Multivalent Interactions by the Set8 Histone Methyltransferase With Its Nucleosome Substrate

Taverekere S. Girish, Robert K. McGinty and Song Tan

Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA

Correspondence to Song Tan: Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, 108 Althouse Laboratory, The Pennsylvania State University, University Park, PA 16802, USA. sxt30@psu.edu

http://dx.doi.org/10.1016/j.jmb.2016.02.025

Edited by K. Luger

Abstract

Set8 is the only mammalian monomethyltransferase responsible for H4K20me1, a methyl mark critical for genomic integrity of eukaryotic cells. We present here a structural model for how Set8 uses multivalent interactions to bind and methylate the nucleosome based on crystallographic and solution studies of the Set8/nucleosome complex. Our studies indicate that Set8 employs its i-SET and c-SET domains to engage nucleosomal DNA 1 to 1.5 turns from the nucleosomal dyad and in doing so, it positions the SET domain for catalysis with H4 Lys20. Surprisingly, we find that a basic N-terminal extension to the SET domain plays an even more prominent role in nucleosome binding, possibly by making an arginine anchor interaction with the nucleosome H2A/H2B acidic patch. We further show that proliferating cell nuclear antigen and the nucleosome compete for binding to Set8 through this basic extension, suggesting a mechanism for how nucleosome binding protects Set8 from proliferating cell nuclear antigen-dependent degradation during the cell cycle.

Introduction

The post-translational methylation of histone H4 on lysine 20 (H4K20me) is critical for the genomic integrity of eukaryotic cells. This modification plays key roles in DNA replication, DNA damage repair, and silenced heterochromatin [1–4]. Consequently, there is a significant interest in the precise mechanistic role of H4K20 methylation and the enzymes responsible for installing the methyl marks.

Like other lysine residues, H4K20 can be mono-, di-, or tri-methylated. Set8 (Pr-Set7, KMT5A) is the only known mammalian monomethyltransferase catalyzing the formation of H4K20me1. H4K20me1 is the preferred substrate for SUV4-20H1 and SUV4-20H2 histone methyltransferases (HMTase) to produce H4K20me2 and H4K20me3 [1–3]. These methyl marks may directly modulate chromatin compaction [5], or they may also recruit trans-acting replication factors in the origin replication complex or chromatin regulators such as 53BP1 and L3MBTL1 [6–8]. For example, recent studies confirm the role of Set8's enzymatic activity to recruit 53BP1 to site of double-strand DNA breaks [9]. The importance of the Set8 enzyme is also highlighted by studies showing that Set8 depletion results in severe changes to the cell, including increased double-stranded DNA breaks and defective cell cycle regulation [10–14]. Furthermore, Set8 is an essential gene in Drosophila and mice, with early embryonic lethality resulting from loss of Set8 [10,11,15,16].

Consistent with Set8's role in cell cycle progression, the enzymatic activity of Set8 is controlled in a cell cycle-dependent manner by regulating the levels of the Set8 protein [4,7,17]. Set8 protein is degraded upon ubiquitylation by the CRL4CDT2 ubiquitin ligase complex during S-phase and following DNA damage in a proliferating cell nuclear antigen (PCNA)-dependent manner [18–22]. Like many other proteins that directly bind the PCNA DNA replication clamp protein, Set8 contains a PCNA-interacting motif or PCNA-interacting peptide (PIP) box that mediates this interaction.

Set8 is a member of the SET domain family of lysine methyltransferases [15,16]. Like other
mammalian members of this family, Set8 contains additional n-SET and c-SET helical regions that flank the central SET domain (Fig. 1a) [24,25]. The SET domain itself is structurally conserved, but the two halves that constitute the domain are separated by an insertion, i-SET, which is variable in length and structure among SET family members [26]. This i-SET region contributes to substrate histone H4 peptide binding by cradling the peptide against the SET domain itself (Fig. 1b).

