A Non-Active-Site SET Domain Surface Crucial for the Interaction of MLL1 and the RbBP5/Ash2L Heterodimer within MLL Family Core Complexes

Stephen A. Shinsky[1], Michael Hu[2], Valarie E. Vought[2], Sarah B. Ng[3], Michael J. Bamshad[3,4], Jay Shendure[3] and Michael S. Cosgrove[1]

*1 - **Department of Biochemistry and Molecular Biology,** SUNY Upstate Medical University, Syracuse, NY 13210, USA*
*2 - **Department of Biology,** Syracuse University, Syracuse, NY 13210, USA*
*3 - **Department of Genome Sciences,** University of Washington Seattle, Seattle, WA 98105, USA*
*4 - **Department of Pediatrics,** University of Washington Seattle, Seattle, WA 98195, USA*

***Correspondence to Michael S. Cosgrove:** Department of Biochemistry and Molecular Biology, SUNY Upstate Medical University, 750 East Adams Street, Syracuse, NY 13210, USA. cosgrovm@upstate.edu*
http://dx.doi.org/10.1016/j.jmb.2014.03.011
Edited by S. Khorasanizadeh

Abstract

The mixed lineage leukemia-1 (MLL1) enzyme is a histone H3 lysine 4 (H3K4) monomethyltransferase and has served as a paradigm for understanding the mechanism of action of the human SET1 family of enzymes that include MLL1–MLL4 and SETd1a,b. Dimethylation of H3K4 requires a sub-complex including WRAD (*W*DR5, *R*bBP5, *A*sh2L, and *D*PY-30), which binds to each SET1 family member forming a minimal core complex that is required for multiple lysine methylation. We recently demonstrated that WRAD is a novel histone methyltransferase that preferentially catalyzes H3K4 dimethylation in a manner that is dependent on an unknown non-active-site surface from the MLL1 SET domain. Recent genome sequencing studies have identified a number of human disease-associated missense mutations that localize to the SET domains of several MLL family members. In this investigation, we mapped many of these mutations onto the three-dimensional structure of the SET domain and noticed that a subset of MLL2 (KMT2D, ALR, MLL4)-associated Kabuki syndrome missense mutations map to a common solvent-exposed surface that is not expected to alter enzymatic activity. We introduced these mutations into the MLL1 SET domain and observed that all are defective for H3K4 dimethylation by the MLL1 core complex, which is associated with a loss of the ability of MLL1 to interact with WRAD or with the RbBP5/Ash2L heterodimer. Our results suggest that amino acids from this surface, which we term the Kabuki interaction surface or *KIS*, are required for formation of a second active site within SET1 family core complexes.

Introduction

The enzymatic conversion from monomethylation to dimethylation of histone H3 lysine 4 (H3K4) is required for the epigenetic maintenance of transcriptionally active states of chromatin in eukaryotes [1–6]. The molecular mechanisms involved are not well understood. Mixed lineage leukemia protein-1 (MLL1, ALL1, KMT2A, HRX, EC 2.1.1.43) is a histone methyltransferase that predominantly catalyzes monomethylation of H3K4 using its evolutionarily conserved ~130-amino-acid SuVar, E(z), Trx or SET domain [7,8]. Dimethylation of H3K4 is dependent on MLL1's interaction with a conserved sub-complex called WRAD [*W*DR5 (tryptophan-aspartate repeat protein 5), *R*bBP5 (retinoblastoma-binding protein 5), *A*sh2L (absent-small-homeotic-2-like), and *D*PY-30 (Dumpy-30)], forming the MLL1 core complex [7,9]. The MLL1 core complex regulates expression of genes required for development [10–12], hematopoiesis [13–15], postnatal neurogenesis [16], and tissue homeostasis [17,18]. MLL1's interaction with WRAD is critical for these biological processes, but due to the absence of a high-resolution three-dimensional structure of the MLL1 core complex, the molecular surfaces involved are poorly understood. This information is necessary for comprehending the molecular mechanisms that regulate the degree of H3K4 methylation,

which, in turn, is crucial for understanding the role of MLL family members in human developmental disorders and malignancies and for the development of potential therapeutics.

Recent data indicate that the MLL1 core complex uses two distinct active sites to catalyze monomethylation and dimethylation of H3K4 in a stepwise manner. Evidence supporting this "two-active-site" model includes (1) the demonstration that the isolated MLL1 SET domain is an intrinsic H3K4 monomethyltransferase [7]; (2) the observation that an H3K4me1 intermediate accumulates during the course of the reaction catalyzed by the MLL1 core complex [7], indicating that H3K4me1 is released from the first active site before rebinding to undergo dimethylation; (3) that WRAD possesses an inherent H3K4 monomethyltransferase activity in the absence of MLL1 [7,19]; and (4) the demonstration that WRAD catalyzes H3K4me2 by preferentially monomethylating the H3K4me1 intermediate within a complex assembled with a catalytically inactive variant of MLL1 [7,20]. These results indicate that WRAD requires a non-active-site surface from MLL1 to form a second active site that preferentially recognizes H3K4me1 as a substrate for dimethylation. The MLL1 surface that interacts with WRAD to form the H3K4 dimethyltransferase active site is unknown.

Recent genome sequencing efforts have identified a large number of nonsense and missense mutations in several MLL family enzymes that are associated with human developmental disorders and cancers. For example, mutations in MLL1 (KMT2A), MLL2 (KMT2D [21], also known as ALR, MLL4), and MLL3 (KMT2C) are associated with Wiedemann-Steiner syndrome [22], Kabuki syndrome (KS) [23–29], and Kleefstra syndrome [30], respectively. MLL2 and MLL3 driver mutations are also associated with non-Hodgkin lymphomas [31,32]; pediatric medulloblastomas [33,34]; and lung [35], renal [36], and prostate carcinomas [37]. While the majority of mutations that cause disease result from loss-of-function nonsense mutations or frameshift-inducing indels, most missense mutations alter evolutionarily conserved amino acid positions with unknown functions; many of which are located in the conserved SET domain [38,39]. To begin understanding the impact of disease-associated missense mutations on the enzymatic activity of MLL family enzymes, we mapped known missense mutations onto the three-dimensional structure of the MLL1 SET domain and noticed that a subset of KS *de novo* dominant missense mutations localize to a common solvent-exposed surface that is distinct from the canonical SET domain active-site cleft. Since these residues are conserved in all MLL family enzymes, we hypothesized that they may constitute the unknown MLL1 surface that interacts with WRAD to form the H3K4 dimethyltransferase active site.

In order to better understand the impact of KS missense mutations on SET domain function, we introduced five KS missense mutations into the MLL1 SET domain and found that all are defective for H3K4 dimethylation when assembled into the MLL1 core complex. In one case, loss of activity is associated with mutation of the conserved arginine in the MLL1 *Win* (*W*DR5 *in*teraction) motif, which we and others have previously shown to be crucial for the interaction between MLL1 and WRAD [40–43]. The other missense mutations cluster on a common non-active-site SET domain surface (Fig. 1), we term the Kabuki interaction surface or *KIS*, and display impaired interactions with the RbBP5/Ash2L/DPY-30 (RAD) sub-complex when assembled within the MLL1 core complex. These results are consistent with the hypothesis that *KIS* amino acids are required for formation of the H3K4 dimethyltransferase active site within the MLL1 core complex. Since *KIS* surface amino acids are conserved from yeast to humans, these results are likely generalizable for all SET1 family core complexes.

Results

Impact of KS mutations on the structure and biochemistry of the isolated MLL1 SET domain

To begin understanding how disease-associated missense mutations affect the biochemistry of SET1 family enzymes, we initially attempted to introduce KS mutations into a recombinant human MLL2 SET domain construct, but due to poor expression, we were unable to obtain enough of each variant for rigorous biophysical comparisons. We therefore introduced five MLL2-associated KS missense substitutions and one non-KS control polymorphism (observed in population-based sequencing) into a similar recombinant human MLL1 construct consisting of amino acid residues 3745–3969 ($MLL1^{3745}$). This construct contains the SET and post-SET domains and the evolutionarily conserved *Win* motif, which is required for interaction with WRAD [41–43]. $MLL1^{3745}$ has 79% sequence similarity to that of a similarly sized construct of MLL2, and KS positions are conserved in all SET1 family SET domains from yeast to humans (Fig. 1). In addition, superposition of the MLL2 SET domain homology model with that of the crystal structure of MLL1 (RMSD, 0.24 Å for backbone atoms) suggests that KS amino acid residues occupy similar positions in the three-dimensional structure of the SET domain (Fig. 2). Furthermore, a recent study demonstrates that purified KMT2C and KMT2D complexes display, like that of the MLL1 core complex, H3K4 monomethyltransferase and dimethyltransferase activity [48], further justifying the use of MLL1 as a model for understanding the impact of KS missense mutations on SET1 family core complexes.

