

A Conserved Arginine-containing Motif Crucial for the Assembly and Enzymatic Activity of the Mixed Lineage Leukemia Protein-1 Core Complex^{*[5]}

Received for publication, August 14, 2008, and in revised form, September 19, 2008 Published, JBC Papers in Press, October 1, 2008, DOI 10.1074/jbc.M806317200

Anamika Patel¹, Valarie E. Vought, Venkatasubramanian Dharmarajan, and Michael S. Cosgrove²

From the Department of Biology, Syracuse University, Syracuse, New York 13244

The mixed lineage leukemia protein-1 (MLL1) belongs to the SET1 family of histone H3 lysine 4 methyltransferases. Recent studies indicate that the catalytic subunits of SET1 family members are regulated by interaction with a conserved core group of proteins that include the WD repeat protein-5 (WDR5), retinoblastoma-binding protein-5 (RbBP5), and the absent small homeotic-2-like protein (Ash2L). It has been suggested that WDR5 functions to bridge the interactions between the catalytic and regulatory subunits of SET1 family complexes. However, the molecular details of these interactions are unknown. To gain insight into the interactions among these proteins, we have determined the biophysical basis for the interaction between the human WDR5 and MLL1. Our studies reveal that WDR5 preferentially recognizes a previously unidentified and conserved arginine-containing motif, called the “Win” or WDR5 interaction motif, which is located in the N-SET region of MLL1 and other SET1 family members. Surprisingly, our structural and functional studies show that WDR5 recognizes arginine 3765 of the MLL1 Win motif using the same arginine binding pocket on WDR5 that was previously shown to bind histone H3. We demonstrate that WDR5’s recognition of arginine 3765 of MLL1 is essential for the assembly and enzymatic activity of the MLL1 core complex *in vitro*.

Eukaryotes have evolved a complex system to regulate access to genomic information by packaging DNA into chromatin. Chromatin can adopt various levels of organization that regulate essential cellular activities such as transcription, DNA replication, recombination, and repair (1). The fundamental repeating unit of chromatin is the nucleosome, in which 147 base pairs of genomic DNA is wrapped around a disc-shaped octamer of histone proteins H2A, H2B, H3, and H4 (2, 3). Nucleosome positioning on DNA is fundamentally involved in controlling gene access and is regulated in part by a diverse array of enzymes that introduce covalent post-translational

modifications on histone proteins (4, 5). An extensive array of histone modifications have been characterized, including lysine acetylation, lysine and arginine methylation, serine and threonine phosphorylation, and lysine ubiquitylation (6, 7). The large number of potential histone modification patterns provides cells with an enormous combinatorial potential for the precise regulation of gene function. This potential is summarized by the histone code hypothesis (8–10), which predicts that patterns of histone modifications are recognized by specialized “effector” domains found in numerous chromatin regulators. It is thought that effector domains help recruit the activities of proteins that either stabilize or remodel specific chromatin states (5, 11, 12).

The combinatorial complexity of histone modifications is further increased by histone lysine residues that can be mono-, di-, or trimethylated, which are correlated with distinct functional outcomes. For example, chromatin immunoprecipitation experiments suggest that nucleosomes occupying the 5′ ends of active genes are trimethylated at H3 lysine 4 (H3K4me3), whereas the body of active genes tend to possess nucleosomes dimethylated at H3K4 (13–17). H3K4me3 is thought to activate transcription through the recruitment of proteins that mobilize nucleosomes by ATP-dependent nucleosome-remodeling enzymes, such as CHD1 (18) or NURF (19). Several enzymes that regulate H3K4 methylation have been identified, all possessing the evolutionarily conserved SET domain, which is required for lysine methylation (20, 21). Members of the SET1 family of SET domain proteins assemble into multisubunit complexes that regulate mono-, di-, and trimethylation of H3K4. In budding yeast, evidence suggests that SET1p is the sole H3K4 methyltransferase (22), which exists in a multisubunit complex called COMPASS (23–25). In contrast, humans encode at least six different SET1 family members: the mixed lineage leukemia proteins MLL1–4 and the SET1a and SET1b proteins (26–34). The most studied human SET1 family member is MLL1 (ALL1, HRX, Htrx), which is required for the regulation of *hox* genes in hematopoiesis and development (29, 35, 36). MLL1 assembles into distinct multisubunit complexes that share several proteins that are conserved among SET1 family complexes from yeast to humans (28, 37, 38).

Recent evidence suggests that the enzymatic activity of SET1 family complexes is regulated by specific protein-protein interactions (30, 39). For example, it was recently shown that the minimal complex required for di- and trimethylation of H3K4 includes MLL1, WDR5, RbBP5, and Ash2L, which together form the MLL1 core complex (39). Evidence suggests that

^{*} This work was supported by Research Grant 5-FY06-585 from the March of Dimes Birth Defects Foundation (to M. S. C.) and by funds from Syracuse University. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[5] The on-line version of this article (available at <http://www.jbc.org>) contains a supplemental figure and supplemental references.

¹ Supported in part by Richard A. Lebkowski Life Science Fellowship.

² To whom correspondence should be addressed: Dept. of Biology, Syracuse University, 107 College Pl., Syracuse, NY 13244. Tel.: 315-443-2964; Fax: 315-443-2102; E-mail: mcosgro@sy.edu.

WDR5, RbBP5, and Ash2L form a stable subcomplex that is capable of interacting with the different members of the SET1 family of SET domain proteins (39). The WD40 repeat protein WDR5 is critical for these interactions, as it has been shown to bridge the interactions between the catalytic SET domain of MLL1 and the RbBP5 and Ash2L regulatory components of the MLL1 core complex (39). RNA interference knockdown of WDR5 in mammalian cells results in the global loss of H3K4me3 (39, 40), which is correlated with decreased expression of *Hox* genes and defects in hematopoiesis and development (40). Similar phenotypes are observed in MLL1^{-/-} and MLL1ΔSET mice (35, 36), suggesting that WDR5 and MLL1 function together to regulate *Hox* gene expression *in vivo* (41). In this and the accompanying investigation (58), we identify and characterize a highly conserved WDR5 interaction motif, or “Win” motif, which is located in the N-SET region of MLL1 and other SET1 family members. In addition, we demonstrate that the *Win* motif is critical for the assembly and enzymatic activity of the MLL1 core complex *in vitro*.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—A human spleen c-DNA library was used to amplify an MLL1 C-terminal fragment encoding residues 3592–3969. PCR subcloning was used to amplify MLL1 SET domain constructs consisting of residues 3745–3969 (MLL³⁷⁴⁵) and residues 3811–3969 (MLL³⁸¹¹). MLL³⁷⁴⁵ and MLL³⁸¹¹ were ligated into the pGST and pMBP parallel vectors, respectively (42). pMCGS7 plasmids encoding full-length WDR5 as N-terminal His₆ fusions were obtained as generous gifts from Alexander Ruthenburg and David Allis. A plasmid containing the RbBP5 clone was obtained as a generous gift from Yali Dou and was subcloned into the pET3a vector without affinity tags. The *ash2L* gene (Clone ID 3921999) was purchased from Open Biosystems and subcloned into the pHis Parallel vector, which encodes tobacco etch virus (TEV)³ cleavable N-terminal His₆ fusion.

All recombinant proteins were overexpressed in *Escherichia coli* (Rosetta II, Novagen) by growing cells containing the plasmids at 37 °C in Terrific Broth medium containing 50 μg/ml carbenicillin to an A₆₀₀ of 1.0. The temperature was then lowered to 16 °C, and cells were induced with isopropyl-1-thio-β-D-galactopyranoside (0.1–1 mM) for 16–18 h. RbBP5 and Ash2L were induced with 0.1 mM and 0.25 mM isopropyl-1-thio-β-D-galactopyranoside, respectively. Cells were harvested, resuspended in lysis buffer (50 mM Tris, pH 7.3, 300 mM NaCl, 10% glycerol, 3 mM dithiothreitol, 0.1 mM phenylmethylsulfonyl fluoride, and EDTA-free protease inhibitor mixture (Roche Applied Science)), lysed with a microfluidizer cell disrupter, and clarified by centrifugation. Cells containing the RbBP5 plasmid were lysed instead using the same lysis buffer containing 1× BugBuster® (Novagen), which minimized RbBP5 degradation upon cell lysis. Clarified supernatants containing the GST-MLL³⁷⁴⁵ protein were passed over a glutathione-Sepharose column (GSTrapTM FF column, GE-Healthcare), and

GST-MLL³⁷⁴⁵ was eluted with a gradient of reduced glutathione. Fractions containing GST-MLL³⁷⁴⁵ were combined, treated with TEV protease, and dialyzed with three changes against lysis buffer (without protease inhibitors). MLL³⁷⁴⁵ was further purified over a glutathione-Sepharose column followed by gel filtration chromatography. MBP-MLL³⁸¹¹ fusion protein was purified by amylose affinity chromatography followed by TEV protease treatment and two rounds of gel filtration chromatography. Full-length WDR5 protein was expressed and purified as described previously (43). RbBP5 protein was purified by ion exchange and gel filtration chromatography. Ash2L was purified by nickel affinity chromatography (HisTrap column, GE Healthcare), dialysis and TEV cleavage to remove imidazole and the His₆ tag, and repurification over a His-trap column. As a final step of purification and for buffer exchange, all proteins were passed through a gel filtration column (Superdex 200, GE Healthcare) pre-equilibrated with 20 mM Tris (pH 7.5), 300 mM NaCl, 1 mM TCEP, and 1 μM ZnCl₂.

Mutagenesis and Peptides—Point mutations were introduced into MLL1 and WDR5 constructs using the QuikChange site-directed mutagenesis kit (Stratagene). Plasmids were sequenced to verify the presence of the intended mutations and the absence of additional mutations. Peptides were synthesized by Global Peptide and Genscript.

Analytical Ultracentrifugation—Analytical ultracentrifugation experiments were carried out using a Beckman Coulter ProteomeLabTM XL-A analytical ultracentrifuge equipped with absorbance optics and an eight-hole An-50 Ti analytical rotor. Sedimentation velocity experiments were carried out at 10 °C and 50,000 rpm (200,000 × g) using 3-mm two-sector charcoal-filled Epon centerpieces with quartz windows. Each sample was scanned at 0-min time intervals for 300 scans. Protein samples in 20 mM Tris-Cl, pH 7.5, 300 mM NaCl, 1 mM TCEP, and 1 μM ZnCl₂ were run at various concentrations, and molar ratios were as described under “Results.” Sedimentation boundaries were analyzed by the continuous distribution (c(s)) method using the program SEDFIT (44). Equilibrium dissociation constants for WDR5-MLL1 complexes were obtained by global fitting sedimentation velocity data acquired at several different protein concentrations using the single-site hetero-association model (A + B ↔ AB) of SEDPHAT (45, 46). The program SEDNTERP, version 1.09 (47), was used to correct the experimental s value (s^{*}) to standard conditions at 20 °C in water (s_{20,w}) and to calculate the partial specific volume of each protein.

Methyltransferase Assays—Radiolabeling assays were conducted by combining 7 μg of MLL³⁷⁴⁵ or MLL³⁸¹¹ with 500 μM histone H3 peptide containing residues 1–20 (with GSK-biotin on the C terminus) and 1 μCi of [³H]methyl-S-adenosylmethionine (GE Healthcare) in 50 mM Tris, pH 8.5, 200 mM NaCl, 3 mM dithiothreitol, 5 mM MgCl₂, and 5% glycerol. The reactions were incubated at 15 °C for 2 h, stopped by the addition of SDS-loading buffer to 1×, and separated by SDS-PAGE on a 4–12% gradient gel (Invitrogen). The gel then was soaked in an autoradiography enhancer solution (Enlightning, PerkinElmer Life Sciences), dried, and exposed to film at –80 °C for 24 h.