While crystallographic studies of Set8 bound to substrate peptides have provided valuable structural insights into substrate specificity [8,27,28], including the basis for Set8 being restricted to monomethylation of H4K20, we do not understand the structural basis for Set8’s preference for a nucleosome substrate. Set8 exhibits remarkably greater enzymatic activity on nucleosomes than on free histone substrates [15,16]. This suggests that Set8 must interact with other surfaces of the nucleosome in addition to the H4 N-terminal region surrounding the targeted H4K20 residue, but we lack information regarding these pertinent regions of Set8 or the nucleosome.

We have tackled the question of how Set8 interacts with the nucleosome through structural and biochemical studies. Our results show that Set8 binds to the nucleosome via multivalent interaction using both the SET domain and a basic region N-terminal to the SET domain. Intriguingly, the basic N-terminal region that we identify includes the PIP box that binds to PCNA. Our results thus explain Set8’s nucleosome specific activity and provide insights into how Set8’s binding to the nucleosome protects it from degradation during the cell cycle.

Results

Set8 histone methylase binding to the nucleosome

Crystal structures and biochemical studies of the Set8 SET domain revealed a structured region from residues 192–352 that is sufficient to monomethylate a histone H4 peptide at Lys20 [24,25]. However, we failed in our attempts to form a complex of Set8(191–352) with the nucleosome. We therefore performed a deletion analysis to determine the minimal domain of Set8 that would methylate and bind a nucleosome substrate. HMTase activity was determined using a filter-binding assay. To measure nucleosome binding, we monitored fluorescence quenching of Oregon Green-488 conjugated to recombinant nucleosomes at histone H4 Q27C. We find that full-length Set8 binds to nucleosomes with an apparent dissociation constant of 5.3 nM. In contrast, Set8(191–352) has 13% of nucleosome methyltransferase activity and we are unable to measure the very weak binding of Set8(191–352) to nucleosomes (Fig. 2). Examination of other Set8 truncations shows that Set8(153–352) has similar methyltransferase activity (113%) and nucleosome binding (Kd = 7.9 nM) to full-length Set8, while Set8(175–352) retains significant methyltransferase activity (102%), but binds nucleosomes much more weakly (Kd = 1.8 μM). A further deletion of 6 residues in Set8(181–352) reduces methyltransferase activity to 79% and binding to 4.8 μM. We also examined the Set8(153–352) H347F mutant that was shown to increase binding to H4 peptide approximately 20-fold [24], but we find that this mutant has similar nucleosome methyltransferase activity and a modest increase in binding affinity to nucleosomes (Kd = 4.5 nM versus 7.9 nM for the unmutated protein). These results indicate that Set8(153–352) constitutes a minimal domain that methylates and binds nucleosomes with significant contributions from residues 153–174 and 181–190 located N-terminal to the structured SET domain.

Use of RCC1 to crystallize the Set8/nucleosome complex

Based on this analysis of Set8 truncations, we selected Set8(153–352) containing the H347F mutation [Set8Δ1(H347F)] for structural studies. While we were able to reconstitute, purify, and concentrate monodisperse Set8Δ1/nucleosome complex, we were not able to grow crystals of the complex. As part of our efforts to explore alternate strategies for crystallizing the Set8/nucleosome complex, we considered using the RCC1 chromatin factor as a co-crystallization agent. This approach was inspired by the packing of individual RCC1/nucleosome complexes in the crystals used to determine the
RCC1/nucleosome X-ray structure [29,30]. The crystal packing leaves the histone H4 tail and regions adjacent to this tail accessible, suggesting the possibility that Set8 could bind to the nucleosome in the context of the crystalline RCC1/nucleosome complex. We first examined if Set8 could bind to RCC1/nucleosome in solution by comparing the elution of Set8\(^{\Delta 1}/\)nucleosome, RCC1/nucleosome, and Set8\(^{\Delta 1}/\)RCC1/nucleosome complexes by Superdex 200 size exclusion chromatography. We observe slightly faster elution of RCC1/nucleosome and nucleosome upon Set8\(^{\Delta 1}\) addition, and the SDS-PAGE analysis of the peak fractions confirms Set8, RCC1, and the nucleosome coelute as a complex (Supplementary Fig. 2).