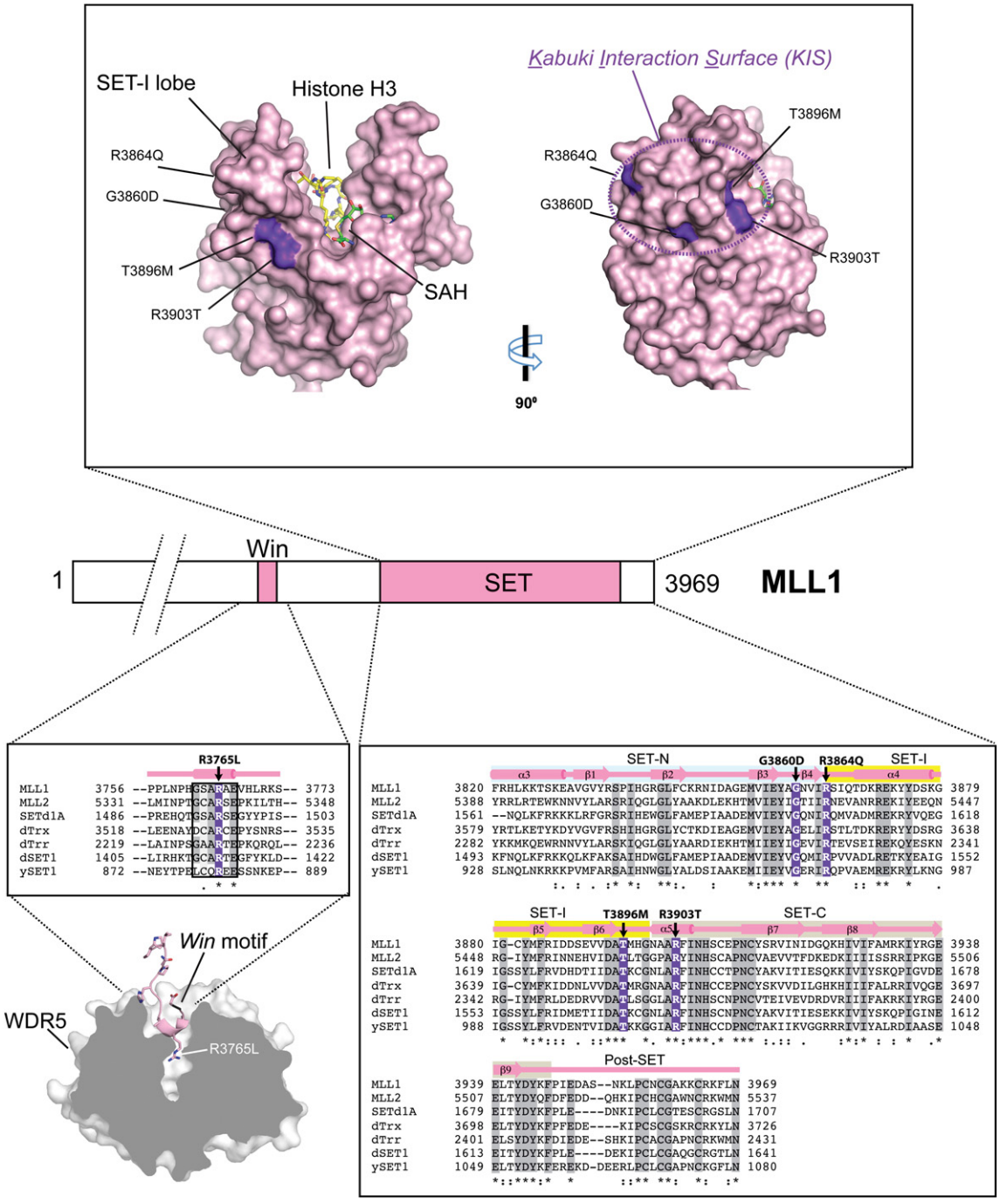

Fig. 1. KS amino acid positions are conserved in SET1 family SET domains and cluster on a common solvent-exposed SET domain surface. In the center is a schematic of the C-terminal portion of MLL1 domain structure with the *Win* motif and SET domain indicated. Above the schematic is a surface representation of the MLL1 SET domain crystal structure (PDB code 2W5Z [44]) with KS amino acid position indicated in purple. The location of histone H3 (yellow) and co-factor product *S*-adenosyl homocysteine (SAH) (green) binding sites are indicated. Below the schematic is a Clustal W [45] sequence alignment of human MLL1, MLL2 (KMT2D, ALR, MLL4), SET1a, *Drosophila* Trx, Trr, dSET1, and yeast SET1 proteins. Missense KS amino acid positions are highlighted in purple. Secondary structure prediction (PSIPRED) [46] is shown above the sequence alignment and the relative positions of SET-N, SET-I, SET-C, and post-SET sub-domains are indicated. Below the sequence alignment on the left is a cutaway view surface representation of WDR5 bound to the MLL1 *Win* motif peptide (PDB code 4ESG [40]) with the position of R3765 indicated. Structure figures were created with the PyMOL Molecular Graphics System, version 1.5.0.4, Schrödinger, LLC.

(a) **Summary of Kabuki syndrome SET domain proximal missense mutations.**

MLL1	MLL2	Sub-domain	Reference
R3765L	R5340L	*Win* motif	22
G3860D	G5428D	SET-N	27
R3864Q	R5432Q	SET-I	28
T3896M	T5464M	SET-I	22
R3903T	R5471T	SET-C	25

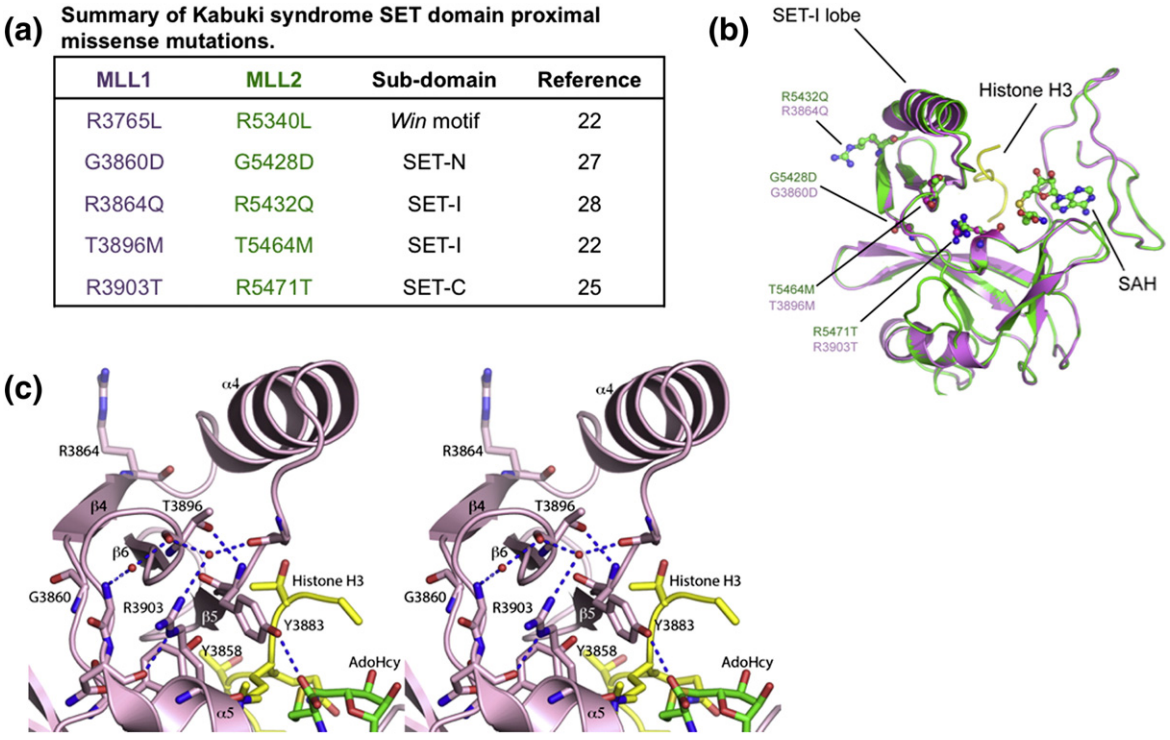

Fig. 2. KS missense amino acids are predicted to occupy similar positions on the MLL1 and MLL2 SET domains and are solvent exposed. (a) Table summarizing MLL2 (KMT2D, ALR, MLL4) KS amino acid positions in green and corresponding MLL1 positions in purple. (b) Superposition of the three-dimensional structure of the MLL1 SET domain (purple) with that of the MLL2 homology model (green), constructed using Modeller [47]. The average RMSD for backbone atoms is 0.24 Å. The positions of the SET-I lobe, histone H3, and AdoHcy are indicted. KS amino acid positions are indicated in purple (MLL1 numbering) and in green (MLL2 numbering). (c) Stereo diagram of the MLL1 SET domain showing solvent-exposed KS amino acid positions and their interactions. Carbon atoms from MLL1 are colored pink; nitrogen atoms, blue; and oxygen atoms, red. Histone H3 is shown with yellow carbon atoms and AdoHcy is shown with green carbons. Hydrogen bond interactions are shown with the blue broken lines. The side chain of R3864 is modeled as it is disordered beyond the C^{β} atom in the crystal structure of the MLL1 SET domain (2W5Z) [44].

Our modeling shows that four KS missense mutations (MLL1 numbering: G3860D, R3864Q, T3896M, R3903T) map in or near a region in the MLL1 SET domain structure known as the SET-I lobe, a region that is thought to contribute to histone substrate specificity [49,50]. However, based on our modeling, KS amino acid side chains are located on a solvent-exposed surface oriented away from the active-site cleft (Figs. 1 and 2). It is therefore unknown how substitution of KS amino acid positions will affect SET domain enzymatic activity of MLL family complexes. A fifth KS missense mutation replaces the evolutionarily conserved arginine in the MLL2 *Win* motif with leucine (R3765L in MLL1) (Fig. 1). Substitution of the homologous arginine with alanine in the MLL1 *Win* motif was previously shown to abolish the interaction between MLL1 and WDR5 [41–43], which results in the loss of the H3K4 dimethyltransferase activity of the MLL1 core complex [40,42]. The MLL1 *Win* motif adopts a three-dimensional structure that is similar to that of the MLL2 *Win* motif when bound to WDR5 [40,51]. We hypothesized that substitution of the *Win* motif arginine with leucine will likewise impair core complex assembly.

Several MLL1 KS variants are defective for methylation of histone H3

To determine if KS substitutions alter the enzymatic activity of the isolated SET domain in the absence of WRAD, we compared histone methylation activities of each variant with that of wild-type MLL1 using a radiolabeling assay. In this assay, the different MLL1 variants were incubated with H3 peptide and ^{3}H-AdoMet, and incorporation of ^{3}H-methyl groups into the H3 peptide was detected by fluorography. As a negative control, we included the N3906A variant of MLL1, which was previously shown to be catalytically inactive due to defective binding of AdoMet [7,20]. We observed ^{3}H-methyl incorporation into the H3 peptide in reactions catalyzed by wild-type MLL1 (Fig. 3a, lanes 2 and 12), the R3746H control variant (lane 6), and the R3765L (lane 8) and R3864Q (lane 16) KS variants.