Mass spectrometry assays were conducted by adapting that previously reported for Dim5 and SET7/9 histone methyltransferases (48). Six micrograms of MLL³⁷⁴⁵ or MLL1 core complex

³ The abbreviations used are: TEV, tobacco etch virus; MALDI-TOF, matrix-assisted laser desorption/ionization-time of flight; TCEP, tris(2-carboxyethyl)phosphine.

WRD5 Recognizes Arg-3765 of MLL1

was incubated with 250 μ M S-adenosyl-methionine and 10 μ M histone H3 peptide (amino acid residues 1–20) at 15 °C in 50 mM Tris-Cl, pH 9.0, 200 mM NaCl, 3 mM dithiothreitol, and 5% glycerol. The reactions were quenched at various time points by the addition of trifluoroacetic acid to 0.5%. The quenched samples were diluted 1:4 with α -cyano-4-hydroxycinnamic acid. MALDI-TOF mass spectrometry was performed on a Bruker AutoFlex mass spectrometer (State University of New York (SUNY), Oswego, NY) operated in reflectron mode. Final spectra were averaged from 100 shots/position at 10 different positions.

Circular Dichroism Spectroscopy—CD spectra were collected on an AVIV 62A DS spectropolarimeter equipped with a Neslab CFT-33 refrigerated circulator at 10 °C using a 0.1-mm path length cell. Spectra for each protein at a concentration of 0.2 mg/ml were collected in a buffer containing 10 mM Tris (pH 7.5), 200 mM NaCl, 1 mM TCEP, and 1 μ M ZnCl₂. The background contribution of the buffer alone was subtracted from each protein spectrum.

RESULTS

WDR5 Forms a Stable Complex with the N-SET Region of MLL1—MLL1 is a large protein of 3,969 residues containing several conserved domains with functions implicated in transcriptional regulation (49) (Fig. 1). Previous efforts to map the interaction region between MLL1 and WDR5 led to the identification of three distinct WDR5 interaction regions in MLL1 (38). The most C-terminal MLL1 fragment that interacts with WDR5 includes residues 3301–3969, which includes the N-SET region and the evolutionarily conserved ~130-amino-acid SET domain at the extreme C terminus. The location of the WDR5 binding site within this region is unknown. Previous immunoprecipitation experiments showed that deletion of the last 149 amino acid residues spanning the MLL1 SET domain prevents co-immunoprecipitation of WDR5, RbBP5, and Ash2L with MLL1 (28), suggesting that the conserved SET domain of MLL1 is important for the interaction with the subcomplex. In contrast, domain-mapping experiments with the human SET1a protein suggest that ~120 amino acid residues in the N-SET region are required for co-immunoprecipitation of WDR5, RbBP5, and Ash2L (33). A similar region of the MLL3 protein was shown to interact with WDR5 in GST pulldown experiments (50). These results suggest that the N-SET region of MLL1 may be recognized by the WDR5 component of the MLL1 core complex. To test this hypothesis, we performed sedimentation velocity analytical ultracentrifugation experiments to compare the hydrodynamic properties of two highly pure MLL1 SET domain constructs in the presence and absence of WDR5.

The first MLL1 SET domain construct consisted of residues 3745–3969 of human MLL1 (hereafter referred to as MLL³⁷⁴⁵), which was shown in previous studies to be the minimal domain required for interaction with the In1 tumor suppressor (51) and for MLL1 SET domain self-association (52). This construct contains 66 amino acid residues of the N-SET region followed by the conserved SET and post-SET domains (Fig. 1a). We show in a separate investigation that MLL³⁷⁴⁵ is the minimal MLL SET domain construct required for the assembly of the

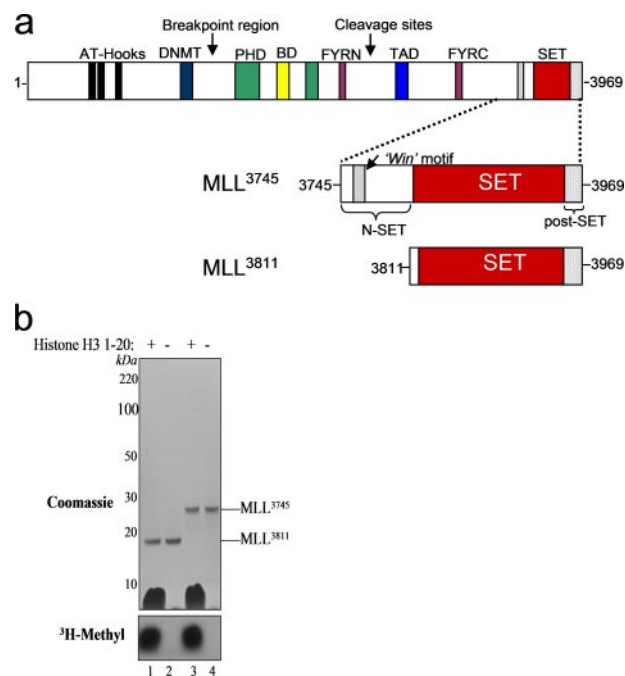


FIGURE 1. Schematic representation of the mixed lineage leukemia protein-1 (MLL1) showing the domain architecture of full-length and recombinant constructs used in this investigation. a, full-length MLL1 contains 3,969 amino acid residues with a conserved SET domain at the C terminus. Other conserved domains include DNA binding AT-hooks, DNA methyltransferase like (DNMT) domain, plant homeodomains (PHD), and a bromodomain (BD), conserved FYRN and FYRC sequences, and a tandem activation domain (TAD). Recombinant MLL1 construct MLL³⁷⁴⁵ consists of residues 3745–3969 and includes the evolutionarily conserved SET and post-SET domains, as well as the Win motif-containing N-SET region. MLL³⁸¹¹ consists of residues 3811–3969 and lacks the N-SET region. b, enzymatic assays with purified MLL³⁸¹¹ (lanes 1 and 2) and MLL³⁷⁴⁵ (lanes 3 and 4) SET domain proteins. The upper panel shows Coomassie Blue-stained SDS-PAGE gel of enzymatic reactions with and without a histone H3 peptide (residues 1–20). The lower panel shows [³H]methyl incorporation into the peptides catalyzed by MLL³⁸¹¹ or MLL³⁷⁴⁵ after 2 h.

full MLL1 core complex *in vitro*.⁴ In the second MLL1 SET domain construct, the N-terminal residues encompassing the N-SET region were deleted. This construct consists of residues 3811–3969 (MLL³⁸¹¹) and encompasses only the conserved SET and post-SET domains (Fig. 1a). Both constructs are catalytically active in methyltransferase assays with histone H3 peptides encompassing residues 1–20 (Fig. 1b).

Sedimentation velocity analyses were first performed with each of the proteins individually and were fitted to a distribution of Lamm equation solutions to determine the diffusion-free sedimentation coefficient distribution (c(s)) (44). Fig. 2, a–c, shows that MLL³⁷⁴⁵, MLL³⁸¹¹, and WDR5 are monodisperse in solution with experimental *s** values independent of protein concentration (not shown), suggesting that each is predominantly monomeric over the concentration range used.⁵

⁴ A. Patel, V. E. Vought, V. Dharmarajan, and M. S. Cosgrove, manuscript in preparation.

⁵ It was previously shown by yeast two-hybrid and GST pulldown experiments that MLL³⁷⁴⁵ self-associates (52). However, we could not detect self-association MLL³⁷⁴⁵ in this investigation by sedimentation velocity analysis. This is probably because the limited solubility of MLL³⁷⁴⁵ in the absence of interacting proteins prevents performing sedimentation velocity experiments at a high enough concentration to observe significant self-association. In support of this, control sedimentation equilibrium experiments show that MLL³⁷⁴⁵ weakly self-associates with a *K*_d of 12.7 μ M (A. Patel and M. S. Cosgrove, unpublished data).

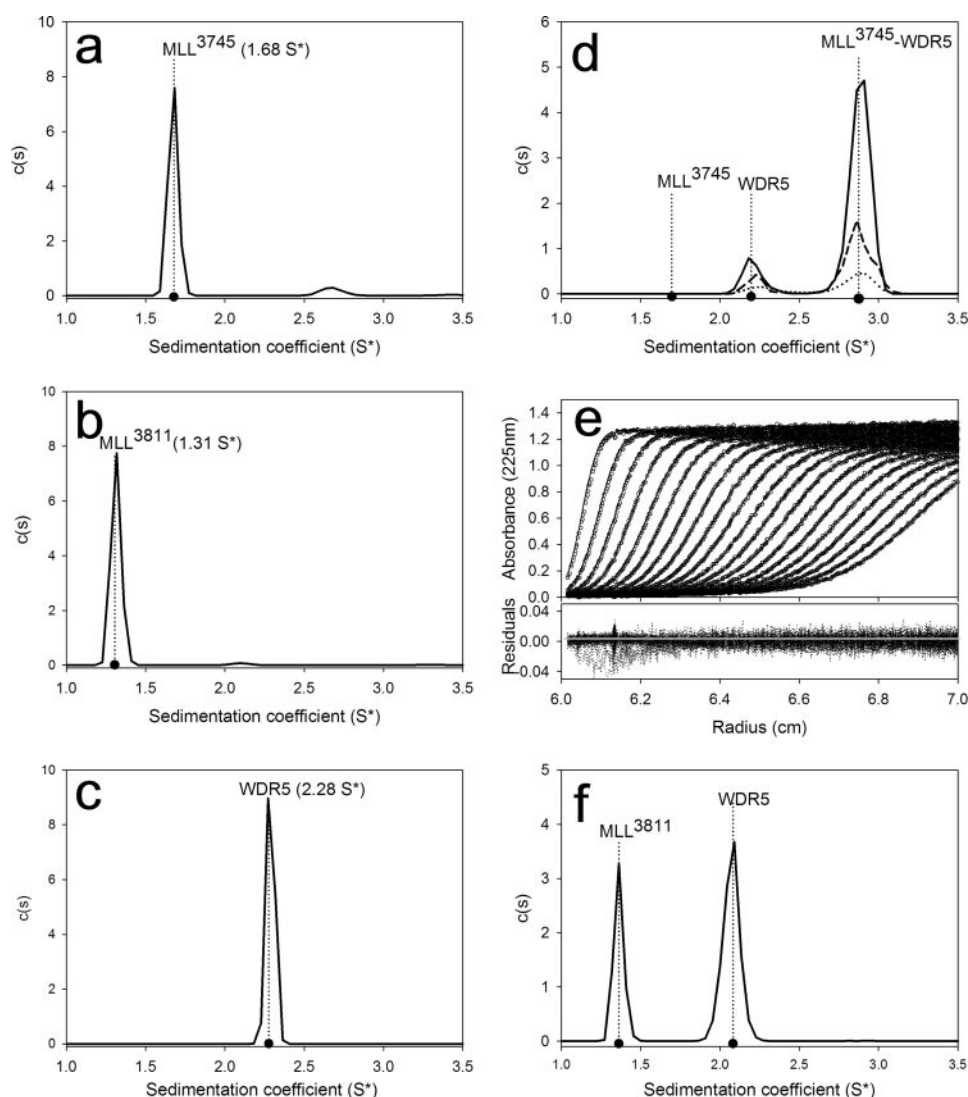


FIGURE 2. Sedimentation velocity analyses of WDR5 and recombinant MLL1 constructs. *a–c*, diffusion-free sedimentation coefficient distributions $c(s)$ derived from sedimentation velocity data of MLL³⁷⁴⁵ (*a*), MLL³⁸¹¹ (*b*), and WDR5 (*c*). Each protein was measured at a concentration of 15 μM . *d*, $c(s)$ distribution of sedimentation velocity data of MLL³⁷⁴⁵-WDR5 complex at the following concentrations: 7 μM (solid black line), 3.5 μM (dashed line), and 1.5 μM (dotted line). *e*, upper panel, direct fitting (solid black line) of sedimentation velocity boundary data (open circles) using a hetero-association model ($A + B \leftrightarrow AB$) in the program SEDPHAT. Lower panel, residuals are plotted as a function of radius. *f*, $c(s)$ distribution from the sedimentation velocity analysis of MLL³⁸¹¹ and WDR5 each at a concentration of 7 μM .