In contrast to our unsuccessful Set8\(^{\Delta 1}/\)nucleosome crystallization trials, our Set8\(^{\Delta 1}/\)RCC1/nucleosome crystallization trials produced single crystals. To confirm the presence of Set8\(^{\Delta 1}\) in such crystals, we trace labeled Set8\(^{\Delta 1}\) with carboxyrhodamine [31–33], purified reconstituted Set8\(^{\Delta 1}/\)RCC1/nucleosome complex by size exclusion chromatography, set up crystallization trials, and examined resulting crystals with a fluorescence microscope (Supplementary Fig. 3a and b). The observed fluorescence indicated Set8\(^{\Delta 1}\) is present in the crystals. To further validate this conclusion, we washed individual crystals and examined the contents by SDS-PAGE. Bands for each of the RCC1, Set8\(^{\Delta 1}\), and individual histone proteins were detected by Coomassie Blue staining, and the carboxyrhodamine labeled the Set8 band was additionally detected by fluorescence (Supplementary Fig. 3c). These results indicated that we had crystallized a ternary complex of Set8\(^{\Delta 1}\), RCC1, and the nucleosome core particle. We note that we obtained crystals of Set8\(^{\Delta 1}/\)RCC1/nucleosome using yeast RCC1 (Srm1), but not the Drosophila RCC1 that we had used in our RCC1/nucleosome crystal structure [29,30].

Initial diffraction studies of these crystals produced diffraction to only 8–10 Å, but we were able to improve order in the crystals through post-crystallization dehydration soaks [30,34]. We collected and processed a 4.5 Å resolution data set, and performed molecular replacement using a polyalanine model of a yeast RCC1/nucleosome structure (Girish, Huang, Makde, and Tan, unpublished) and a polyalanine model of the Set8 SET domain (PDB coordinates 1ZKK, chain A). Two potential molecular replacement (MR) solutions were obtained with logarithm likelihood gain of 446 and 380 for MR models 1 and 2, respectively. Other MR solutions were not considered further since they included substantial steric clashes. Both models contain a pseudosymmetric half of the nucleosome and one molecule each of RCC1 and Set8 in the asymmetric unit. We manually corrected close contacts in COOT and performed rigid body refinement followed by restrained refinement of the models using REFMAC5 [35,36] to \(R_{\text{work}}/R_{\text{free}}\) of 33.9% and 38.7% for model 1 and \(R_{\text{work}}/R_{\text{free}}\) of 45.0%/47.0% for model 2. The electron density for RCC1 and the
histone cores is continuous and well defined in model 1 but significantly less so in model 2, while the electron density for Set8 is discontinuous and poorly defined in both models 1 and 2 (Supplementary Fig. 4). Attempts to further refine these crystallographic models were not successful.

The structure of yeast RCC1/nucleosome in Set8/RCC1/nucleosome MR models 1 and 2 are similar and resemble the Drosophila RCC1/nucleosome in RCC1 switchback loop/histone dimer acidic patch interactions (Fig. 3) [29]. While Set8 binds directly to the nucleosome in both models, the orientation of Set8 in the MR models 1 and 2 is very different, with different regions of Set8 in the two models interacting with the nucleosome. In both cases, no interactions are observed between Set8 and RCC1.

Set8/RCC1/nucleosome MR model 1

In MR model 1, the Set8 i-SET helix, which contains a cluster of three basic residues (K256, K257, and R258), is positioned in the DNA major groove of the nucleosome at SHL ± 1 (Fig. 3a and c). Resting above the adjacent DNA minor groove is the c-SET domain containing the basic residue K341 in position to interact with the DNA backbone. In contrast, the n-SET helix projects from the face of the SET domain opposite the nucleosome interface and makes no contact with histones or nucleosomal DNA. Complementing the Set8 i-SET and c-SET interactions with the nucleosome DNA are Set8–histone protein–protein interactions by the Set8 loop between strands β6 and β7 located above histone H4 α2 helix. Potential contacts include charged interactions between Set8 Lys280 and histone H4 Glu52, and additional hydrogen bonding interactions.