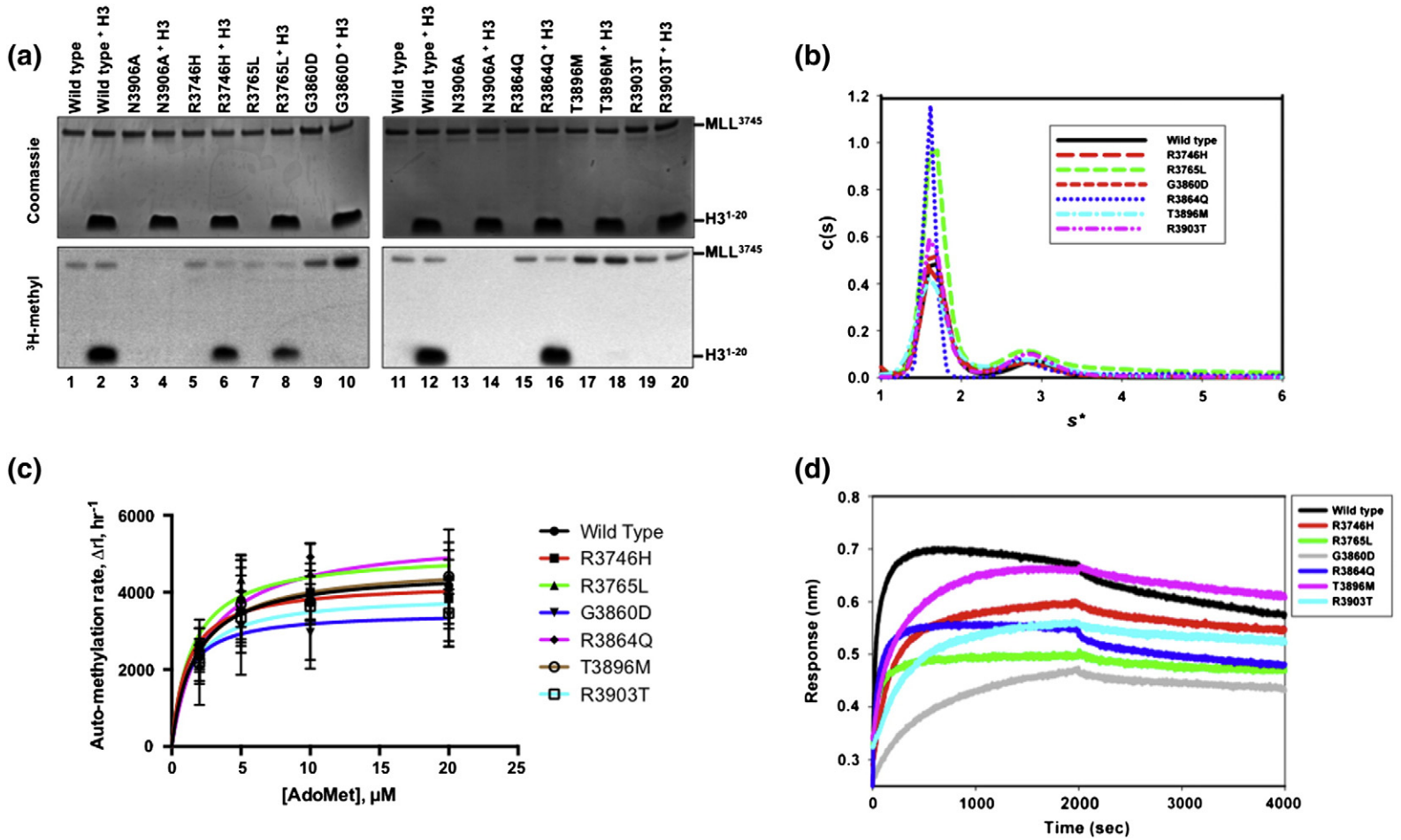

Fig. 3. Characterization of KS missense substitutions in the isolated MLL1 SET domain. (a) Comparison of the histone methyltransferase activity of wild type and KS variant MLL^{3745} SET domains. Upper panels show Coomassie blue-stained gels, and lower panels show ^{3}H-methyl incorporation after 8 h as determined by fluorography. (b) Comparison of wild type and KS variant diffusion-free sedimentation coefficient distributions [$c(s)$] derived from SV-AUC. (c) Comparison of apparent K_m curves for MLL1 catalyzed auto-methylation reactions with different concentrations of AdoMet. Error bars represent the standard deviation from two independent experiments. Fits were obtained using the Michaelis–Menten equation as described in Materials and Methods. (d) Comparison of histone H3 peptide (1–20) binding to wild type and KS variant MLL1 SET domains using biolayer interferometry. Biotinylated histone H3 was immobilized on streptavidin-coated sensors. Wild type or KS variant MLL1 SET domain association is shown between 0 and 2000 s, followed by dissociation between 2000 and 4000 s.

These results indicate that the R3765L and R3864Q variants are capable of at least monomethylation of H3K4. In contrast, no methylation was observed in reactions incubated with the KS variants G3860D (lane 10), T3896M (lane 18), and R3903T (lane 20) under the same conditions.

KS missense mutations do not significantly alter the SET domain fold

Since loss of enzymatic activity could be due to protein misfolding, aggregation, defective ligand binding, or removal of catalytic residues, we compared hydrodynamic and ligand binding properties of each KS variant with that of wild-type MLL1. To determine if KS substitutions alter the conformation or monodispersity of the SET domain fold, we compared hydrodynamic properties of each variant protein with that of wild-type MLL1 using sedimentation velocity analytical ultracentrifugation (SV-AUC). Amino acid substitutions that induce gross changes in protein hydrodynamic shape display altered sedimentation coefficients and frictional ratios in SV-AUC experiments [52]. However, in this investigation, each variant displayed an experimental sedimentation coefficient (*s**, uncorrected for buffer density, temperature, and viscosity) that is identical with that of wild-type MLL1 at 1.7 *s** (Fig. 3b and Table 1), indicating that KS amino acid substitutions do not severely alter the hydrodynamic shape of the SET domain. In addition, while each protein showed an additional peak with an *s** value of 2.2, which likely represents a glutathione *S*-transferase (GST) dimer contaminant, each variant was relatively monodisperse with no indication of aggregation. These results suggest that KS missense mutations do not significantly alter the SET domain fold.

All KS mutations bind AdoMet with similar affinities

To determine if KS substitutions alter SET domain AdoMet binding, we compared MLL1 SET domain auto-methylation activities of the KS variants with that of wild-type MLL1. We previously showed that the MLL1 SET domain undergoes a robust auto-methylation reaction in the absence of histone H3 with an apparent K_m value for AdoMet that is similar to that of the histone H3 methylation reaction [20]. As shown in Fig. 3a, unlike the AdoMet binding defective variant N3906A, which does not auto-methylate (lanes 3 and 13), each KS variant catalyzes the auto-methylation reaction at a level that is qualitatively similar to that of wild-type MLL1 (Fig. 3a, lanes 5, 7, 9, 15, 17, and 19). Despite being inactive for the methylation of histone H3 in the radiolabeling assay, the G3860D, T3896M, and R3903T variants all display auto-methylation activity (lanes 9, 17, and 19), with the G3860D and T3896M variants appearing to have increased activity (lanes 9 and 17). However, when we compared apparent AdoMet K_m values in auto-methylation reactions among wild type and all KS variant MLL1 SET domains, no significant differences were observed (Fig. 3c and Table 2, last column). These results suggest that KS missense mutations do not alter the ability of the SET domain to bind AdoMet.

Table 1. Summary of SV-AUC experiments

Sedimentation coefficients (*s**)[a]				
	M	MW	MWRAD$^{4\mu M}$	MWRAD$^{2\mu M}$
M^{WT}	1.7	2.8 ± 0.04	5.3 ± 0.02	5.3 ± 0.01
M^{R3746H}	1.7	2.8 ± 0.08	5.2 ± 0.09	5.3 ± 0.03
M^{R3765L}	1.7	NC[b]	4.6 ± 0.05	4.3 ± 0.02
M^{G3860D}	1.7	2.8 ± 0.03	4.7 ± 0.04	4.4 ± 0.01
M^{R3864Q}	1.7	2.8 ± 0.06	4.7 ± 0.03	4.4 ± 0.03
M^{T3896M}	1.7	2.8 ± 0.12	4.9 ± 0.04	4.7 ± 0.02
M^{R3903T}	1.7	2.8 ± 0.09	4.7 ± 0.03	4.5 ± 0.02

[a] Experimental sedimentation coefficient (*s**, uncorrected for buffer density, temperature, and viscosity) ± standard deviation from two experimental replicates.
[b] NC, no complex observed under the experimental conditions tested.

Histone H3 binding

To determine if KS substitutions alter the binding of histone H3 to the MLL1 SET domain, we compared binding constants to a biotinylated histone H3 peptide containing residues 1–20 using biolayer interferometry. Biotinylated histone H3 peptide was loaded onto streptavidin-coated sensors and dipped into a solution containing 1 or 2 μM MLL1. All wild type and KS variant SET domains showed histone H3-dependent association with sensors, with the G3860D variant showing more than an order of magnitude slower rate of association (Fig. 3d and Table 2). The dissociation rate of MLL1 from the sensors was slow but could be increased by raising the pH of the dissociation buffer (data not shown), arguing against irreversible aggregation on the sensor. Fitting of the association and dissociation

Table 2. Summary of steady kinetics and thermodynamic binding experiments

SET domain	K_D[a] (nM)	K_d[a] × 10^{-4} (s^{-1})	K_{on}[a] (M^{-1} s^{-1})	K_m[b] AdoMet (μM)
M^{WT}	107 ± 14	7.2 ± 0.3	6900 ± 915	1.4 ± 0.3
M^{R3746H}	161 ± 29	8.1 ± 0.8	5500 ± 1280	4.1 ± 1.3
M^{R3765L}	162 ± 28	13.9 ± 3.8	8200 ± 1120	2.1 ± 0.5
M^{G3860D}	1580 ± 410	8.6 ± 0.8	590 ± 91	1.1 ± 0.3
M^{R3864Q}	143 ± 30	10.5 ± 0.5	8400 ± 2540	3.0 ± 1.8
M^{T3896M}	264 ± 43	8.3 ± 1.6	3160 ± 364	2.2 ± 0.3
M^{R3903T}	467 ± 77	7.9 ± 0.6	1780 ± 247	2.3 ± 0.4

[a] Standard error of measurement from three independent experiments.
[b] K_m ± standard error of the fit to the Michaelis–Menten equation.

profiles at pH 7.5 revealed that the wild type, R3746H, R3765L, R3864Q, T3896M, and R3903T MLL1 variants have similar apparent dissociation constants (K_d), which range from 107 to 467 nM. In contrast, the G3860D variant showed a dissociation constant that was approximately an order of magnitude greater than that of wild-type MLL1 (Table 2). Thus, all KS variants associate with histone H3, with one (G3860D) showing weaker binding, which may at least partially account for its loss of histone H3 methylation activity.