TABLE 1

Summary of sedimentation coefficients derived from sedimentation velocity analyses

Protein	s^* ^a	$s_{20,w}$ ^b	Calculated molecular mass	Expected molecular mass
			kDa	kDa
MLL ³⁷⁴⁵	1.68 ± 0.03	1.76	26.1 ± 0.47	26.07
MLL ³⁸¹¹	1.31 ± 0.02	1.38	18.1 ± 0.28	18.2
WDR5	2.28 ± 0.04	2.41	37.1 ± 0.65	36.5
MLL ³⁷⁴⁵ -WDR5	2.87 ± 0.02	3.08	62.0 ± 0.43	62.6
WDR5-RbBP5-Ash2L	3.98 ± 0.04	4.25	150.0 ± 1.50	152.9
MLL ³⁷⁴⁵ -WDR5-RbBP5-Ash2L	5.01 ± 0.03	5.36	180.0 ± 1.08	179.0

^a Experimental sedimentation coefficients (s^*) \pm standard error of measurement determine from three independent experiments.

^b Standard sedimentation coefficient ($s_{20,w}$) after correcting for water at 20 $^{\circ}\text{C}$.

When run individually, MLL³⁷⁴⁵, MLL³⁸¹¹, and WDR5 sediment with s^* values of 1.68 ± 0.03 ($1.76 s_{20,w}$), 1.31 ± 0.02 ($1.38 s_{20,w}$), and 2.28 ± 0.05 ($2.41 s_{20,w}$), respectively. The molecular

masses calculated from these $s_{20,w}$ values are within error ($<3\%$), similar to that of the expected masses for each protein (Table 1).

When WDR5 is mixed with MLL³⁷⁴⁵ at a 1:1 stoichiometric ratio, a complex is formed that sediments with an s^* value of 2.87 ($3.08 s_{20,w}$) (Fig. 2*d*). The calculated molecular mass of this complex is 62 kDa, which is within error similar to the expected molecular mass for a 1:1 complex between WDR5 and MLL³⁷⁴⁵ (Table 1). Varying the concentration of the complex at equimolar concentrations shows a relatively constant peak position (Fig. 2*d*), suggesting that the complex is stable on the time scale of sedimentation (45). Global fitting of the data using a hetero-association model ($A + B \leftrightarrow AB$) in the program SEDPHAT (46) gives a dissociation constant (K_d) of $0.12 \pm 0.02 \mu\text{M}$ (\pm standard error from three independent experiments) with a k_{off} value of $1.2 \times 10^{-5} \text{ s}^{-1}$ (Fig. 2*e*, Table 2). These data indicate that WDR5 forms a high affinity complex with MLL³⁷⁴⁵ that is stable on the time scale of sedimentation.

To further map the interaction between WDR5 and MLL1, sedimentation velocity experiments were conducted with an equimolar mixture of WDR5 and MLL³⁸¹¹, which lacks the N-terminal N-SET region (Fig. 1*a*). As shown in Fig. 2*f*, the $c(s)$ distribution suggests that WDR5 and MLL³⁸¹¹ sediment as non-interacting species with derived molecular masses consistent

with the monomeric forms of each protein. The same result is observed when the experiment is conducted with a 1:5 stoichiometric excess of WDR5 (not shown). These results indicate that the 66-residue N-SET region of MLL1 is required for the interaction of WDR5 with MLL1.

WDR5 Recognizes a Conserved Arginine-containing Sequence in the N-SET Region of MLL1—To identify amino acid residues in the N-SET region that may be involved in the interaction between MLL1 and WDR5, we performed a sequence alignment of mammalian SET1 family members that are known to interact with WDR5 (Fig. 3 and supplemental Fig. 1) (28, 32, 33, 38, 39, 50). Although the N-SET region is much less conserved than the catalytic SET and post-SET domains, a conserved 6-residue sequence containing the MLL1 residues GSARAE stands out in the N-SET region. To determine whether these residues are important for the interaction of WDR5 with MLL1,

TABLE 2

Summary of sedimentation velocity experiments with wild-type and mutant MLL³⁷⁴⁵ and WDR5 proteins

Protein	$s^*, s_{20,w}$	f/fo	$s_{20,w}$ of complex with WDR5	$K_{d, \mu M}$ A+B \rightleftharpoons AB
MLL³⁷⁴⁵				
Wild type	1.7, 1.8	1.72	3.08	0.12
S3763A	1.7, 1.8	1.71	3.02	0.27
R3765A	1.7, 1.8	1.70	NI ^a	NI
E3767A	1.7, 1.8	1.71	3.00	0.41
WDR5				
Wild type	2.3, 2.4	1.51	3.08	0.12
S91K	2.1, 2.3	1.61	NI	NI
F133A	2.2, 2.4	1.53	NI	NI
Y191F				
Monomer	2.3, 2.4	1.49	3.04	0.35
Dimer	2.9, 3.1	1.89	NI	NI

^a NI = no interaction observed.

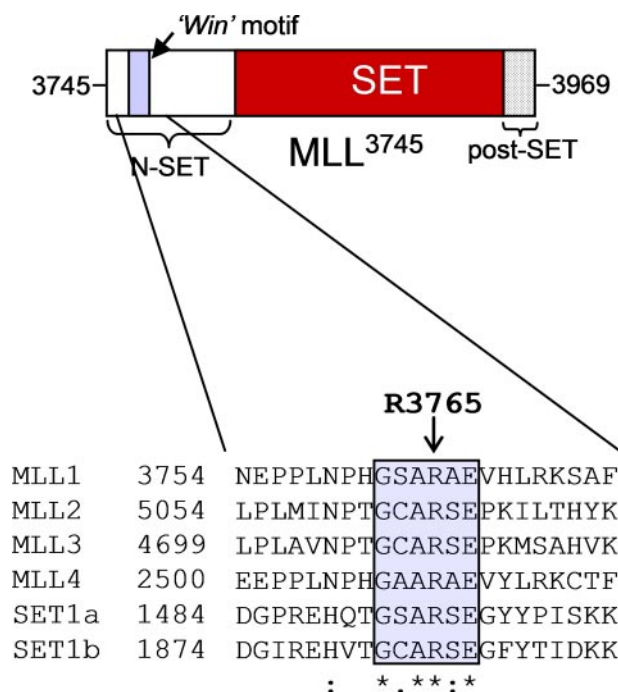


FIGURE 3. ClustalW2 multiple sequence alignment (60) of the C-terminal SET domains of human SET1 family members MLL1 (Swiss Protein data base accession number Q03164), MLL2 (O14606), MLL3 (AAK00583), MLL4 (Q9UMN6), SET1a (O15047), and SET1b (K1AA1076) showing only a subgroup of residues from the N-SET region (for larger alignment, see supplemental Fig. 1). The BLOSUM matrix (61) was used for the alignment. The conserved Win motif in the N-SET region is highlighted in blue. Conserved residues are denoted underneath the alignment by an asterisk, conservative substitutions are denoted by a colon, and semiconservative substitutions are denoted by a period.

we individually replaced Ser-3763, Arg-3765, and Glu-3767 of MLL1 with alanine and compared the wild-type and mutant MLL³⁷⁴⁵ proteins for their ability to interact with WDR5 in sedimentation velocity experiments. The results show that although the S3763A and E3767A mutant proteins only modestly affect the interaction of MLL³⁷⁴⁵ with WDR5 (Fig. 4, *a* and *c*, and Table 2), the replacement of Arg-3765 with alanine abolishes the interaction (Fig. 4*b*). No interaction is detected even when the experiment is conducted at a 1:5 stoichiometric ratio

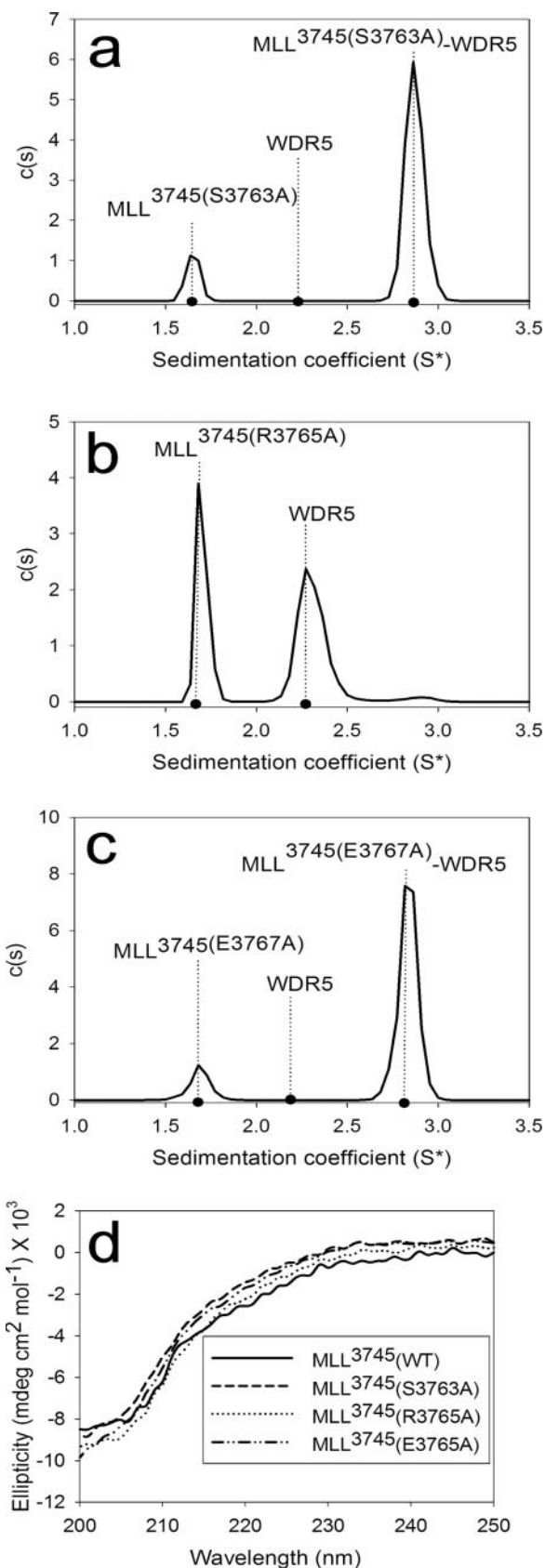


FIGURE 4. Arginine 3765 of MLL1 is critical for the interaction of WDR5 with MLL1. a–c, sedimentation velocity analyses (c(s)) of wild-type (WT) WDR5 with mutant MLL³⁷⁴⁵ constructs: S3763A (a), R3765A (b), and E3767A (c). Each protein was run at a concentration of 7 μM. d, circular dichroism spectra of wild-type and mutant MLL³⁷⁴⁵ constructs at 0.2 mg/ml.

