. B. A. Knutson, J. Luo, J. Ranish, S. Hahn, Architecture of the *Saccharomyces cerevisiae* RNA polymerase I Core Factor complex. *Nat. Struct. Mol. Biol.* **21**, 810–816 (2014).
66. B. A. Knutson, M. L. Smith, A. E. Belkevich, A. M. Fakhouri, Molecular topology of RNA polymerase I upstream activation factor. *Mol. Cell Biol.* **40**, e00056-20 (2020).
67. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
68. T. Grant, A. Rohou, N. Grigorieff, *cstEM*, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018).
69. S. Q. Zheng *et al.*, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
70. A. Rohou, N. Grigorieff, CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
71. S. H. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
72. M. van Heel, M. Schatz, Fourier shell correlation threshold criteria. *J. Struct. Biol.* **151**, 250–262 (2005).
73. E. D. Zhong, T. Bepler, B. Berger, J. H. Davis, CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).
74. Z. Yang *et al.*, UCSF Chimera, MODELLER, and IMP: An integrated modeling system. *J. Struct. Biol.* **179**, 269–278 (2012).
75. J. C. Phillips *et al.*, Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
76. J. Huang, A. D. MacKerell Jr., CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
77. K. Hart *et al.*, Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BI conformational equilibrium. *J. Chem. Theory Comput.* **8**, 348–362 (2012).
78. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
79. B. Webb, A. Sali, Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **54**, 5.6.1–5.6.3.7 (2016).
80. G. C. P. van Zundert *et al.*, The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725 (2016).
81. N. A. Hoffmann, S. Rahman, E. J. Worden, C. Wolberger, Complex between MLL1-WRAD and an H2B-ubiquitinated nucleosome. RCSB. <http://www.rcsb.org/pdb/explore/explore.do?structureId=7UD5>. Deposited 18 March 2022.
82. N. A. Hoffmann, S. Rahman, E. J. Worden, C. Wolberger, Complex between RbBP5-WDR5 and an H2B-ubiquitinated nucleosome. RCSB. <http://www.rcsb.org/pdb/explore/explore.do?structureId=8DU4>. Deposited 26 July 2022.
83. N. A. Hoffmann, S. Rahman, E. J. Worden, C. Wolberger, Complex between MLL1-WRAD and an H2B-ubiquitinated nucleosome- State 1. EMD. <https://www.emdataresource.org/EMD-27802>. Deposited 9 August 2022.
84. N. A. Hoffmann, S. Rahman, E. J. Worden, C. Wolberger, Complex between MLL1-WRAD and an H2B-ubiquitinated nucleosome- State 2. EMD. <https://www.emdataresource.org/EMD-27715>. Deposited 26 July 2022.
85. N. A. Hoffmann, S. Rahman, E. J. Worden, C. Wolberger, Complex between MLL1-WRAD and an H2B-ubiquitinated nucleosome- State 3. EMD. <https://www.emdataresource.org/EMD-27803>. Deposited 9 August 2022.
86. N. A. Hoffmann, S. Rahman, E. J. Worden, C. Wolberger, Complex between MLL1-WRAD and an H2B-ubiquitinated nucleosome- State 4. EMD. <https://www.emdataresource.org/EMD-26454>. Deposited 18 March 2022.
87. S. Okuda *et al.*, iPOSTrepo: An international standard data repository for proteomes. *Nucleic Acids Res.* **45**, D1107–D1111 (2017).
88. B. A. Knutson, Multistate structures of the MLL1-WRAD complex bound to H2B-ubiquitinated nucleosome. JPOST. <https://repository.jpostdb.org/entry/JPOST001677>. Deposited 28 June 2022.