Taken together, these results suggest that the loss of enzymatic activity observed in the MLL1 variants cannot be explained by SET domain misfolding, large conformational changes, defective AdoMet, or defective histone H3 binding with the exception of the G3860D variant.

Impact of KS mutations on the assembly and enzymatic activity of the MLL1 core complex

All KS core complexes are defective for dimethylation of H3K4

MLL family enzymes interact with a conserved sub-complex of proteins that include WRAD, which is required for multiple H3K4 methylation [7,9]. Previous experiments revealed that the MLL1 SET domain catalyzes predominantly monomethylation of H3K4 [7]. When WRAD complexes with the MLL1 SET domain, two significant changes in enzymatic activity are observed under single-turnover conditions: (1) a 600-fold increase in the rate of H3K4 monomethylation, which accumulates in a non-processive manner and rapidly decays due to (2) a significant increase in the rate of H3K4 dimethylation [7] (Fig. 4). To determine how KS missense mutations impact these two activities, we assembled MLL1 core complexes (MWRAD) with wild type or KS variant MLL1 SET domains and compared single-turnover kinetic profiles using quantitative matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometry. After reactions were incubated for 24 h, MALDI TOF spectra showed that MWRAD complexes assembled with wild type and the control polymorphism R3746H SET domains converted most of the histone H3 peptide into the dimethyl form of H3K4 (Fig. 4a and b), with similar overall reaction kinetics (Fig. 4h and i) (Table 3). Conversely, all KS substitutions either abolished or nearly abolished H3K4 dimethylation by the MLL1 core complex (Fig. 4c–g). MWRAD complexes assembled with the R3765L and R3864Q variants displayed slow monomethylation activity with pseudo-first-order rate constants (k_1) that were each reduced compared to the wild-type complex by ~170-fold (Table 3). No monomethylation or dimethylation activity was detected in complexes assembled with the G3860D and R3903T MLL1 variants. In contrast, the T3896M variant, which was shown to be inactive in the absence of WRAD (Fig. 3a), showed significantly stimulated monomethylation activity in the presence of WRAD but displayed only with a small amount of dimethylation activity after 24 h (Fig. 4f). The first-order rate constant for monomethylation for the complex assembled with the T3896M variant was reduced 14-fold compared to that of the complex assembled with wild-type MLL1 (Table 3).

All KS variants within the SET domain interact with WDR5

Since differences in enzymatic activity among MLL1 core complexes assembled with KS variants cannot be fully explained by improper folding of the SET domain or defective ligand binding, we set out to evaluate the impact of KS missense substitutions on MLL1 core complex assembly. We previously established the pathway for complex assembly and the apparent dissociation constants for each interaction using SV-AUC [7], which is in agreement with the results from pull-down experiments with other SET1 family complexes [9,54]. The WDR5 subunit engages in a direct, high-affinity interaction with the MLL1 *Win* motif, which serves as a bridge for interaction with the relatively stable RAD sub-complex [40–43]. To determine if KS substitutions alter the interaction of MLL family enzymes with WDR5, we compared sedimentation coefficients of the MLL1–WDR5 binary complex assembled with wild type or KS variant MLL1 SET domains using SV-AUC. As expected, replacement of R3765 with leucine in the MLL1 *Win* motif abolishes the interaction between MLL1 and WDR5, as two non-interacting species are observed in the $c(s)$ distribution with s^* values similar to that of unbound MLL1 (1.7 s^*) and WDR5 (2.2 s^*) (Fig. 5c). In contrast, substitution of all other KS variant positions located within the SET domain did not alter the interaction between MLL1 and WDR5, as all displayed sedimentation coefficients that were identical with that of the wild type and control polymorphism MW complexes at 2.8 s^* (Fig. 5a, b, and d–g and Table 1).

Taken together, these results indicate that the loss of enzymatic activity for complexes assembled with the KS variants located within the SET domain cannot be explained by a loss of interaction with WDR5. In contrast, replacement of R3765 with leucine in the MLL1 *Win* motif abolishes the interaction between MLL1 and WDR5, which likely accounts for the loss of the H3K4 dimethylation activity observed by the core complex assembled with the R3765L variant.

KS mutations alter interaction with the RAD sub-complex

To determine the impact of KS missense substitutions on the assembly of the holo-MLL1 core complex, we assembled WRAD with wild type and

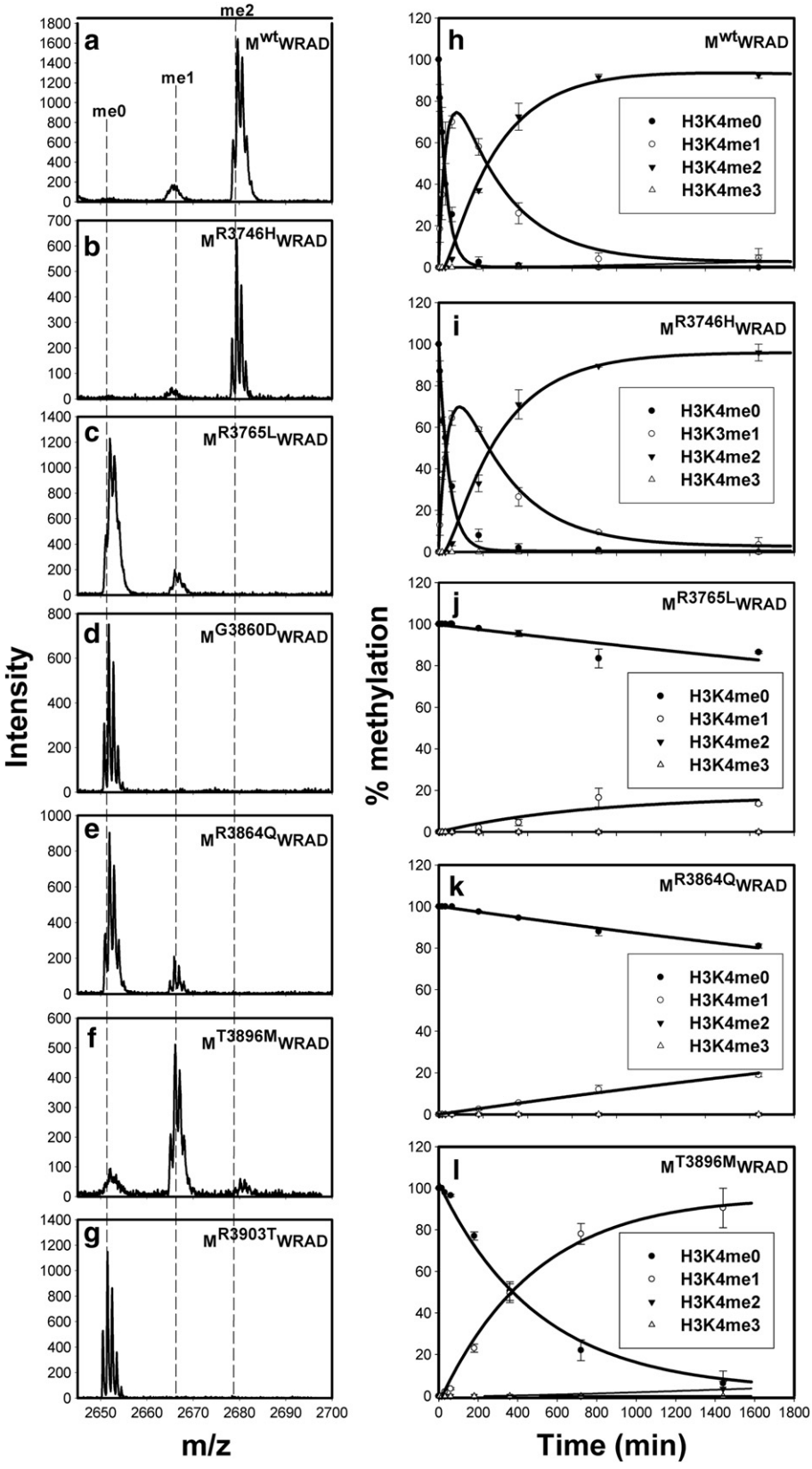
a
M^wtWRAD
me0
me1
me2
b
M^R3746HWRAD
c
M^R3765LWRAD
d
M^G3860DWRAD
e
M^R3864QWRAD
f
M^T3896MWRAD
g
M^R3903TWRAD
Intensity
m/z
h
M^wtWRAD
i
M^R3746HWRAD
j
M^R3765LWRAD
k
M^R3864QWRAD
l
M^T3896MWRAD
H3K4me0
H3K4me1
H3K3me1
H3K4me2
H3K4me3
% methylation
Time (min)

Table 3. Summary of single-turnover rate constants for core complexes assembled with wild type or variant MLL1 SET domains

Complex	Rate constants[a] (h^{-1})	
	k_1	k_2
M^{WT}WRAD	1.68 ± 0.10	0.25 ± 0.02
M^{R3746H}WRAD	1.32 ± 0.11	0.25 ± 0.02
M^{R3765L}WRAD	0.01 ± 0.00[b]	NA[c]
M^{G3860D}WRAD	NA	NA
M^{R3864Q}WRAD	0.01 ± 0.00[b]	NA
M^{T3896M}WRAD	0.12 ± 0.01	NA
M^{R3903T}WRAD	NA	NA

[a] Rate constants from single-turnover experiments were obtained by global fitting the average methylation intensities ± the standard error over time using Dynafit 4.0.
[b] Sample could not be globally fit, and rate reflects fit to exponential decay model [Eq. (2)].
[c] NA, no activity observed.

KS variant MLL1 SET domains and compared *s** values using SV-AUC. The complex assembled with the MLL1 control polymorphism R3746H sediments with an *s** value of 5.2, which is highly similar to that of the complex assembled with wild-type MLL1 at 5.3 *s** (Fig. 5h and i) (Table 1). Conversely, when the R3765L MLL1 variant was mixed with WRAD, peaks at 1.7 *s** (data not shown) and 4.6 *s** were observed (Fig. 5j). The peak at 1.7 *s** corresponds to free MLL1, whereas the peak at 4.6 *s** is larger than that expected for the free WRAD complex, which sediments under similar conditions at 4.2 *s** [19]. These data indicate that the peak at 4.6 *s** likely represents an equilibrium mixture between free WRAD and the MWRAD complex that cannot be resolved within the signal-to-noise ratio of the data. Since the R3765L variant does not appreciably interact with WDR5 in pairwise experiments (Fig. 5c), it is likely that the shifted WRAD *s** value results from a weak interaction between MLL1 and the RAD sub-complex. In contrast, KS variants G3860D, R3864Q, T3896M, and R3903T form complexes with WRAD, but with altered *s** values that range between 4.7 and 4.9 *s**. The magnitude of these *s** value differences between wild type and KS complexes significantly exceeds the expected variability of repeated measurements of any individual complex carried out under identical conditions with the same instrument, which is typically $\leq$0.1 *s** units (Table 1).