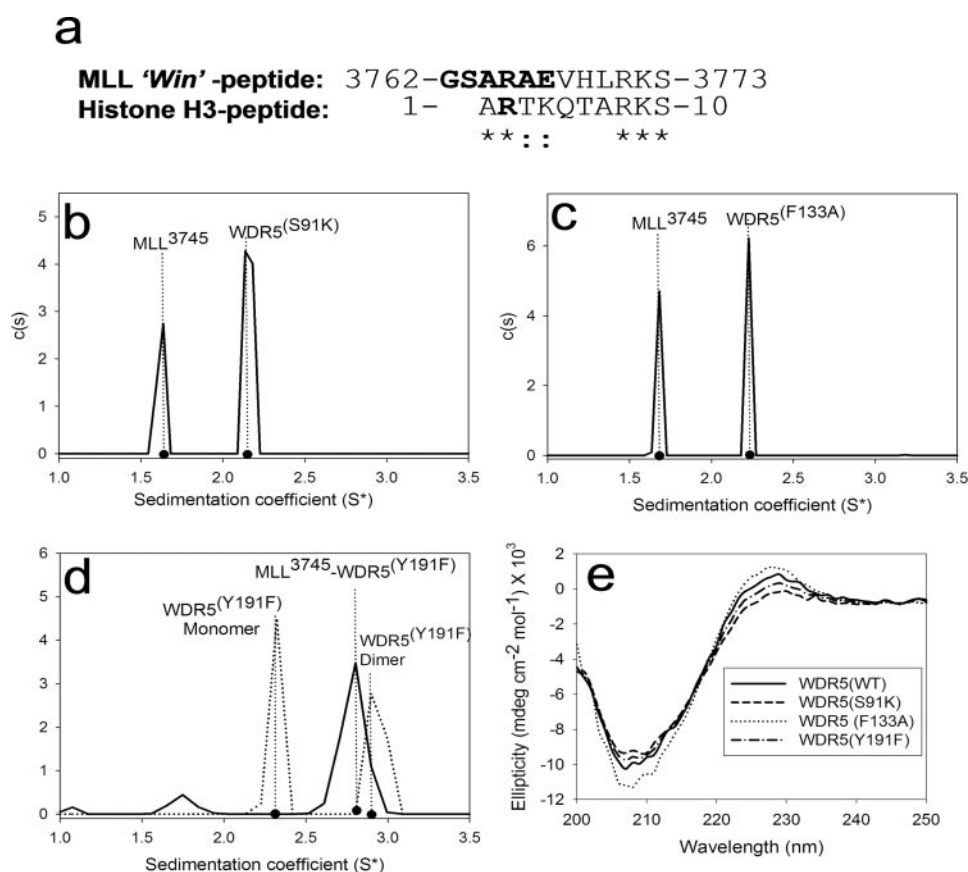


FIGURE 5. The arginine binding cavity of WDR5 is required for the interaction between MLL³⁷⁴⁵ and WDR5. *a*, ClustalW sequence alignment of the human MLL1 (residues 3762–3773) with the human histone H3 N-terminal tail (residues 1–10). Conserved residues are denoted underneath the alignment by an asterisk, conservative substitutions are denoted by a colon, and semiconservative substitutions are denoted by a period. *b–d*, sedimentation velocity analyses (*c*(*s*)) of wild-type (WT) MLL³⁷⁴⁵ with WDR5 mutations: S91K (*b*), F133A (*c*), and Y191F (*d*). The dotted line in *d* represents the *c*(*s*) distribution for the Y191F WDR5 protein alone. The solid line represents the *c*(*s*) distribution when the Y191F protein was run with MLL³⁷⁴⁵. *e*, circular dichroism spectra of wild-type and mutant WDR5 constructs at 0.2 mg/ml.

(data not shown). To determine whether the R3765A mutation disrupts the overall conformation of MLL³⁷⁴⁵, we compared far-UV CD spectra among the wild-type and mutant MLL³⁷⁴⁵ proteins (Fig. 4*d*). The results show that wild-type and mutant protein spectra are essentially superimposable, suggesting that the mutations do not alter the secondary structure of MLL³⁷⁴⁵. In addition, the frictional coefficients (*f*/*f*₀) derived from sedimentation velocity experiments are essentially unchanged between the wild-type and mutant MLL1 proteins (Table 2), suggesting that the mutations do not significantly alter the overall hydrodynamic shape of MLL³⁷⁴⁵. Taken together, these results suggest that Arg-3765 of MLL1 is critical for the interaction with WDR5.

The Arginine Binding Cleft of WDR5 Is Required for Interaction with MLL1—Previous x-ray structural studies of WDR5 bound to histone H3 peptides reveal that arginine 2 of H3 is critical for the interaction with WDR5 (41, 53–55). The side chain of H3R2 inserts into a central cavity of WDR5 and is stabilized by hydrogen bond, hydrophobic, and cation- π interactions. Since the sequence surrounding Arg-3765 of MLL1 is similar to that surrounding Arg-2 of H3 (Fig. 5*a*), we hypothesized that WDR5 interacts with MLL1 by a similar mechanism. To test this hypothesis, we constructed point mutations in

WDR5 and analyzed the mutant proteins for their ability to interact with wild-type MLL³⁷⁴⁵ in sedimentation velocity experiments. It was previously reported that the S91K and F133A WDR5 mutations prevent histone H3 binding to free WDR5 (39, 41, 53). Both Ser-91 and Phe-133 line the surface of the central arginine binding cavity, and the replacement of Ser-91 with lysine is expected to sterically interfere with the insertion of the arginine side chain into the cavity. Phe-133 makes extensive hydrophobic and cation- π interactions with the guanidinium of arginine and was previously shown to be critical for the interaction of histone H3 with WDR5 (41, 53). As expected, both the S91K and the F133A WDR5 proteins are incapable of interacting with MLL³⁷⁴⁵ in sedimentation velocity experiments (Fig. 5, *b* and *c*). This occurs without significant changes in the frictional coefficient or CD spectra of either mutant protein (Fig. 5*e*, and Table 2), suggesting that the overall conformation of the S91K and F133A mutant proteins is similar to that of wild-type WDR5. These results suggest that the central arginine binding cavity of WDR5 is essential for the interaction with MLL1.

Previous results suggest that histone H3 binding to WDR5 is significantly reduced by the Y191F mutation in WDR5 (41). This is because Tyr-191 interacts with Thr-6 in the histone H3-WDR5 crystal structure (Protein Data Bank (PDB) code 2co0) (41, 53–55). To determine whether Tyr-191 of WDR5 is important for the interaction of WDR5 with MLL1, we constructed Y191F WDR5 and determined its ability to interact with MLL³⁷⁴⁵ in sedimentation velocity experiments. Intriguingly, unlike wild-type WDR5, which exists as a monomer in solution, the replacement of Tyr-191 with phenylalanine results in significant self-association, as the sedimentation velocity profile in the absence of MLL³⁷⁴⁵ shows two peaks characteristic of a mono-dimer equilibrium at 2.32 and 2.9 *s*^{*} (Fig. 5*d*, dotted line). At 0.5 mg/ml, ~50% of WDR5 exists as a monomer and dimer species, respectively. In contrast, at 1 mg/ml, ~70% of WDR5 exists in the dimer form (data not shown), consistent with self-association. However, in the presence of MLL³⁷⁴⁵, the dimer dissociates and forms a 1:1 complex with MLL³⁷⁴⁵ that is similar to that of the wild-type complex at 2.8 *s*^{*} (Fig. 5*d*, solid line). The apparent dissociation constant for the binding of WDR5 to MLL³⁷⁴⁵ is increased 3-fold when Tyr-191 of WDR5 is replaced with phenylalanine (*K_d* = 0.35 μ M, Table 2). These results suggest that Tyr-191 of WDR5 does not

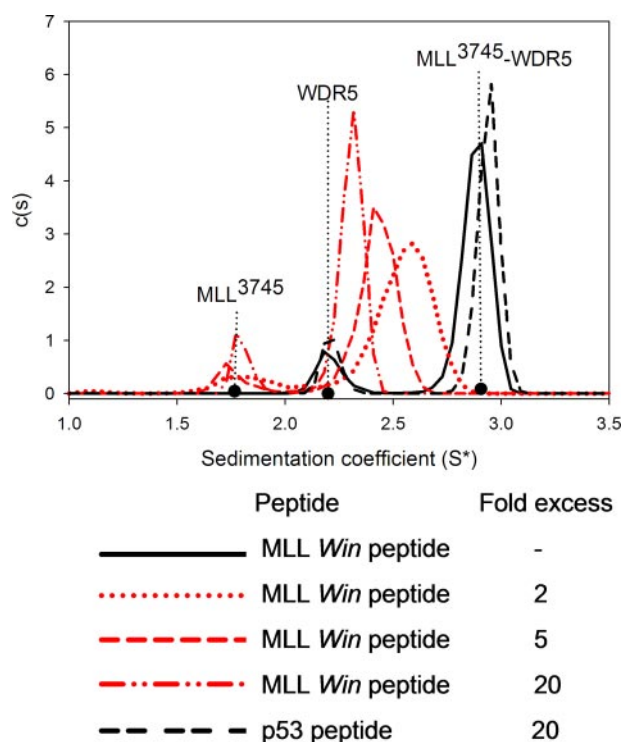


FIGURE 6. **MLL1 Win peptide disrupts MLL³⁷⁴⁵-WDR5 complex.** Diffusion-free sedimentation coefficient distribution ($c(s)$) of the WDR5-MLL³⁷⁴⁵ complex (solid black line) in the presence of the MLL1 Win peptide (red lines) or p53 peptide (³⁶⁵HSSHLKSKKGQSTSRHKK³⁸²) (dashed black line) are shown. The concentration of complex in each experiment was 7 μ M.

significantly contribute to the interaction with MLL³⁷⁴⁵ and that the modest increase in the dissociation constant for the complex is likely due to competition between self-association and MLL³⁷⁴⁵ binding near the arginine binding cavity of WDR5.