Although electron density for the histone H4 tail around target Lys20 is not visible, we can model the H4 tail based on crystal structures of the Set8 catalytic domain complexed with an H4 peptide spanning residues 16–23 [24,25]. The modeled H4 peptide is sandwiched between the Set8 SET domain on top and the nucleosome DNA at SHL ± 1.5 below, and is bordered on the sides by the i-SET and c-SET domains, which interact with nucleosomal DNA (Fig. 3e). The C-alpha atom of H4 Arg23 is 13 Å from the C-alpha atom of H4 Asp24 in this model, suggesting that if this model is correct, at least four residues starting from H4 D24 would need to be unraveled from the short helix extension to the H4 histone fold core. Aside from this, it appears that there are no other constraints limiting the H4 tail to bind in the Set8 active site.

Set8/RCC1/nucleosome MR model 2

The predominant interaction of Set8 with the nucleosome in MR model 2 is Set8’s n-SET domain, which lies across the nucleosome DNA major groove at SHL ± 2 (Fig. 3b and d). In addition, the loop between Set8 strands β9 and β10 may interact with the adjacent minor groove, and the C-terminal end of the n-SET domain is positioned over the histone H3 α1 helix. Neither the Set8 i-SET nor the c-SET domains are close to the nucleosome, and no
interactions between Set8 with the histone H4 α2 helix are evident.

The H4(16–23) peptide modeled in MR model 2 is located on a Set8 surface that faces away from the nucleosome and is solvent exposed (Fig. 3f), in contrast to the H4 peptide modeled in MR model 1. The H4 Arg23 C-alpha atom is 23 Å from the H4 Asp24 C-alpha atom, but even more problematic than this distance are the relative locations of these two residues. The Set8 SET domain is positioned between the H4 histone fold and the modeled H4 peptide, blocking a path for the H4 tail to bind to Set8, thus increasing the effective distance between the nucleosome surface and the Set8 active site, and raising the question of how the H4 peptide around Lys20 could reach into the Set8 active site in MR model 2.

Basic residues in i-SET and c-SET helices are important for Set8 methyltransferase activity and nucleosome binding

Our crystallography studies provided two distinct structural models for how Set8 binds to the nucleosome, but these studies did not unambiguously distinguish between the two models. While MR model 1 appears more plausible because the Set8 position is consistent with previous structural information for how Set8 binds to its histone H4 substrate and because of the better crystallographic statistics and electron density, we sought additional evidence to distinguish between the two models. Since Set8 employs the n-SET, the i-SET, and the c-SET regions differently to bind to the nucleosome in the two structural models, we engineered mutations into these regions (Fig. 4a and b) and tested the activity of the mutant Set8 proteins in nucleosome methyltransferase and nucleosome binding assays (Fig. 5). We used the same Set8(153–352) H347F construct used for our crystallographic studies for this analysis. We focused on basic residues because each of these three regions interacts with nucleosomal DNA in the structural models. We find that mutating basic residues in the n-SET, i-SET, and c-SET domains slightly decrease nucleosome binding affinity by a factor of only 3.5- to 5-fold (Fig. 5). Mutations in the n-SET and c-SET domains did not significantly affect Set8’s nucleosome methyltransferase activity, but mutations in the i-SET domain decrease methyltransferase activity by a factor of two. Combining the i-SET and c-SET mutations further decreased methyltransferase activity (32% of wild type) and significantly weakened nucleosome binding ($K_d = 330 \text{ nM}$). We also prepared hSet8∆1(K280A,H347F), but the
aggregation of this mutant protein prevented us from testing the predicted contact between Set8 Lys280 and histone H4 Glu52. Since model 1 is more consistent with how Set8 binds the histone H4 tail substrate (model 2 does not provide a plausible means for Set8 to bind the H4 tail) and since mutations designed to disrupt model 1 interfaces have a larger effect on Set8 enzymatic and substrate binding activity, we believe that model 1 better describes how Set8 binds to the nucleosome. Model 1’s direct involvement of the Set8 i-SET region in nucleosome binding is also consistent with the variable i-SET region contributing substrate specificity among the SET domain proteins.