These results suggest that KS substitutions either alter the hydrodynamic shape of the fully assembled MLL1 core complex or weaken the interaction of the MLL1 SET domain with WRAD or a sub-complex thereof. To distinguish these hypotheses, we compared *s** values of complexes assembled with wild type or KS variant MLL1 SET domains at two different concentrations using SV-AUC. Variant complexes with altered hydrodynamic shapes but with unaltered binding equilibria among subunits display different *s** values that are independent of complex concentration. In contrast, mutations that alter binding equilibria among subunits display concentration-dependent *s** value shifts. As shown in Fig. 5 and in Table 1, a 2-fold dilution of complexes assembled with wild-type MLL1 and the R3746H control polymorphism showed identical values at 5.3 *s** (Fig. 5h and i), indicating that these complexes are stable on the timescale of sedimentation. In contrast, 2-fold dilution of all complexes assembled with KS variants showed concentration-dependent shifts to lower *s** values ranging from 4.4 to 4.7. These results indicate that one or more binding equilibria have been altered in each KS variant complex so that the complexes are no longer stable on the timescale of sedimentation. With the exception of R3765L in the MLL1 *Win* motif, KS substitutions do not alter the binding equilibrium between MLL1 and WDR5. We therefore conclude that each KS substitution within the MLL1 SET domain weakens the binding interaction between the MW sub-complex and the RAD sub-complex.

Discussion

The ~130-amino-acid SET domain comprises only 2–12% of the primary sequence of SET1 family proteins; however, its enzymatic activity has been suggested to underlie a general mechanism for chromatin-mediated transcriptional regulation [8]. The important role of the enzymatic activity of the SET domain in development was first suggested by the discovery that the Trithorax (Trx) mutant allele Trx^{Z11}, which results from a single glycine-to-serine substitution at a conserved position in the Trx SET domain [55], causes larval/pupal lethality and homeotic transformations [56]. We recently demonstrated with the recombinant Trx SET domain that the Trx^{Z11} G3601S variant is catalytically inactive [57], suggesting that this

Fig. 4. Impact of KS missense substitutions on the enzymatic activity of the MLL1 core complex. (a–g) MALDI TOF spectra of enzymatic assays of MLL1 core complexes assembled with wild type or KS variant MLL1 SET domains after 24 h. (h–l) Reaction progress curves showing the kinetic progress of different states of H3K4 methylation under single-turnover conditions. Each time point represents the mean percentage of total integrated area for each species in MALDI TOF reactions. Error bars represent the ±standard error from duplicate measurements. Data were global fitted to an irreversible consecutive reactions model [Eqs. (1–3)] as previously described [7] using Dynafit 4.0 [53]. The kinetic plots for the complexes assembled with the G3860D and R3903T variants are not shown because they displayed no monomethylation or dimethylation activity.

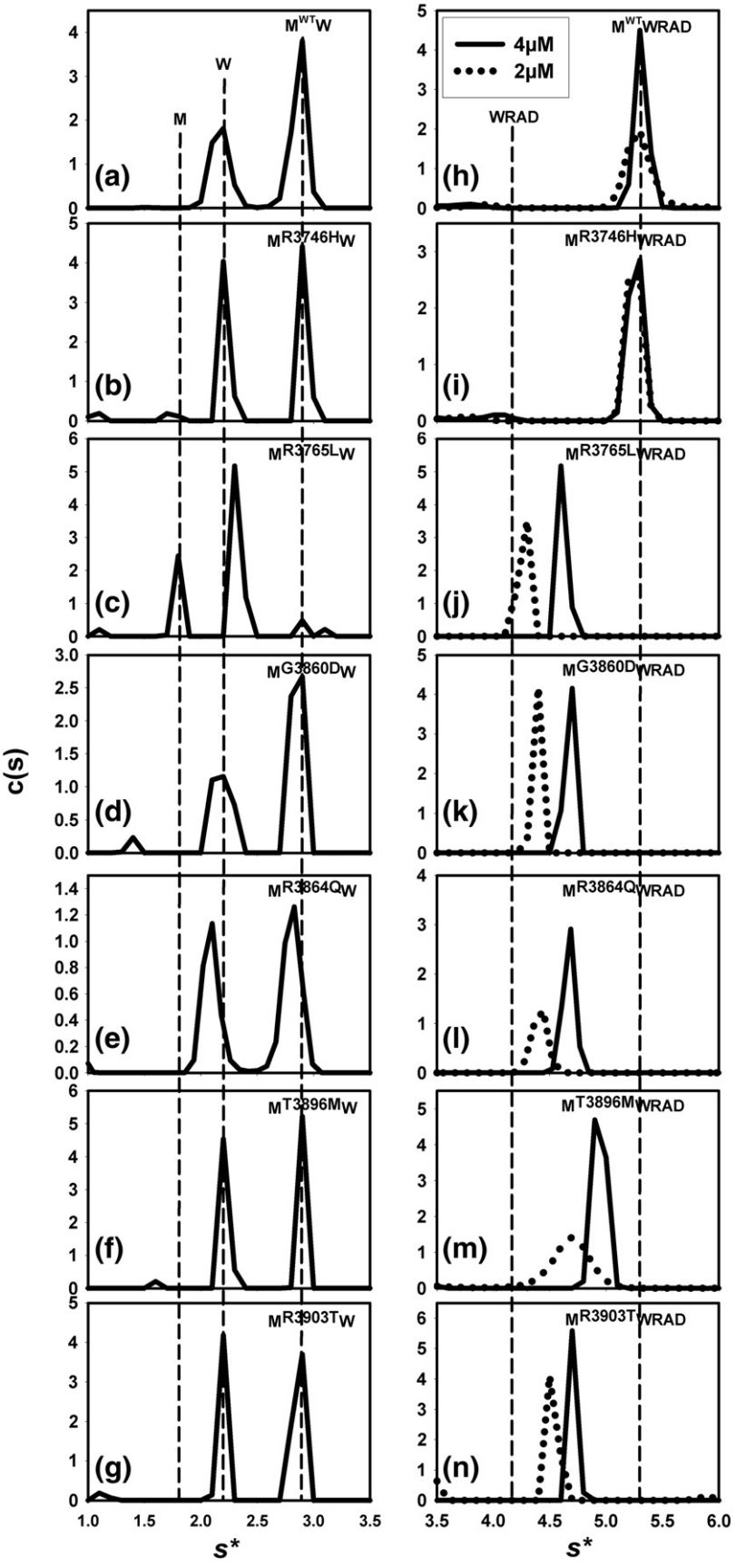

Fig. 5. Impact of KS missense substitutions on the assembly of the MLL1 core complex. (a–g) The *c*(*s*) distributions for the interaction of wild type or KS variant MLL1 SET domains with WDR5. (h–n) Comparison of *c*(*s*) distributions of wild type or KS variant MLL1 SET domains when assembled with WRAD. Continuous line shows *c*(*s*) distributions at an MLL1 core complex concentration of 4 μM. Dotted line represents *c*(*s*) distributions of the MLL1 core complex at a concentration of 2 μM.

phenotype may be related to loss of enzymatic activity, possibly because it is required for histone binding [58]. Likewise, targeted disruption of the murine MLL1 locus results in posterior homeotic transformations and altered *Hox* gene expression patterns similar to that observed in *Drosophila* [10]. In addition, although less severe, comparable phenotypes are observed in mice lacking the evolutionarily conserved SET domain at the C-terminus of MLL1 [12]. These results support the hypothesis that the histone methyltransferase activity of SET1 family members plays a crucial role in the maintenance of gene expression in development.

The recent discovery that mutations in MLL2 are associated with ~70% of cases with the human developmental disorder KS [23] appears to be consistent with this hypothesis. However, the majority of identified mutations are nonsense or frameshift-inducing indels that occur throughout the open reading frame, making it difficult to distinguish competing mechanisms for loss of function. For example, while loss of histone methyltransferase activity of MLL2 might be a common mechanism for all KS nonsense mutations that delete the SET domain, it is also possible that they could disrupt some other function of MLL2, such as gene targeting or recruitment of other factors. In addition, it is unclear if the observed missense mutations result in a similar loss of function because they alter amino acid positions of unknown function. The results from this investigation demonstrate that KS missense variants located in or around the SET domain display several functional properties consistent with correctly folded protein; however, all show defects in enzymatic activity, particularly when assembled with components of the core complex. These results suggest that KS amino acid positions are important for the assembly of SET1 family core complexes.

Since multi-subunit assemblies are required for full activity, identification of all the protein interactions that make up the SET1 family core complexes will be essential for a complete understanding of how H3K4 methylation is regulated in cells. Progress has been made in the crystallization and structural characterization of several MLL1 core complex sub-domains [41,43,59–62], and domain-mapping experiments have elucidated the details of key interactions between MLL1 and WDR5 [42,43], WDR5 and RbBP5 [63,64], and DPY-30 and Ash2L [65]. Cryo-electron microscopy studies have provided useful information on the overall shape of the budding yeast COMPASS (*Comp*lex *as*sociated with *S*ET1) and the human MLL1 core complex [66]. However, in the absence of a high-resolution crystal structure of the holo-complex, many of the detailed interactions remain to be elucidated. In particular, it is still unclear how the SET domain interacts with RbBP5 and Ash2L and why this interaction is required for the H3K4 dimethylation activity of the MLL1 core complex.