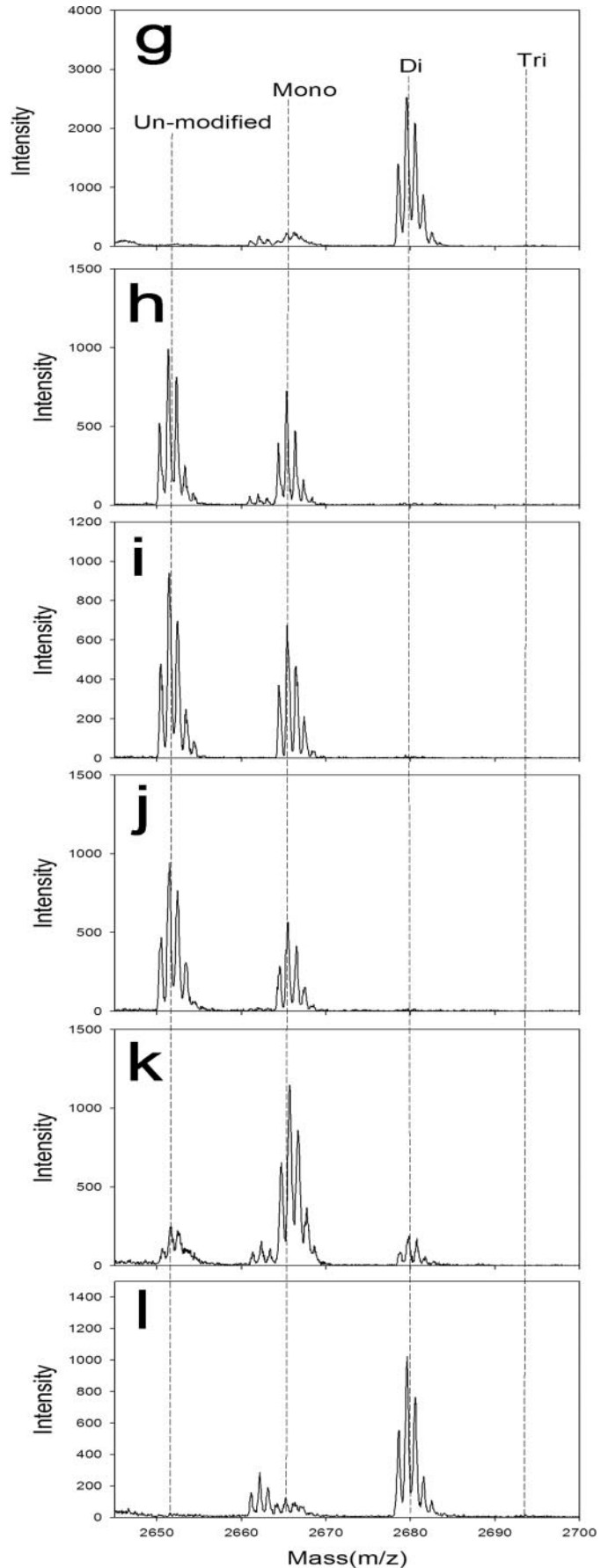
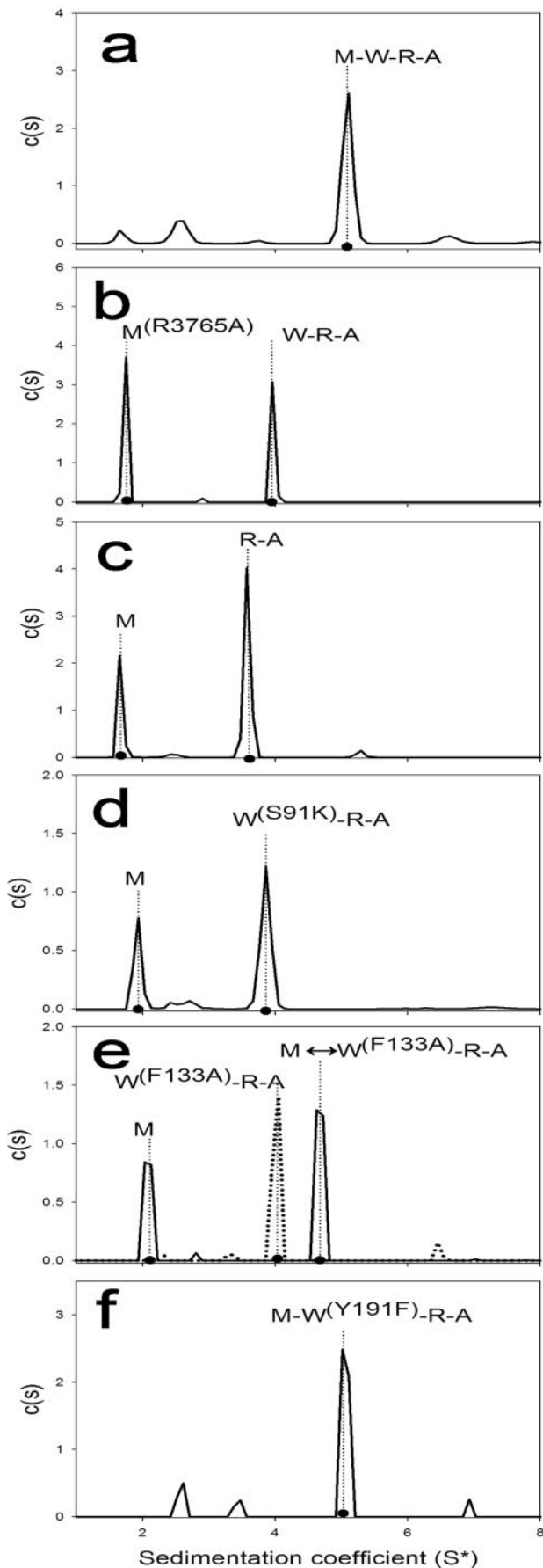
Peptide Derived from MLL1 Win Motif Disrupts WDR5-MLL³⁷⁴⁵ Complex—The results presented above suggest that a peptide containing the MLL1 Win motif should inhibit the formation of the complex between MLL³⁷⁴⁵ and WDR5. To test this hypothesis, we compared sedimentation velocity profiles of the WDR5-MLL³⁷⁴⁵ complex in the presence of increasing amounts of a 12-residue synthetic peptide derived from the MLL1 Win motif. The MLL1 peptide consists of residues 3762–3773 and contains the conserved GSARAE motif plus an additional 6 residues (called the MLL1 Win peptide). This peptide was synthesized with acetyl- and amide-capping groups on the N and C termini, respectively, to eliminate the effects of unnatural N- and C-terminal charges on the binding. As a control, we performed the same experiments using an arginine-containing peptide from the C terminus of the human p53 tumor suppressor (56). The results show that even at a 2-fold molar excess of the MLL1 Win peptide, the WDR5-MLL³⁷⁴⁵ complex no longer sediments as a stable complex, as its sedimentation peak is broadened and shifted to 2.58 s^* (Fig. 6). It is likely that this peak represents the equilibrium distribution between free WDR5 and the MLL³⁷⁴⁵-WDR5 complex, which cannot be resolved within signal-to-noise ratio of the data. Consistent with this, increasing the MLL1 Win peptide concentration causes the complex to increasingly dissociate with sedimentation values

for each protein approaching that expected for non-interacting species. In addition, a peak corresponding to free MLL³⁷⁴⁵ appears at 1.7 s^* , the concentration of which also increases with increasing MLL1 Win peptide. When the same experiments were conducted with a 20-fold molar excess of the control p53-peptide, the WDR5-MLL³⁷⁴⁵ complex is unaffected (Fig. 6). These results are consistent with a model in which MLL1 Win peptide specifically competes with MLL1 for the central arginine binding groove of WDR5.

Recognition of Arg-3765 of MLL1 Is Essential for the Assembly and Enzymatic Activity of the MLL1 Core Complex—Previous co-immunoprecipitation experiments indicate that WDR5 functions to bridge the interaction between the catalytic subunit of MLL1 and the non-catalytic regulatory components of the MLL1 core complex: namely RbBP5 and Ash2L (39). Given that Arg-3765 of MLL1 is essential for the interaction between MLL³⁷⁴⁵ and WDR5, we hypothesized that the same residue will be critical for the assembly and activity of the full MLL1 core complex. To test this hypothesis, we compared wild-type and R3765A MLL³⁷⁴⁵ proteins for their ability to assemble into a complex with WDR5, RbBP5, and Ash2L using purified recombinant components. Sedimentation velocity analytical ultracentrifugation was used to monitor complex assembly. In addition, we used MALDI-TOF mass spectrometry to measure methylation of a histone H3 peptide (amino acids 1–20).

We show in a separate investigation that an enzymatically active MLL1 core complex can be readily assembled with wild-type MLL³⁷⁴⁵, WDR5, RbBP5, and Ash2L.⁴ Sedimentation velocity studies show that these components assemble into a 180-kDa complex with an s^* value of 5.01 (5.36 $s_{20,w}$) and with 1:1:1:1 stoichiometry. Furthermore, enzymatic assays reveal that wild-type MLL³⁷⁴⁵ is a slow monomethyltransferase in the absence of interacting proteins but becomes a fast dimethyltransferase in the presence of WDR5, RbBP5, and Ash2L.⁴ However, in the current investigation, when the core components were mixed with the R3765A MLL³⁷⁴⁵ protein, the 5 s^* complex was not detected in sedimentation velocity experiments (compare Fig. 7, *a* and *b*). Instead, we observed two separate peaks that correspond to free MLL³⁷⁴⁵ at 1.7 s^* and the WDR5-RbBP5-Ash2L (W-R-A) subcomplex at 4 s^* , which is where the subcomplex sediments in the absence of MLL1 (Table 1). Consequently, the H3K4 dimethylation activity of the wild-type MLL1 core complex is lost when Arg-3765 of MLL1 is replaced with alanine (compare Fig. 7, *g* and *h*). Instead, only monomethylation of histone H3 is observed. A similar result was observed when we attempted to assemble the complex in the absence of WDR5 (Fig. 7, *c* and *i*). These results suggest that WDR5's recognition of Arg-3765 of MLL1 is critical for the assembly and H3K4 dimethylation activity of the MLL1 core complex.

To further test this hypothesis, we reasoned that mutations in the arginine binding pocket of WDR5 should also disrupt the assembly of the MLL1 core complex *in vitro* and diminish the H3K4 dimethylation activity. We mixed wild-type MLL³⁷⁴⁵, RbBP5, and Ash2L proteins with the S91K and F133A WDR5 proteins and compared sedimentation velocity profiles and enzymatic activity with those of the wild-type core complex. As a control, we performed the same experiments with a WDR5 mutant protein where Tyr-191 was replaced with phenylala-



WRD5 Recognizes Arg-3765 of MLL1

nine. Tyr-191 resides outside the arginine binding pocket of WDR5 and has been suggested to interact with histone H3 (41).

Sedimentation velocity results show that both the S91K and the F133A WDR5 proteins are capable of forming the 4 *s** WDR5-RbBP5-Ash2L subcomplex in solution (Fig. 7, *d* and *e*, respectively). However, the interaction of MLL³⁷⁴⁵ with these subcomplexes is either abolished or significantly impaired. When Ser-91 of WDR5 is replaced with lysine, MLL³⁷⁴⁵ and the WDR5^{S91K}-RbBP5-Ash2L subcomplex sediment as two non-interacting species in solution, with *s** values corresponding to free MLL³⁷⁴⁵ and the unbound WDR5^{S91K}-RbBP5-Ash2L subcomplex at 4 *s** (Fig. 7*d*). As observed with the R3765A mutation of MLL1, the dimethylation activity of the MLL1 core complex is lost when assembled with the S91K WDR5 protein (Fig. 7*j*).

Similarly, when the complex is assembled with F133A WDR5, two peaks are observed but with shifted *s** values at 2.02 and 4.63 (Fig. 7*e*, *solid black line*). The *s** value at 2.02 is slightly larger than that of free MLL³⁷⁴⁵, whereas the peak at 4.63 is between that expected for the 4 *s** WDR5-RbBP5-Ash2L subcomplex and the 5 *s** MLL1 core complex. It is likely that the single peak at 4.63 *s** represents an equilibrium mixture of the 4 and 5 *s** species that cannot be resolved within the signal-to-noise ratio of the data. In support of this, the WDR5^{F133A}-RbBP5-Ash2L subcomplex in the absence of MLL³⁷⁴⁵ sediments with an *s** value similar to that of the wild-type subcomplex at 4 *s** (Fig. 7*e*, *dotted line*). These results are consistent with a system in which the dissociation rate constant is rapid when compared with the time scale of sedimentation ($k_{\text{off}} \sim 0.01 \text{ s}^{-1}$) (45). These results suggest that the interaction between MLL1 and WDR5 within the MLL1 core complex is significantly weakened when Phe-133 of WDR5 is replaced with alanine. Consistent with a weak interaction, the H3K4 dimethylation activity of wild-type MLL1 core complex is diminished by $\sim 80\%$ when Phe-133 of WDR5 is replaced with alanine (Fig. 7*k*). These results suggest that there is a correlation between MLL1's interaction with the WDR5 component of the MLL1 core complex and its ability to catalyze H3K4 dimethylation.

In contrast, when the complex was assembled with the Y191F mutant WDR5 protein, no changes were detected in the sedimentation behavior or enzymatic activity of the MLL1 core complex when compared with that of the wild-type complex (Fig. 7, *f* and *l*). Taken together, these results demonstrate the critical nature of the interaction between the arginine binding pocket of WDR5 and Arg-3765 of the MLL1 *Win* motif for the assembly and dimethylation activity of the MLL1 core complex *in vitro*.