Role of Set8 basic region preceding SET domain in nucleosome substrate binding

Although the Set8 construct we used for our crystallization studies, Set8(153–352) H347F, was extended by 40 residues N-terminal to the SET domain (residues 193–352), we did not observe electron density for this region. Nevertheless, our deletion analysis suggests that this region is important for interaction with the nucleosome since Set8(175–352) and Set(181–352) were approximately 200- to 600-fold weaker in nucleosome binding than Set8(153–352), and Set8(191–352) fails to bind the nucleosome altogether (Fig. 2). The major feature of Set8 between residues 153 and 192 is its basic nature, with 14 Arg or Lys residues (35% of 40 residues) and only 2 acidic residues (5% of 40 residues) (Fig. 6a). Among the basic residues in the Set8(175–191) region critical for nucleosome binding are R179, K180, R188, and R189. We find that the Set8(153–352) R188A,R189A mutant has reduced nucleosome methyltransferase and much lower nucleosome binding activity (56% of full-length, wild-type Set8 enzymatic activity, and 2.5 μM nucleosome binding dissociation constant or over 300-fold lower binding affinity compared to the equivalent wild-type protein) (Fig. 6b). In contrast, the effect of the R179A,K180A mutations is much more muted with similar enzymatic activity as the wild-type equivalent and about 2.5× decrease in nucleosome binding affinity. The combination of R179A,K180A with R188A,R189A produced similar enzymatic activity and even weaker nucleosome binding as compared to the R188A,R189A mutant. We note that the effect of the R188A,R189A mutations on nucleosome binding was larger than the effect of the combined i-SET and c-SET basic mutations in the structured SET domain. Thus, our results suggest a major role for R188,R189 residues and a minor role for R179,K180 residues in nucleosome substrate recognition by Set8. Binding experiments using the fluorescent probe installed on a different position (H4 G56C within the histone fold) produced similar trends in binding for the subset of Set8 variants examined, suggesting that our conclusions are not dependent on the location of the probe at H4 Q27C near the H4 N-terminal tail (Supplementary Table 2).

Fig. 5. Site-directed mutagenesis of Set8 n-SET, i-SET, and c-SET regions to distinguish between molecular replacement models 1 and 2. The combinations of mutations targeting the i-SET and c-SET regions predicted to bind to nucleosomal DNA in model 1 significantly reduced HMTase and nucleosome binding activities, whereas mutations in the n-SET region implicated in model 2 had only minor effects.
interactions and not due to interactions within the SET domain [24,25], Set8 residues 190–202 could bridge the distance to the acidic patch as an extended polypeptide chain. In contrast, in model 2, the N terminus of the Set8 n-SET domain is pointing away from the histone disk toward nucleosomal DNA (Fig. 4d). We therefore examined the effect of removing this histone dimer acidic patch on the Set8’s activity. We find that full-length Set8 has only 3.8% of its HMTase activity on nucleosomes containing the quadruple H2A E61A,E64A,D90,E92A mutation compared to wild-type nucleosomes (Fig. 6b). This is a larger effect on methyltransferase activity than any of the Set8 truncations or mutations that we analyzed. Similarly, Set8Δ1 lacking the N-terminal 152 residues has 5.4% HMTase activity on nucleosomes with the same acidic patch mutation. The acidic patch mutations had a significantly greater effect on nucleosome binding for the Set8Δ1 truncation (Kd = 1.3 μM or 160-fold weaker binding compared to wild-type nucleosomes) than it did for full-length Set8 (Kd = 60 nM or 11-fold weaker binding compared to wild-type nucleosomes). We find that combining the Set8Δ1(R188A,R189A) basic mutations with the nucleosome acidic patch mutations results in only slightly weaker binding to the nucleosome compared to the R188A,R189A mutations or the nucleosome acidic patch mutations on their own. The finding that the Set8 R188A,R189A basic mutations and the nucleosome acidic patch mutations each severely affect nucleosome binding but the combination of the mutations is not additive suggests the possibility that the Set8 R188A,R189A basic region interacts with the nucleosome acidic patch.