The topological order of subunits of the MLL1 core complex is MLL1/WDR5/RbBP5/Ash2L/DPY-30 (MWRAD), as determined based on pairwise analytical ultracentrifugation [7] and immunoprecipitation [9] experiments. A similar topology of MLL3 and MLL4 core complexes was suggested based on GST pull-down experiments [54]. We and others previously discovered that WDR5 interacts directly with MLL1 through recognition of a conserved arginine-containing sequence called the *Win* motif [41–43]. Replacement of R3765 within the MLL1 *Win* motif with alanine abolishes the interaction of MLL1 and WRAD *in vitro* and *in vivo* [40,42], which results in loss in the ability to form the H3K4 dimethyltransferase active site [42] and loss in the ability of the complex to recognize nucleosomes as a substrate [19]. The three-dimensional structure of an MLL1 *Win* motif peptide bound to WDR5 shows that R3765 is inserted into the central tunnel of the WDR5 β-propeller structure [41,43]. The guanidinium moiety of R3765 is positioned between a pair of conserved phenylalanine residues on WDR5, which is stabilized by cation–π, van der Walls, and hydrogen bond interactions. A similar architecture is observed in crystal structures of WDR5 bound to MLL2 *Win* motif peptides [40,51]. However, it was unclear whether replacement of the *Win* motif arginine with leucine, as observed in MLL2 in KS [23], will have similar effects. Here we show that substitution of R3765 of MLL1 with leucine abolishes the interaction between MLL1 and WDR5 resulting in the complete loss of the H3K4 dimethylation activity of the MLL1 core complex. It is likely that the binding energy of the complex is dominated by the cation–π interactions between R3765 and the conserved phenylalanine pair in WDR5. Loss of the cation–π interaction would explain why the *Win* motif arginine cannot be replaced by leucine for insertion into a relatively hydrophobic pocket on WDR5.

While a direct interaction between MLL1 and RbBP5 or Ash2L could not be detected in pairwise experiments [7,54], evidence suggests that they do interact within the context of the holo-complex. MLL1 monomethylation activity is stimulated by RbBP5 or the Ash2L/DPY-30 sub-complex in the absence of WDR5 [59]. RbBP5 modestly stimulates monomethylation of H3K4 by ~2- to 6-fold when added to the MLL1–WDR5 complex [7,59]. Moreover, the addition of RbBP5, Ash2L, and DPY-30 to the MLL1–WDR5 complex results in ~600-fold stimulation in the rate of H3K4 monomethylation [7]. In addition, an interaction between MLL1 and Ash2L is suggested by our recent demonstration that MLL1 methylates Ash2L in an intra-molecular manner within the core complex [20]. In this investigation, we demonstrate that, while the R3765L MLL1 variant does not show an interaction with WDR5, it does show a weak interaction with WRAD, suggesting that MLL1 interacts weakly with RAD sub-complex in the

absence of the *Win* motif–WDR5 interaction. Other studies with MLL1 paralogs SETd1a,b indicate that the RbBP5/Ash2L heterodimer remains bound to the SET domain in the absence of WDR5, which is stabilized by a KRKK sequence located in the SET-N region of the SET domain [67]. However, a similar sequence is absent in MLL1, likely explaining the requirement for WDR5 for stable association between MLL1 and the RbBP5/Ash2L heterodimer [9,42]. The results of this investigation are consistent with this hypothesis as all KS-associated missense mutations within the MLL1 SET domain specifically weaken the interaction with the RbBP5/Ash2L heterodimer without altering their ability to interact with WDR5.

One of the common consequences of all KS mutations studied in this investigation is the loss of the H3K4 dimethylation activity of the MLL1 core complex. A conceptual framework that is consistent with our experimental observations is that KS mutations disrupt the formation and/or mechanism of action of the dimethyltransferase active site (Fig. 6). We have hypothesized that the MLL1 core complex uses two distinct active sites to catalyze monomethylation and dimethylation of H3K4 in a stepwise manner, based on non-processive kinetic behavior, the monomethylation activity of the MLL1 SET domain, and the demonstration that WRAD catalyzes H3K4 dimethylation in a complex assembled with a catalytically inactive variant of MLL1 [7,20]. Consistent with this hypothesis, we recently showed that automethylation reactions catalyzed by the MLL1 SET domain are inhibited by unmodified histone H3, but not by histones previously monomethylated at H3K4, suggesting that they do not bind to the SET domain during the dimethylation reaction [20].

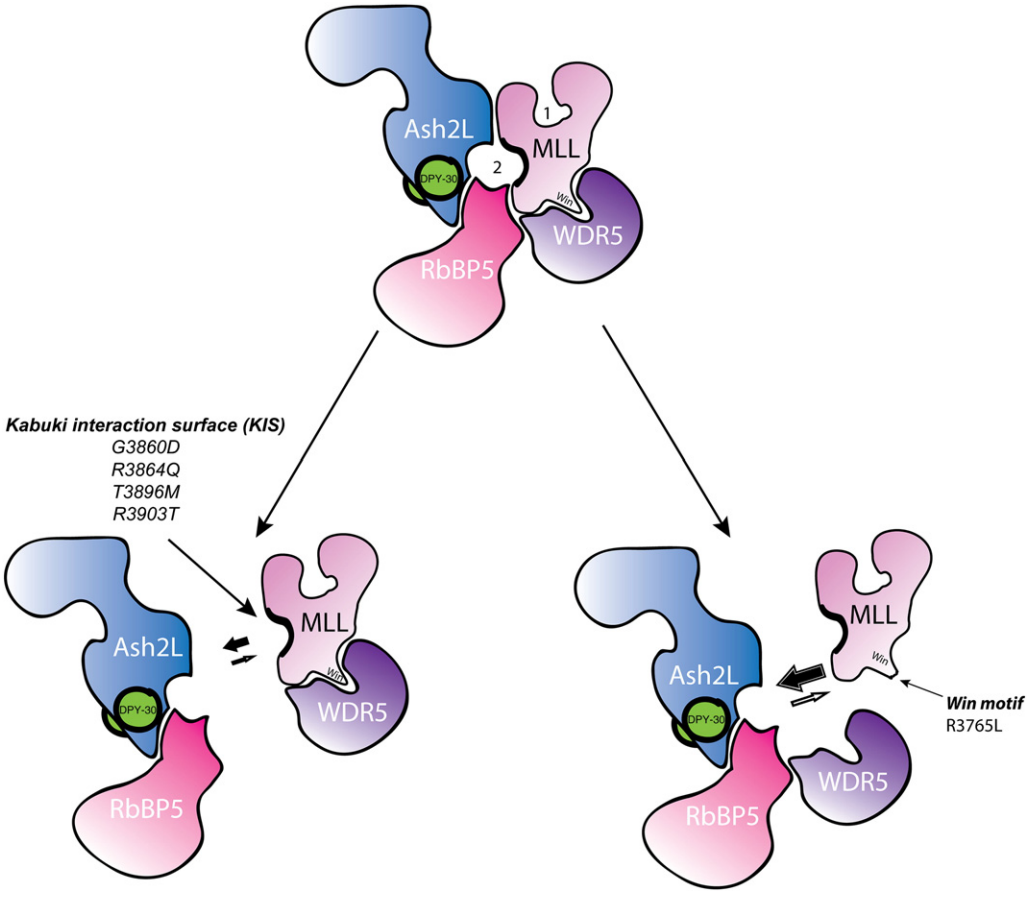

Fig. 6. KS missense mutations define a *KIS* motif that is crucial for interaction with the RbBP5/Ash2L heterodimer and for the H3K4 dimethylation activity of MLL family complexes. (Upper panel) The two-active-site model for multiple lysine methylation by MLL family complexes. The MLL core complex monomethylates H3K4 at the canonical SET domain active site (indicated by the number 1), which is followed by H3K4 dimethylation at a second active site formed at the interface between MLL and WRAD (indicated by number 2). *KIS* localized KS mutations in the MLL SET domain weaken the interaction with RbBP5 and Ash2L resulting in the loss of H3K4 dimethylation activity (lower left panel). Alternatively, substitution of R3765 with leucine in the *Win* motif destabilizes the interaction between MLL and WRAD thus also preventing formation of the dimethylation active site (lower right panel).

Because WRAD possesses no homology to known methyltransferases, the location of the second active has remained elusive. In addition, WRAD catalyzes only monomethylation of H3K4 in the absence of MLL1 [7,19] but acquires the ability to preferentially monomethylate the H3K4me1 substrate when assembled with a catalytically inactive variant of MLL1 [20]. These results suggest that WRAD requires a surface from MLL1, distinct from the SET domain active site, to complete formation of a functional dimethyltransferase active site. Results from this investigation are consistent with this hypothesis. Several disease-associated missense mutations alter conserved amino positions that cluster on a common solvent-exposed surface oriented away from canonical SET domain active-site cleft. Substitution of these amino acids in the MLL1 SET domain all demonstrate defects in the ability to form a stable complex with the RAD sub-complex, resulting in the loss of the H3K4 dimethylation. The fact that these mutations do not alter the hydrodynamic shape of the MLL1 SET domain nor do they significantly affect the ability of MLL1 to bind AdoMet, histone H3 (with the exception of G3860D), or WDR5 argues against the hypothesis that the loss of activity is associated with a global perturbation in the SET domain structure.

While loss of the ability to form the dimethyltransferase active site provides a plausible explanation for loss of the H3K4 dimethylation activity of the MLL1 core complex, the reasons for loss of H3K4 monomethylation activity in the G3860D and R3903T variants are less clear. Analysis of the crystal structure reveals that substitution of G3860 with Asp causes steric clashes with every possible aspartate rotamer, suggesting local structural rearrangements. This would likely have the effect of wedging apart the β4 and β6 strands, limiting SET-I lobe flexibility and altering the position of the invariant Y3858 side chain that is oriented in the active site and located two residues away (Fig. 2c). Indeed, it was previously shown that replacement of Y3858 with alanine or phenylalanine significantly reduces activity of the MLL1 SET domain [59]. Alteration of the position of Y3858 in the G3860D variant could explain reduced histone H3 binding and loss of H3K4 monomethylation activity by the MLL1 core complex. In contrast, substitution of R3903 with threonine is expected to disrupt a hydrogen bonding network that stabilizes the α5–β6 loop and its connection to the α4–β5 loop, which positions invariant Y3883 for a hydrogen bond with AdoMet (Fig. 2c). Loss of the hydrogen bond between Y3883 and AdoMet would be expected to reduce AdoMet affinity or result in incorrect positioning of AdoMet for the histone H3 methylation reaction. The fact that the R3903T variant undergoes a robust auto-methylation reaction suggests that latter is more likely. A similar mechanism can be invoked to explain loss of activity in the T3896M SET domain. The T3896 side-chain hydroxyl forms a hydrogen bond with the main chain amide of Y3883, an interaction that is expected to be lost when T3896 is replaced with methionine. However, in contrast to the R3903T variant, the defect observed in the T3896M variant is partially overcome by the addition of WRAD, as the core complex assembled with the T3896M variant regains stimulation in monomethylation activity. This stimulation may be related to the observation that the complex assembled with the T3896M variant is hydrodynamically more similar to the wild-type complex compared to the complexes assembled with the other KS variants (Fig. 5 and Table 1).