The MLL1 *Win* Peptide Specifically Inhibits H3K4 Dimethylation Activity of the MLL1 Core Complex—As shown above, a peptide derived from the MLL1 *Win* motif specifically com-

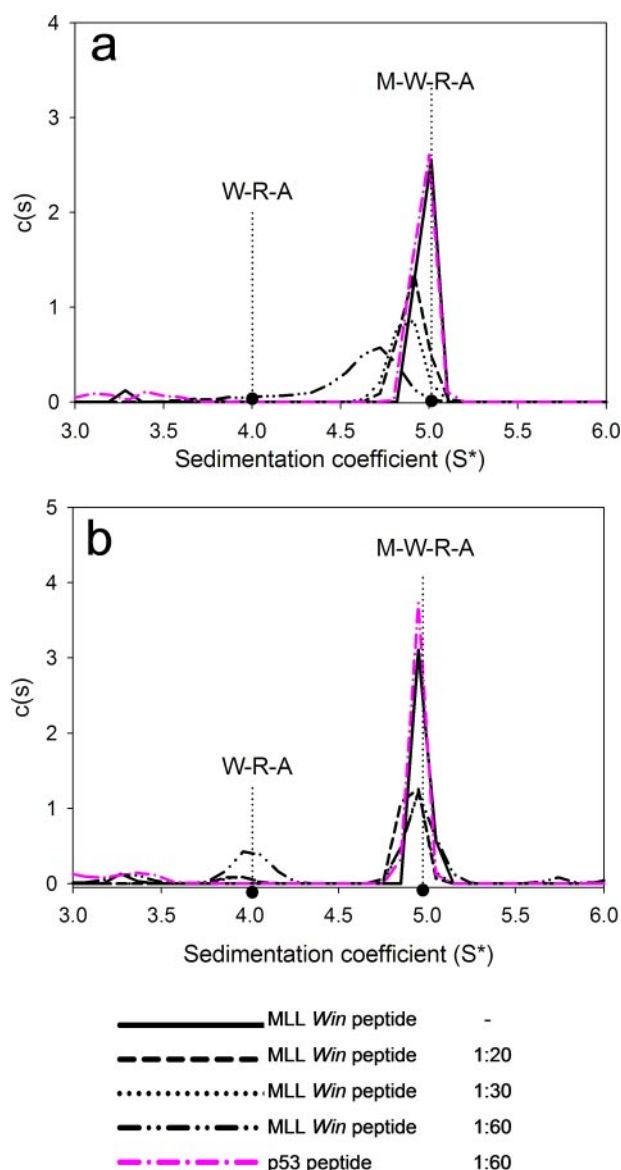


FIGURE 8. The MLL1 *Win* motif peptide inhibits the assembly of the MLL1 core complex. *a*, *c*(*s*) distributions of the MLL1 core complex with increasing concentrations of the MLL1 *Win* motif peptide (black lines) or p53 peptide (pink line). The data were fit with maximum entropy regularization with uniform prior expectations. *b*, Bayesian analysis of *c*(*s*) distributions in *a* using the 4 and 5 *s** peaks as prior expectations for the subcomplex and holocomplex, respectively. The line patterns and colors are presented as in *a*.

petes with MLL1 for binding to WDR5 (Fig. 6). To determine whether the same peptide can inhibit the assembly and activity of the MLL1 core complex *in vitro*, we compared sedimentation velocity profiles and enzymatic activity of the MLL1 core complex in the presence and absence of increasing amounts of the MLL1 *Win* peptide. As above, we used the arginine-containing p53 peptide as a control. As shown in Fig. 8*a*, the addition of a

FIGURE 7. Recognition of Arg-3765 of MLL1 by WDR5 is essential for the assembly and activity of the MLL1 core complex. *a–f*, diffusion-free sedimentation coefficient distributions (*c*(*s*)) for the MLL1 core complexes assembled with wild-type or mutant MLL³⁷⁴⁵, WDR5, RbBP5, and Ash2L proteins; denoted as M, W, R, and A, respectively. *g–l*, MALDI-TOF mass spectrometry of enzymatic reactions after 24 h. Each MALDI-TOF spectra corresponds to the enzymatic activity of the complex displayed in the *c*(*s*) panel on the left (*a–f*). The complexes were assembled with wild-type components (*a* and *g*); with R3765A MLL1 (*b* and *h*); without WDR5 (*c* and *i*); with S91K WDR5 (*d* and *j*); with F133A WDR5 (*e* and *k*); or with Y191F WDR5 (*f* and *l*). In panel *e*, the *solid black line* represents the *c*(*s*) profile of M \leftrightarrow W^{F133A}-R-A, whereas the *dotted line* shows the *c*(*s*) profile for the same proteins but without MLL³⁷⁴⁵. *Mono*, monomethylation; *Di*, dimethylation; *Tri*, trimethylation.

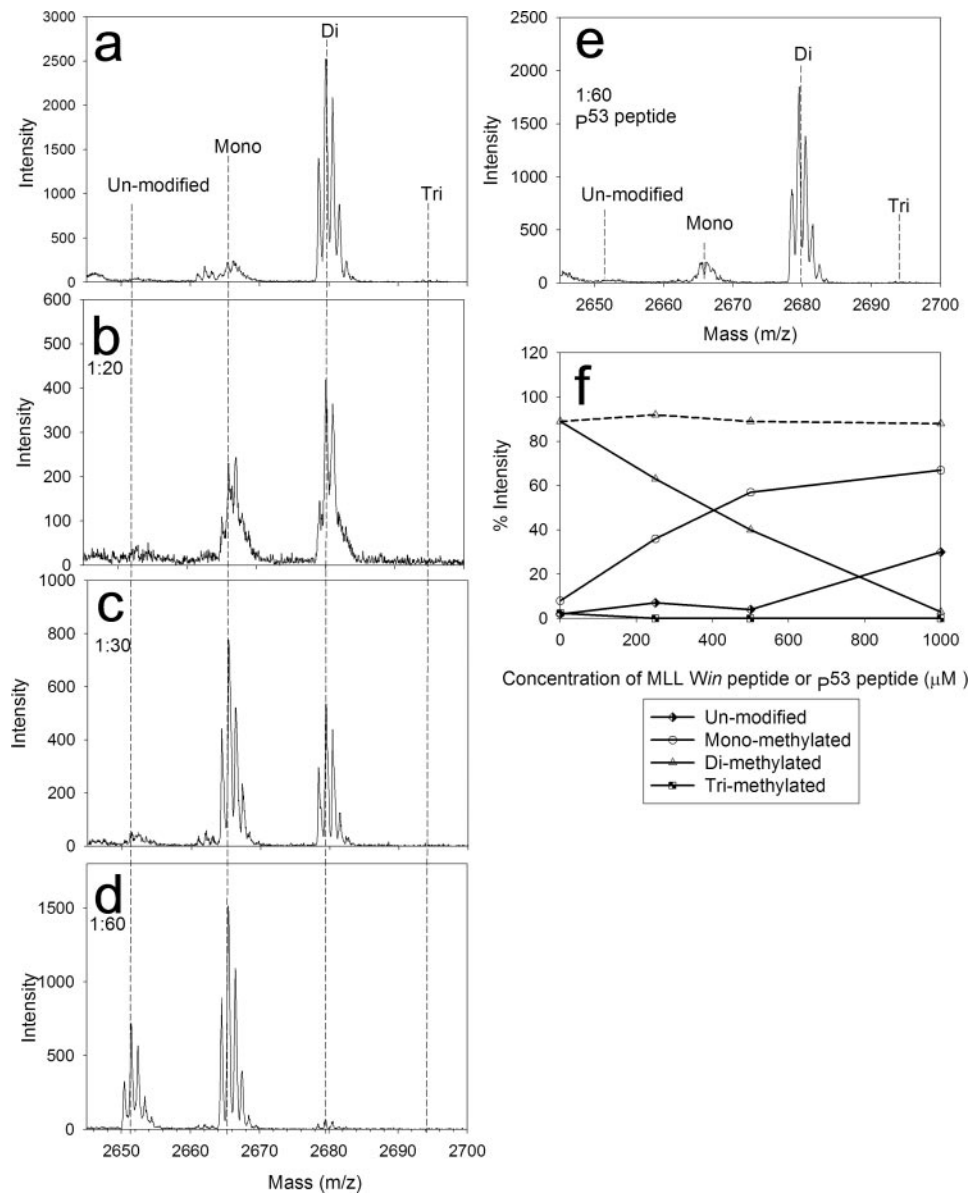


FIGURE 9. The MLL1 *Win* motif peptide inhibits the H3K4 dimethylation activity of the MLL1 core complex. *a–e*, MALDI-TOF mass spectrometry of enzymatic activity catalyzed by the wild-type MLL1 core complex in the absence (*a*) or presence (*b–d*) of a 20-fold excess (*b*), 30-fold excess (*c*), or 60-fold stoichiometric excess (*d*) of MLL1 *Win* peptide. *Mono*, monomethylation; *Di*, dimethylation; *Tri*, trimethylation. *e*, MALDI-TOF mass spectrometry of enzymatic assay catalyzed by wild-type MLL1 core complex in the presence of a 60-fold molar excess of the p53 peptide. *f*, summary of unmodified, mono-, di-, and trimethylation activity of the MLL1 core complex as a function of MLL1 *Win* peptide concentration (solid black lines) or p53 peptide concentration (dashed black line).

20-, 30-, and 60-fold stoichiometric excess of the MLL1 *Win* peptide causes the 5 *s*^{*} sedimentation peak of the MLL1 core complex to increasingly broaden and shift to lower sedimentation values. For example, at a 60-fold excess of the MLL1 *Win* peptide, the sedimentation profile of the complex is significantly broadened with a peak at 4.6 *s*^{*} (Fig. 8*a*), similar to that observed with complexes assembled with the F133A WRD5 protein (Fig. 7*e*). This is consistent with a model in which the MLL1 *Win* peptide competes with MLL1 for binding to the WRD5 component of the WRD5-RbBP5-Ash2L subcomplex, producing a broad c(s) peak intermediate between the 4 *s*^{*} subcomplex and the 5 *s*^{*} holocomplex that is not resolvable within the signal-to-noise ratio of the data. To test this hypothesis,

we next examined the enzymatic activity of the MLL1 core complex in the presence of increasing amounts of the MLL1 *Win* peptide using MALDI-TOF mass spectrometry. The results show that the H3K4 dimethylation activity of the MLL1 core complex is increasingly inhibited in a manner that is closely dependent on peptide concentration (Fig. 9). For example, with a 20-fold molar excess of the MLL1 *Win* peptide, the relative amount of dimethylation catalyzed by the MLL1 core complex decreases by 31% after 24 h (compare Fig. 9, *a* and *b*). The amount of H3K4 dimethylation decreases by 55% in the presence of a 30-fold excess (Fig. 9*c*) and by 97% in the presence of a 60-fold molar excess of the MLL1 *Win* peptide (Fig. 9*d*). As a consequence, the monomethylation activity of the complex is correspondingly increased (summarized in Fig. 9*f*). Because of the effect of the MLL1 *Win* peptide on the sedimentation behavior of the MLL1 core complex (Fig. 8) and because the

instead of fitting the data with a maximum entropy regularization with uniform prior knowledge, we reanalyzed the data, incorporating prior knowledge derived from the *s* values of the subcomplex and holocomplex determined from independent sedimentation experiments using the Bayesian implementation in SEDFIT (57). The resulting sedimentation coefficient distribution shows that the two species can be resolved with amplitudes that depend on the MLL1 *Win* peptide concentration (Fig. 8*b*). These results suggest that increasing amounts of the MLL1 *Win* peptide increasingly shifts the equilibrium of the complex toward dissociation of MLL1³⁷⁴⁵. We note that the concentration of *Win* peptide required to observe this effect is significantly higher than that required to dissociate the MLL1³⁷⁴⁵-WRD5 complex in solution (Fig. 6), suggesting that RbBP5 and Ash2L contribute additional interactions with MLL1 that stabilize the core complex.

In contrast, when the same experiments were conducted with the p53 peptide, we did not observe a significant change in the sedimentation behavior of the MLL1 core complex, even with a 60-fold excess of the p53 peptide (Fig. 8, *a* and *b*, pink line). These results suggest that the MLL1 *Win* peptide specifically inhibits the formation of the MLL1 core complex.

We next examined the enzymatic activity of the MLL1 core complex in the presence of increasing amounts of the MLL1 *Win* peptide using MALDI-TOF mass spectrometry.

WDR5 Recognizes Arg-3765 of MLL1

MLL1 *Win* peptide does not inhibit the monomethylation activity of MLL³⁷⁴⁵ in the absence of interacting proteins (not shown), it is likely that the MLL1 *Win* peptide inhibits dimethylation of H3K4 by preventing the association of MLL1 with the WDR5 component of the WDR5-RbBP5-Ash2L subcomplex, which is required for H3K4 dimethylation.

In contrast, when the same experiments were conducted with the p53 peptide, no change in the H3K4 dimethylation activity of the MLL1 core complex was observed, even with a 60-fold excess of peptide (Fig. 9e). These results are summarized in Fig. 9f, where we plotted the percent intensity of unmethylated and methylated histone peptides as a function of MLL1 *Win* peptide (solid lines) or p53 peptide (dashed line) concentrations. These results show that the concentration of the MLL1 *Win* peptide that is required to inhibit 50% of the H3K4 dimethylation activity of the MLL1 core complex (IC₅₀) is ~400 μ M.

DISCUSSION

WDR5 is a common component of complexes containing members of the SET1 family of H3K4 methyltransferases. Previous research suggests that WDR5 functions as a structural platform that bridges the interaction between the catalytic and regulatory subunits of the MLL1 core complex (39, 40). In particular, it has been proposed that WDR5, RbBP5, and Ash2L form a subcomplex that interacts interchangeably with the different members of the SET1 family of histone methyltransferases, which is required for the regulation of mono-, di-, and trimethylation of histone H3 at Lys-4 (39). The molecular mechanisms by which the WDR5-RbBP5-Ash2L structural platform recognizes different catalytic subunits of the SET1 family are unknown. Here we describe the identification of a conserved arginine-containing motif, called the *Win* or *WDR5* interaction motif, which is located in the N-SET region of MLL1 and other SET1 family members. Our functional characterization of this motif, along with our crystal structure (Protein Data Bank number 3EG6) described in the accompanying investigation (58), demonstrates that the recognition of Arg-3765 of the MLL1 *Win* motif by WDR5 is critical for the assembly and enzymatic activity of the MLL1 core complex *in vitro*.

These results are particularly surprising given current models for the role of WDR5 in the MLL1 core complex. For example, in addition to its role as a structural platform, recent research suggests that WDR5 also functions as a histone binding domain (39–41, 53–55). The histone effector domain hypothesis suggests that WDR5 functions within the MLL1 core complex to recognize dimethylated H3K4 (40). Alternatively, it has been suggested that WDR5 binds histone H3 and “presents” the H3K4 side chain for further methylation by the catalytic SET domain of MLL1 (39, 41). Both of these models predict that histone H3 and MLL1 bind simultaneously to WDR5 at distinct sites (Fig. 10a).

Unexpectedly, our structural and functional results show that Arg-3765 of MLL1 interacts with the same arginine binding pocket of WDR5 that was previously shown to bind Arg-2 of histone H3 (41, 53–55). This suggests a new model for the role of WDR5 in SET1 family complexes (Fig. 10b)

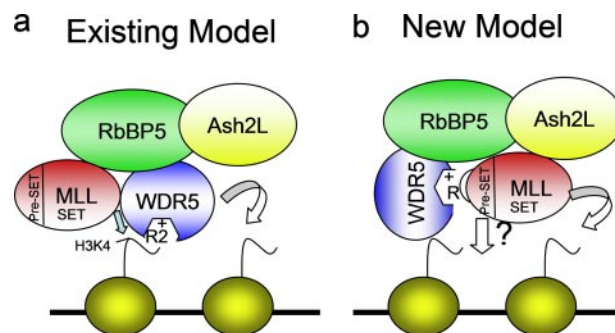


FIGURE 10. Schematic models for the role of WDR5 in the MLL1 core complex. Nucleosomes are shown in yellow with the histone H3 N-terminal tail indicated. *a*, histone binding model where WDR5 recognizes Arg-2 of histone H3 and facilitates H3 methylation by presenting the Lys-4 side chain to the SET domain of MLL1. *b*, a new model based on the present data in which WDR5's recognition of Arg-3765 of the MLL1 *Win* motif is required for the assembly and H3K4 dimethylation activity of the MLL1 core complex.

and also presents a paradox. How does WDR5 simultaneously interact with MLL1 and histone H3 if they bind to the same site on WDR5? The current information adds a new dimension to the problem, which must be accounted for by any model for the role of WDR5 in SET1 family complexes.

The histone binding domain model has arisen in part because of the observation that histone H3 peptides bind to free WDR5, which subsequently led to the suggestion that histone H3 binds to WDR5 that is incorporated into the MLL1 core complex (40). Although a number of investigations have confirmed the binding of histone H3 peptides to free WDR5 (40, 41, 53–55), there is little evidence that supports a direct interaction of histones with WDR5 while bound to MLL1. Evidence cited in support of this hypothesis includes peptide pulldown experiments that suggest that WDR5 is required for the ability to pull down a partially purified MLL2 complex (40). However, it is not clear from these experiments whether the complex is pulled down by virtue of an interaction of the peptides with WDR5 or with the catalytic SET domain of MLL2. Indeed, the latter is a significant possibility, as it was previously shown that histone H3 peptides can efficiently pull down a SET domain fragment derived from the *Drosophila* trithorax protein (59). Therefore, an unambiguous demonstration of a direct interaction of histone peptides with WDR5 while bound to the MLL1 core complex is lacking.

An alternative possibility is that there could be a complex interplay between binding MLL1 and histones during the course of the reaction. However, our results with the R3765A mutation in MLL1 argue against this possibility. This is because if the arginine binding pocket of WDR5 is required for histone binding and presentation within the complex, then an increase in the enzymatic activity of the MLL1 core complex would be expected upon removal of Arg-3765 in MLL1. However, our results clearly show that replacing Arg-3765 with alanine disrupts MLL1's binding to the subcomplex and abolishes the H3K4 dimethylation activity of the MLL1 core complex. This occurs without detectable changes in the secondary structure, hydrodynamic properties, or H3K4 monomethylation activity of the R3765A MLL³⁷⁴⁵ enzyme

when compared with that of the wild-type enzyme. In addition, our results suggest that WDR5 preferentially interacts with MLL1, which forms a stable complex in solution with a dissociation constant (K_d) of 120 nM (Fig. 2 and Table 2). This suggests that WDR5 binds to MLL1 with at least ~40-fold greater affinity than that for the binding of free WDR5 to a dimethylated histone H3 peptide in solution, which binds with a K_d of 4.5 μ M as determined by isothermal titration calorimetry (53). Indeed, the interaction between WDR5 and MLL1 is the strongest pairwise interaction among the members of the MLL1 core complex.⁴ Dissociation of WDR5 from MLL1 during the course of the reaction is therefore less likely.

Furthermore, it was previously suggested that WDR5 functions as a histone presenter domain based on binding experiments with the Y191F mutation in WDR5 (41). In a previous crystal structure (PDB number 2hbn), Tyr-191 interacts with Thr-6 of histone H3, and binding experiments suggest that the replacement of Tyr-191 with phenylalanine reduces histone H3 binding to WDR5 by a factor of 10 (41). However, we observed no changes in the dimethylation activity of the MLL1 core complex when assembled with the Y191F WDR5 protein when compared with that of the wild-type complex (Fig. 7l). Furthermore, we observe that the Y191F WDR5 protein by itself significantly dimerizes in solution, whereas the wild-type protein behaves as a monomer. This suggests that decreased histone binding to Y191F WDR5 may be an artifact due to the involvement of the histone H3 binding site in the dimerization. Consistent with this, Y191F WDR5 dimerization is easily displaced by the addition of MLL³⁷⁴⁵, which sediments with Y191F WDR5 as a 1:1 complex, albeit with a modestly increased dissociation constant (Table 2). It is likely that MLL³⁷⁴⁵ binding to Y191F WDR5 outcompetes the weaker self-association of the Y191F mutant when present at a 1:1 stoichiometry.

Taken together, these results argue against a role of WDR5 as a histone presenter domain, at least while WDR5 is bound to MLL1. However, they do not rule out a possible role for WDR5 in binding histone H3 in the absence of MLL1. This possibility deserves further exploration.

In contrast, several previously published observations support an interaction of MLL1 with the arginine binding pocket of WDR5. Using a recombinant 180-kDa C-terminal fragment of MLL1 (MLL-C), which contains the catalytic SET domain, Dou *et al.* (39) demonstrated that WDR5 is important for the assembly of the minimal MLL1 core complex, which includes MLL-C, WDR5, RbBP5, and Ash2L. Specifically, they showed that MLL-C's interaction with the WDR5-RbBP5-Ash2L subcomplex is abolished in the absence of WDR5. They also showed evidence that the arginine binding pocket of WDR5 is critical for the interaction of MLL1 with the WDR5-RbBP5-Ash2L subcomplex. They showed that the S91K and F133A mutations in WDR5, both of which reside in the arginine binding pocket, disrupt MLL-C's co-immunopurification with the rest of the complex (39). Although it was previously suggested that the loss of MLL-C from these complexes was due to local conformational changes that prevent MLL-C binding to a distinct site on WDR5 (39), our structural and functional results are not consistent with this hypothesis. In particular, we show that the frictional coefficients of S91K and F133A WDR5 proteins in

sedimentation velocity experiments are nearly identical to that of wild-type WDR5 (Table 2), suggesting that each protein has a similar overall hydrodynamic shape. Furthermore, CD spectra for the wild-type and mutant WDR5 proteins are highly similar (Fig. 5d), consistent with similar secondary structure content. We also show that the S91K and F133A WDR5 proteins are each capable of forming the 4S* WDR5-RbBP5-Ash2L subcomplex *in vitro* (Fig. 7, d and e). These results suggest that the overall conformation of the S91K and F133A WDR5 proteins is similar to that of wild-type WDR5 and that the loss of MLL-C in co-immunopurification experiments is likely due to the inability of the mutant WDR5 proteins to recognize Arg-3765 of MLL1.

This is supported by the results of this investigation. We show that the replacement of Arg-3765 with alanine prevents the interaction of MLL³⁷⁴⁵ with WDR5 (Fig. 4b), as well as with the WDR5-RbBP5-Ash2L subcomplex (Fig. 7b). The functional significance of Arg-3765 of MLL1 is demonstrated by the loss of H3K4 dimethylation activity of the MLL1 core complex when assembled with the R3765A mutant MLL³⁷⁴⁵ protein (Fig. 7h) or with the S91K mutant WDR5 protein (Fig. 7, d and j). A similar result was observed when we attempted to assemble the MLL1 core complex in the absence of WDR5 (Fig. 7, c and i).