We note that the defects observed in the complexes assembled with the G3860D and R3903T variants are in contrast to what is observed when WRAD is assembled with the catalytically inactive N3906A variant of MLL1 [7], which forms a core complex with a sedimentation coefficient that is identical with that of wild-type MLL1 [7]. Unlike the G3860D and R3903T variants, the complex assembled with the N3906A variant displays weak monomethylation activity with the H3K4me0 substrate, which we have attributed to the activity of WRAD within the complex [7]. The lack of similar activity in complexes assembled with the G3860D and R3903T variants suggests that the WRAD active site may be partially occluded within their respective complexes.

Despite the differences among the KS variants as outlined above, all KS variants share the properties that they would be expected to alter a conserved surface on the SET-I sub-domain that faces away from the SET domain active-site cleft (Fig. 1). As a consequence, the complementarity of this surface for interaction with RbBP5 and Ash2L is likely to be altered, which may explain why all KS variants display defects in their ability to interact with the RAD sub-complex. While a subset of variants may induce local structural perturbations (G3860D, R3903T, T3896M) that could induce unexpected allosteric effects that are propagated to other surfaces on MLL, a structural basis for this allostery is not evident from the available crystallographic data. Furthermore, two of the variants (R3765L and R3864Q) that are not expected to induce structural changes are indeed similar to wild-type MLL1 in every respect except for their ability to catalyze H3K4 dimethylation and to interact with WRAD or RAD, respectively.

In summary, in this investigation, we utilized MLL1 as a model to determine the impact of KS missense mutations on the biochemistry of SET1 family enzymes. We found that the KS mutations under study map to a surface on the SET domain that is required for correct positioning of RAD within the core complex (Fig. 6). This conclusion is supported by our results that show that (1) KS mutations cluster on a common non-active-site SET domain surface that is solvent exposed, (2) KS variants have a diminished ability to interact with RAD, and (3) MLL1

core complexes assembled with KS variants exhibit a loss of H3K4 dimethyltransferase activity. Thus, we have named this MLL1 surface as the *KIS* domain. Because RAD interacts only with SET domains from the SET1 family, it is reasonable for the interaction to take place at a surface on the SET-I sub-domain that is unique to the SET1 family. We therefore conclude that the SET-I sub-domain not only is required for histone substrate specificity but also participates in the formation of the unique interactions that make up SET1 family core complexes.

Materials and Methods

Protein expression/purification

A human MLL1 construct consisting of amino acid residues 3745–3969 in a pGST vector was subjected to site-directed mutagenesis (QuikChange II XL, Stratagene). MLL1 mutants and full-length wild-type WRAD were individually expressed in *Escherichia coli* [Rosetta 2 (DE3) pLysS; Novagen] and purified by affinity chromatography as previously described [7,42]. WRAD components were further purified and buffer exchanged by gel-filtration chromatography (Superdex 200) using 20 mM Tris (pH 7.5), 300 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine, and 1 μM $ZnCl_2$. MLL mutants were purified by GST affinity chromatography in 50 mM Tris (pH 7.3), 300 mM NaCl, 3 mM dithiothreitol, 10% glycerol, and 1 μM $ZnCl_2$. Mutants were dialyzed into the buffer containing 20 mM Tris (pH 7.5), 300 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine, and 1 μM $ZnCl_2$.

Peptides

Biotinylated histone H3 peptides (hereafter referred to as "H3 peptide") were synthesized by Pi Proteomics, LLC, and contained residues 1–20 of histone H3 (ARTKQTARKSTGGKAPRKQL) followed by GG(K[biotin]). The peptides were also amidated on the C-terminus making the molecular mass 2651.144 and were >95% pure.

MALDI TOF mass spectrometry methyltransferase assays

An *in vitro* reaction containing the MLL1 core complex at 7 μM was incubated with 250 μM *S*-adenosyl methionine (AdoMet) and 10 μM H3 peptide at 15 °C for 24 h. At various time points, aliquots of the reaction were quenched with 0.5% trifluoroacetic acid. Samples were diluted 1:5 in α-cyano-4-hydroxycinnamic acid and shot on a Bruker Autoflex III mass spectrometer (SUNY College of Environmental Science and Forestry, Syracuse, NY) in reflectron mode. Relative methylation levels were quantitated using mMass [68] and fit with an irreversible consecutive reactions model [Eqs. (1–3)] using Dynafit [53] as previously described [7].

$$[A] = [A]_\circ \exp(-k_1 t) \tag{1}$$

$$[B] = \frac{[A]_\circ k_1}{k_2 - k_1[\exp(-k_1 t) - \exp(-k_2 t)]} \tag{2}$$

$$[C] = [A]_\circ \left\{ 1 + \frac{1}{k_1 - k_2[k_2 \exp(-k_1 t) - k_1 \exp(-k_2 t)]} \right\} \tag{3}$$

Analytical ultracentrifugation

MLL1 variants were mixed in stoichiometric amounts with WRAD (2 or 4 μM) and the samples were loaded into 3-mm two-sector charcoal-filled Epon centerpieces with quartz windows. All experiments were carried out at 10 °C using a Beckman Coulter ProteomLab XL-A analytical ultracentrifuge equipped with absorbance optics and a 4-hole An-60 Ti rotor at 60,000 RPM that was pre-equilibrated at 10 °C. The samples were scanned at 0-min time intervals for 300 scans and analyzed by the continuous distribution method $c(s)$ in the program SEDFIT [69].

^{3}H-Methyltransferase assays

Histone H3 methylation assays were conducted by incubating 5 μM MLL1 or MLL1 variant with 0.5 μM ^{3}H-AdoMet (Perkin Elmer, Inc.) and 100 μM H3 peptide for 8 h at 15 °2C. Reactions were quenched with SDS loading buffer, separated by 4–12% Bis-Tris SDS PAGE (Invitrogen) in 4-morpholineethanesulfonic acid buffer at 200 V for 30 min. Coomassie brilliant blue-stained gels were photographed and soaked for 30 min in autoradiography enhancer solution (Enlightning, Perkin Elmer, Inc.), dried for 2 h at 72 °C under constant vacuum, and exposed to film (Kodak BioMax MS Film) at −80 °C for 16–18 h.

AdoMet binding assays

Apparent K_m values for AdoMet in MLL1 auto-methylation reactions were determined by incubating wild type or variant MLL1 at a concentration of 14 μM with various concentrations of [^{3}H]AdoMet (0–50 μM) at 15 °C. The samples were quenched with SDS loading buffer at 1- and 3-h time points and separated by 4–12% Bis-Tris SDS PAGE and stained with Coomassie brilliant blue. Gels were soaked in an enhancer solution (Enlightning, Perkin Elmer, Inc.) for 30 min then dried and exposed to film for 7 days at −80 °C. Bands were quantified by densitometry using Image J [70], and relative intensity values were plotted as a function of time and fitted by linear regression to determine methylation rates. Methylation rates were then plotted as a function of AdoMet concentration and fitted by nonlinear least-squares regression to the Michaelis–Menten Eq. (4). We note that that all variants tested displayed a pattern of substrate inhibition above 20 μM AdoMet; therefore, apparent K_m values are reported from the data fitted over the concentration range from 0 to 20 μM AdoMet.

$$v = v_{max} \times \frac{[\text{AdoMet}]}{K_m + [\text{AdoMet}]} \tag{4}$$

Histone H3 binding assays

Histone H3 binding was measured using biolayer interferometry (BLI) using a ForteBio Octet Red instrument. MLL1 or MLL1 SET domain variants at a concentration of 1 or 2 μM were prepared in a buffer containing 20 mM Tris (pH 7.5), 300 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine, 1 μM $ZnCl_2$, 10% glycerol, and 0.5 mg/mL bovine serum albumin. The biotin-labeled H3 peptide was prepared at a concentration of 25 nM in the same buffer and immobilized onto streptavidin-coated sensors (ForteBio) until a threshold signal of 0.4 nm was reached. MLL1 or an MLL1 variant was allowed to associate to the ligand bound sensor for 2000 s then dissociated with buffer for 2000 s. All steps were performed at 25 °C with sensors dipped into 200 μL of sample and stirred at 1000 RPM. Reference sensors (without H3 peptide) were used for each variant tested to correct for background binding and baseline drift. Data were processed using the ForteBio Data Analysis Program (version 6.4) utilizing reference subtraction and Savitzky-Golay filtering. Association and dissociation profiles were fit to determine binding constants.

Acknowledgements

We thank Steven Hanes, Nilda Alicea-Velazquez, and Susan Viggiano for helpful comments on this manuscript. We also thank Thomas Duncan for his guidance and Nilda Alicea-Velazquez for her help with biolayer interferometry. This work is supported in part by National Institutes of Health grant 1R01CA140522 (to M.S.C.).

Received 7 February 2014;
Received in revised form 14 March 2014;
Accepted 20 March 2014
Available online 27 March 2014

Keywords:
Kabuki syndrome;
mixed lineage leukemia;
histone methylation;
SET domain;
RbBP5/Ash2L

Abbreviations used:
MLL, mixed lineage leukemia; H3K4, histone H3 lysine 4; KS, Kabuki syndrome; SV-AUC, sedimentation velocity analytical ultracentrifugation; GST, glutathione S-transferase; MALDI, matrix-assisted laser desorption/ionization; TOF, time-of-flight; RAD, RbBP5/Ash2L/DPY-30.

References

[1] Strahl BD, Ohba R, Cook RG, Allis CD. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. Proc Natl Acad Sci U S A 1999;96:14967–72.

[2] Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, et al. Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci U S A 2002;99:8695–700.

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

[4] Cosgrove MS, Wolberger C. How does the histone code work? Biochem Cell Biol 2005;83:468–76.

[5] Strahl BD, Allis CD. The language of covalent histone modifications. Nature 2000;403:41–5.