Intriguingly, our results suggest that there is a correlation between the strength of interaction between MLL³⁷⁴⁵ and WDR5 and the ability of the MLL1 core complex to catalyze dimethylation of H3K4. For example, when Phe-133 of WDR5 was replaced with alanine, the interaction of MLL1 with the subcomplex is significantly reduced, and the dimethylation activity of the MLL1 core complex is correspondingly reduced (Fig. 7, e and k). Furthermore, similar results were observed using the MLL1 *Win* motif peptide, which disrupts the interaction of MLL³⁷⁴⁵ with WDR5 (Fig. 6) and results in a similar loss of H3K4 dimethylation activity by the full MLL1 core complex *in vitro*, the extent to which is dependent on the dose of the peptide (Figs. 8 and 9). In the accompanying investigation (58), our crystal structure (Protein Data Bank number 3EG6) of WDR5 bound to the MLL1 *Win* motif peptide shows that Arg-3765 of the MLL1 *Win* motif indeed binds in the central arginine binding pocket of WDR5. We also show that the MLL1 *Win* peptide binds to WDR5 with high affinity and makes several previously unobserved interactions with conserved residues on WDR5. Since the *Win* motif is conserved among metazoan SET1 family members, it is possible that a similar mode of recognition is involved in the assembly of all metazoan SET1 family complexes.

In summary, these results suggest that WDR5's recognition of Arg-3765 of MLL1 is crucial for bringing together the catalytic and regulatory subunits of the MLL1 core complex, which is required for the H3K4 dimethylation activity of MLL1 *in vitro*. Future studies that probe the molecular mechanisms by which the regulatory components of the subcomplex increase the enzymatic activity of MLL1, both *in vitro* and *in vivo*, will be instrumental in our understanding of how the SET1 family of lysine methyltransferases contributes to regulation of gene expression in eukaryotes.

Acknowledgments—We thank Anthony Ouellette for training and access to the MALDI-TOF mass spectrometer at SUNY Oswego. We thank Stewart Loh for access to a CD spectropolarimeter and Cynthia Wolberger for p53 peptide. We thank Peter Schuck for advice regarding the Bayesian analysis of our data and Dave Skalnik for a critical reading of this manuscript. We are grateful to Dave Allis, Alex Ruthenberg, Robert Roeder, and Yali Dou for plasmids. We thank Benny Howard, Ilona Savich, Gina Duggleby, and Diana Zhang for useful discussions.

REFERENCES

- Jenuwein, T., and Allis, C. D. (2001) *Science* **293**, 1074–1080
- Luger, K. (2003) *Curr. Opin. Genet. Dev.* **13**, 127–135
- Kornberg, R. D., and Lorch, Y. (1999) *Cell* **98**, 285–294
- Cosgrove, M. S., Boeke, J. D., and Wolberger, C. (2004) *Nat. Struct. Mol. Biol.* **11**, 1037–1043
- Cosgrove, M. S., and Wolberger, C. (2005) *Biochem. Cell Biol.* **83**, 468–476
- Cosgrove, M. S. (2007) *Expert Rev. Proteomics* **4**, 465–478
- Peterson, C. L., and Laniel, M. A. (2004) *Curr. Biol.* **14**, R546–R551
- Strahl, B. D., and Allis, C. D. (2000) *Nature* **403**, 41–45
- Turner, B. M. (2000) *BioEssays* **22**, 836–845
- Fischle, W., Wang, Y., and Allis, C. D. (2003) *Nature* **425**, 475–479
- Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. (2007) *Nat. Struct. Mol. Biol.* **14**, 1025–1040
- Ruthenburg, A. J., Li, H., Patel, D. J., and Allis, C. D. (2007) *Nat. Rev. Mol. Cell Biol.* **8**, 983–994
- Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002) *Nature* **419**, 407–411
- Krogan, N. J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M. A., Dean, K., Ryan, O. W., Golshani, A., Johnston, M., Greenblatt, J. F., and Shilatifard, A. (2003) *Mol. Cell* **11**, 721–729
- Ng, H. H., Robert, F., Young, R. A., and Struhl, K. (2003) *Mol. Cell* **11**, 709–719
- Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., III, Gingeras, T. R., Schreiber, S. L., and Lander, E. S. (2005) *Cell* **120**, 169–181
- Schubeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D. E., O'Neill, L. P., Turner, B. M., Dellow, J., Bell, S. P., and Groudine, M. (2004) *Genes Dev.* **18**, 1263–1271
- Sims, R. J., III, Chen, C. F., Santos-Rosa, H., Kouzarides, T., Patel, S. S., and Reinberg, D. (2005) *J. Biol. Chem.* **280**, 41789–41792
- Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D., and Patel, D. J. (2006) *Nature* **442**, 91–95
- Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000) *Nature* **406**, 593–599
- Dillon, S. C., Zhang, X., Trievel, R. C., and Cheng, X. (2005) *Genome Biol.* **6**, 227
- Briggs, S. D., Bryk, M., Strahl, B. D., Cheung, W. L., Davie, J. K., Dent, S. Y., Winston, F., and Allis, C. D. (2001) *Genes Dev.* **15**, 3286–3295
- Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W. W., Wilm, M., Aasland, R., and Stewart, A. F. (2001) *EMBO J.* **20**, 7137–7148
- Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J. F., and Shilatifard, A. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12902–12907
- Nagy, P. L., Griesbeck, J., Kornberg, R. D., and Cleary, M. L. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 90–94
- Hughes, C. M., Rozenblatt-Rosen, O., Milne, T. A., Copeland, T. D., Levine, S. S., Lee, J. C., Hayes, D. N., Shanmugam, K. S., Bhattacharjee, A., Biondi, C. A., Kay, G. F., Hayward, N. K., Hess, J. L., and Meyerson, M. (2004) *Mol. Cell* **13**, 587–597
- Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N., and Herr, W. (2003) *Genes Dev.* **17**, 896–911
- Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D. J., Kitabayashi, I., Herr, W., and Cleary, M. L. (2004) *Mol. Cell Biol.* **24**, 5639–5649
- Milne, T. A., Briggs, S. D., Brock, H. W., Martin, M. E., Gibbs, D., Allis, C. D., and Hess, J. L. (2002) *Mol. Cell* **10**, 1107–1117
- Steward, M. M., Lee, J. S., O'Donovan, A., Wyatt, M., Bernstein, B. E., and Shilatifard, A. (2006) *Nat. Struct. Mol. Biol.* **13**, 852–854
- Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufte-land, K. R., Aasland, R., Anastassiadis, K., Ang, S. L., and Stewart, A. F. (2006) *Development (Camb.)* **133**, 1423–1432
- Lee, J. H., and Skalnik, D. G. (2005) *J. Biol. Chem.* **280**, 41725–41731
- Lee, J. H., Tate, C. M., You, J. S., and Skalnik, D. G. (2007) *J. Biol. Chem.* **282**, 13419–13428
- Prasad, R., Zhadanov, A. B., Sedkov, Y., Bullrich, F., Druck, T., Rallapalli, R., Yano, T., Alder, H., Croce, C. M., Huebner, K., Mazo, A., and Canaani, E. (1997) *Oncogene* **15**, 549–560
- Terranova, R., Agherbi, H., Boned, A., Meresse, S., and Djabali, M. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6629–6634
- Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A., and Korsmeyer, S. J. (1995) *Nature* **378**, 505–508
- Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C. M., and Canaani, E. (2002) *Mol. Cell* **10**, 1119–1128
- Dou, Y., Milne, T. A., Tackett, A. J., Smith, E. R., Fukuda, A., Wysocka, J., Allis, C. D., Chait, B. T., Hess, J. L., and Roeder, R. G. (2005) *Cell* **121**, 873–885
- Dou, Y., Milne, T. A., Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., Allis, C. D., and Roeder, R. G. (2006) *Nat. Struct. Mol. Biol.* **13**, 713–719
- Wysocka, J., Swigut, T., Milne, T. A., Dou, Y., Zhang, X., Burlingame, A. L., Roeder, R. G., Brivanlou, A. H., and Allis, C. D. (2005) *Cell* **121**, 859–872
- Ruthenburg, A. J., Wang, W., Graybosch, D. M., Li, H., Allis, C. D., Patel, D. J., and Verdine, G. L. (2006) *Nat. Struct. Mol. Biol.* **13**, 704–712
- Sheffield, P., Garrard, S., and Derewenda, Z. (1999) *Protein Expression Purif.* **15**, 34–39
- Couture, J., Collazo, E., and Trievel, R. C. (2006) *Nat. Struct. Mol. Biol.* **13**, 1606–1619
- Schuck, P. (2000) *Biophys. J.* **78**, 1606–1619
- Schuck, P. (2005) in *Analytical Ultracentrifugation: Techniques and Methods* (Scott, D. J., Harding, S. E., and Rowe, A. J., eds) pp. 26–50, Royal Society of Chemistry, Cambridge, UK
- Dam, J., Velikovskiy, C. A., Mariuzza, R. A., Urbanek, C., and Schuck, P. (2005) *Biophys. J.* **89**, 619–634
- Laue, T. M., Shah, B., Ridgeway, T. M., and Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., and Horton, L. C., eds) pp. 90–125, Royal Society of Chemistry, Cambridge
- Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., and Cheng, X. (2003) *Mol. Cell* **12**, 177–185
- Rasio, D., Schichman, S. A., Negrini, M., Canaani, E., and Croce, C. M. (1996) *Cancer Res.* **56**, 1766–1769
- Cho, Y. W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G. R., Copeland, T. D., Kalkum, M., and Ge, K. (2007) *J. Biol. Chem.* **282**, 20395–20406
- Rozenblatt-Rosen, O., Rozovskaia, T., Burakov, D., Sedkov, Y., Tillib, S., Blechman, J., Nakamura, T., Croce, C. M., Mazo, A., and Canaani, E. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 4152–4157
- Rozovskaia, T., Rozenblatt-Rosen, O., Sedkov, Y., Burakov, D., Yano, T., Nakamura, T., Petruck, S., Ben-Simchon, L., Croce, C. M., Mazo, A., and Canaani, E. (2000) *Oncogene* **19**, 351–357
- Couture, J. F., Collazo, E., and Trievel, R. C. (2006) *Nat. Struct. Mol. Biol.* **13**, 698–703
- Han, Z., Guo, L., Wang, H., Shen, Y., Deng, X. W., and Chai, J. (2006) *Mol. Cell* **22**, 137–144
- Schuetz, A., Allali-Hassani, A., Martin, F., Loppnau, P., Vedadi, M., Bochkarev, A., Plotnikov, A. N., Arrowsmith, C. H., and Min, J. (2006) *EMBO J.* **25**, 4245–4252

56. Avalos, J. L., Celic, I., Muhammad, S., Cosgrove, M. S., Boeke, J. D., and Wolberger, C. (2002) *Mol. Cell* **10**, 523–535
57. Brown, P. H., Balbo, A., and Schuck, P. (2007) *Biomacromolecules* **8**, 2011–2024
58. Patel, A., Dharmarajan, V., and Cosgrove, M. S. (2008) *J. Biol. Chem.* **283**, 32158–32161
59. Katsani, K. R., Arredondo, J. J., Kal, A. J., and Verrijzer, C. P. (2001) *Genes Dev.* **15**, 2197–2202
60. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) *Bioinformatics (Oxf.)* **23**, 2947–2948
61. Henikoff, S., and Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10915–10919

A Conserved Arginine-containing Motif Crucial for the Assembly and Enzymatic Activity of the Mixed Lineage Leukemia Protein-1 Core Complex

Anamika Patel, Valarie E. Vought, Venkatasubramanian Dharmarajan and Michael S. Cosgrove

J. Biol. Chem. 2008, 283:32162-32175.

doi: 10.1074/jbc.M806317200 originally published online October 1, 2008

Access the most updated version of this article at doi: [10.1074/jbc.M806317200](https://doi.org/10.1074/jbc.M806317200)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2008/10/01/M806317200.DC1>

This article cites 58 references, 19 of which can be accessed free at

<http://www.jbc.org/content/283/47/32162.full.html#ref-list-1>