[6] Kim T, Buratowski S. Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5′ transcribed regions. Cell 2009;137:259–72.

[7] Patel A, Dharmarajan V, Vought VE, Cosgrove MS. On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. J Biol Chem 2009;284:24242–56.

[8] Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, Allis CD, et al. MLL targets SET domain methyltransferase activity to *Hox* gene promoters. Mol Cell 2002;10:1107–17.

[9] Dou Y, Milne TA, Ruthenburg AJ, Lee S, Lee JW, Verdine GL, et al. Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat Struct Mol Biol 2006;13:713–9.

[10] Yu BD, Hess JL, Horning SE, Brown GA, Korsmeyer SJ. Altered Hox expression and segmental identity in Mll-mutant mice. Nature 1995;378:505–8.

[11] Ingham PW. A clonal analysis of the requirement for the trithorax gene in the diversification of segments in *Drosophila*. J Embryol Exp Morphol 1985;89:349–65.

[12] Terranova R, Agherbi H, Boned A, Meresse S, Djabali M. Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. Proc Natl Acad Sci U S A 2006;103:6629–34.

[13] Jude CD, Climer L, Xu D, Artinger E, Fisher JK, Ernst P. Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. Cell Stem Cell 2007;1:324–37.

[14] McMahon KA, Hiew SY, Hadjur S, Veiga-Fernandes H, Menzel U, Price AJ, et al. Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal. Cell Stem Cell 2007;1:338–45.

[15] Gan T, Jude CD, Zaffuto K, Ernst P. Developmentally induced Mll1 loss reveals defects in postnatal haematopoiesis. Leukemia 2010;24:1732–41.

[16] Lim DA, Huang YC, Swigut T, Mirick AL, Garcia-Verdugo JM, Wysocka J, et al. Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells. Nature 2009;458:529–33.

[17] Yamashita M, Hirahara K, Shinnakasu R, Hosokawa H, Norikane S, Kimura MY, et al. Crucial role of MLL for the maintenance of memory T helper type 2 cell responses. Immunity 2006;24:611–22.

[18] Diehl F, Rossig L, Zeiher AM, Dimmeler S, Urbich C. The histone methyltransferase MLL is an upstream regulator of endothelial-cell sprout formation. Blood 2007;109:1472–8.

[19] Patel A, Vought VE, Dharmarajan V, Cosgrove MS. 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 2011;286:3359–69.

[20] Patel A, Vought VE, Swatkoski S, Viggiano S, Howard B, Dharmarajan V, 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 2013;289:868–84.

[21] Bogershausen N, Bruford E, Wollnik B. Skirting the pitfalls: a clear-cut nomenclature for H3K4 methyltransferases. Clin Genet 2013;83:212–4.
[22] Jones WD, Dafou D, McEntagart M, Woollard WJ, Elmslie FV, Holder-Espinasse M, et al. *De novo* mutations in MLL cause Wiedemann-Steiner syndrome. Am J Hum Genet 2012;91:358–64.
[23] Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 2010;42:790–3.
[24] Banka S, Howard E, Bunstone S, Chandler KE, Kerr B, Lachlan K, et al. *MLL2* mosaic mutations and intragenic deletion–duplications in patients with Kabuki syndrome. Clin Genet 2013;83:467–71.
[25] Banka S, Veeramachaneni R, Reardon W, Howard E, Bunstone S, Ragge N, et al. How genetically heterogeneous is Kabuki syndrome? MLL2 testing in 116 patients, review and analyses of mutation and phenotypic spectrum. Eur J Hum Genet 2012;20:381–8.
[26] Hannibal MC, Buckingham KJ, Ng SB, Ming JE, Beck AE, McMillin MJ, et al. Spectrum of MLL2 (ALR) mutations in 110 cases of Kabuki syndrome. Am J Med Genet A 2011;155A: 1511–6.
[27] Li Y, Bogershausen N, Alanay Y, Simsek Kiper PO, Plume N, Keupp K, et al. A mutation screen in patients with Kabuki syndrome. Hum Genet 2011;130:715–24.
[28] Paulussen AD, Stegmann AP, Blok MJ, Tserpelis D, Posma-Velter C, Detisch Y, et al. MLL2 mutation spectrum in 45 patients with Kabuki syndrome. Hum Mutat 2011;32: E2018–25.
[29] Kokitsu-Nakata NM, Petrin AL, Heard JP, Vendramini-Pittoli S, Henkle LE, dos Santos DV, et al. Analysis of MLL2 gene in the first Brazilian family with Kabuki syndrome. Am J Med Genet A 2012;158A:2003–8.
[30] Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, Vissers LE, et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. Am J Hum Genet 2012;91:73–82.
[31] Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 2011;476:298–303.
[32] Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet 2011;43:830–7.
[33] Pugh TJ, Weeraratne SD, Archer TC, Pomeranz Krummel DA, Auclair D, Bochicchio J, et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature 2012;488:106–10.
[34] Jones DT, Jager N, Kool M, Zichner T, Hutter B, Sultan M, et al. Dissecting the genomic complexity underlying medulloblastoma. Nature 2012;488:100–5.
[35] Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 2012;150:1107–20.
[36] Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature 2010;463:360–3.
[37] Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature 2012;487:239–43.
[38] Morgan MA, Shilatifard A. *Drosophila* SETs its sights on cancer: Trr/MLL3/4 COMPASS-like complexes in development and disease. Mol Cell Biol 2013;33:1698–701.
[39] Bogershausen N, Wollnik B. Unmasking Kabuki syndrome. Clin Genet 2013;83:201–11.
[40] Dharmarajan V, Lee JH, Patel A, Skalnik DG, Cosgrove MS. Structural basis for WDR5 interaction (Win) motif recognition in human SET1 family histone methyltransferases. J Biol Chem 2012;287:27275–89.
[41] Patel A, Dharmarajan V, Cosgrove MS. Structure of WDR5 bound to Mixed Lineage Leukemia Protein-1 peptide. J Biol Chem 2008;283:32158–61.
[42] Patel A, Vought V, Dharmarajan V, Cosgrove MS. A conserved arginine containing motif crucial for the assembly and enzymatic activity of the Mixed Lineage Leukemia protein-1 core complex. J Biol Chem 2008;283:32162–75.
[43] Song JJ, Kingston RE. WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. J Biol Chem 2008;283:35258–64.
[44] Southhall SM, Wong P, Odho Z, Roe SM, Wilson JR. Structural basis for the recruitment of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. Mol Cell 2009;33:181–91.
[45] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–8.
[46] Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195–202.
[47] Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
[48] Lee JE, Wang C, Xu S, Cho YW, Wang L, Feng X, et al. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. Elife 2013;2:e01503.
[49] Marmorstein R. Structure of SET domain proteins: a new twist on histone methylation. Trends Biochem Sci 2003;28:59–62.
[50] Dillon SC, Zhang X, Trievel RC, Cheng X. The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol 2005;6:227.
[51] Zhang P, Lee H, Brunzelle JS, Couture JF. The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. Nucleic Acids Res 2012;40:4237–46.
[52] Newell JO, Schachman HK. Amino acid substitutions which stabilize aspartate transcarbamoylase in the R state disrupt both homotropic and heterotropic effects. Biophys Chem 1990;37:183–96.
[53] Kuzmic P. Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. Anal Biochem 1996;237:260–73.
[54] Cho YW, Hong T, Hong S, Guo H, Yu H, Kim D, et al. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. J Biol Chem 2007;282:20395–406.
[55] Stassen MJ, Bailey D, Nelson S, Chinwalla V, Harte PJ. The *Drosophila* trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. Mech Dev 1995;52:209–23.
[56] Breen TR. Mutant alleles of the *Drosophila* trithorax gene produce common and unusual homeotic and other developmental phenotypes. Genetics 1999;152:319–44.
[57] Tie F, Banerjee R, Saiakhova AR, Howard B, Monteith KE, Scacheri PC, et al. Trithorax monomethylates H3K4 and interacts directly with CBP to promote H3K27 acetylation and

antagonize Polycomb silencing. Development 2014;141: 1129–39.
[58] Katsani KR, Arredondo JJ, Kal AJ, Verrijzer CP. A homeotic mutation in the trithorax SET domain impedes histone binding. Genes Dev 2001;15:2197–202.
[59] Southall SM, Wong PS, Odho Z, Roe SM, Wilson JR. Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. Mol Cell 2009;33:181–91.
[60] Chen Y, Cao F, Wan B, Dou Y, Lei M. Structure of the SPRY domain of human Ash2L and its interactions with RbBP5 and DPY30. Cell Res 2012;22:598–602.
[61] Chen Y, Wan B, Wang KC, Cao F, Yang Y, Protacio A, et al. Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding. EMBO Rep 2011;12:797–803.
[62] Sarvan S, Avdic V, Tremblay V, Chaturvedi CP, Zhang P, Lanouette S, et al. Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain. Nat Struct Mol Biol 2011;18:857–9.
[63] Avdic V, Zhang P, Lanouette S, Groulx A, Tremblay V, Brunzelle J, et al. Structural and biochemical insights into MLL1 core complex assembly. Structure 2011;19:101–8.
[64] Odho Z, Southall SM, Wilson JR. Characterization of a novel WDR5-binding site that recruits RbBP5 through a conserved motif to enhance methylation of histone H3 lysine 4 by mixed lineage leukemia protein-1. J Biol Chem 2010;285:32967–76.
[65] South PF, Fingerman IM, Mersman DP, Du HN, Briggs SD. A conserved interaction between the SDI domain of Bre2 and the Dpy-30 domain of Sdc1 is required for histone methylation and gene expression. J Biol Chem 2010;285:595–607.
[66] Takahashi YH, Westfield GH, Oleskie AN, Trievel RC, Shilatifard A, Skiniotis G. Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. Proc Natl Acad Sci U S A 2011;108:20526–31.
[67] Mersman DP, Du HN, Fingerman IM, South PF, Briggs SD. Charge-based interaction conserved within histone H3 lysine 4 (H3K4) methyltransferase complexes is needed for protein stability, histone methylation, and gene expression. J Biol Chem 2012;287:2652–65.
[68] Niedermeyer TH, Strohalm M. mMass as a software tool for the annotation of cyclic peptide tandem mass spectra. PLoS One 2012;7:e44913.
[69] Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J 2000;78:1606–19.
[70] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012;9: 671–5.