The Set8 R188A,R189A basic region mediates competitive binding of Set8 to PCNA and to the nucleosome

Set8 protein levels are regulated by ubiquitylation-mediated proteolysis during the cell cycle by binding to PCNA [18–22]. Set8 contains two PIP or PCNA interaction boxes, but only the second PIP box is required for PCNA-dependent ubiquitylation of Set8 [18,21,22]. This second PIP box is located in the region that includes Set8 residues 178 to 190 (Fig. 6a). This directly overlaps the same Set8 basic region that we identified as playing a critical role in the Set8 enzyme binding to its nucleosome substrate. The crystal structure of the homologous PIP box peptide from the p21 protein bound to PCNA shows extensive interaction between the PIP box and PCNA[43]. Specifically, PCNA interacts with the p21 peptide residues that correspond to Set8(178–190), including the p21 R155 equivalent of Set8 R189. Prompted by this observation that the same Set8 basic region apparently interacts with both PCNA and the nucleosome, we asked if Set8 can bind simultaneously with both PCNA and the nucleosome.

We used the same nucleosome-binding assay that monitors quenching of Oregon Green-488 conjugated to H4 Q27C nucleosomes and examined the ability of PCNA to compete Set8 from the nucleosome. In this assay, quenching is observed upon Set8 binding to the nucleosome. If Set8 is competed

Fig. 6. Critical role of Set8 basic regions N-terminal to the SET domain. (a) Amino acid sequence of Set8(153–200). The PIP box is shown as a red box and the conserved PIP residues are underlined in red. Basic residues R179, K180, R188, and R189 within the PIP box are shown in blue, while other basic residues, N-terminal to the n-SET helix, are shown in black. The location of the n-SET helix observed in crystal structures is shown as a blue box. (b) Mutations in the Set8 basic region highlight the critical role of Set8 R188 and R189 and the nucleosome acidic patch in Set8/nucleosome interactions. Acidic patch mutant nucleosomes contain an N-terminal hexahistidine tag. Similarly tagged wild-type nucleosomes have no effect on nucleosome binding or methylation (data not shown).
off the nucleosome by the addition of PCNA, a
dequenching of this quenched fluorescence should
be observed, leading to fluorescence levels compara-
table to the nucleosome alone in the absence of
Set8. As expected, adding PCNA to the nucleo-
somes did not change the fluorescence of the
labeled nucleosomes since PCNA is not anticipated
to interact with the nucleosome. When a saturating
amount of Set8 is added to nucleosomes, the
binding of Set8 is detected as quenching of the
nucleosome fluorescence signal (normalized fluo-
rescence change of 1.0 in Fig. 7). Titrating PCNA
into this pre-formed Set8/nucleosome complex
results in an increase in the fluorescence (returning
the relative fluorescence change to 0 in Fig. 7). We
confirmed that this fluorescence dequenching was
due to the direct binding of PCNA to Set8’s PIP
domain using a Set8 PIP domain F184A,Y185A
mutant shown to be defective for binding to PCNA
[18,21]. In contrast to wild-type Set8, PCNA was
severely compromised in its ability to compete this
PIP mutant Set8 off the nucleosome, while the Set8
mutant’s nucleosome binding was only minimally
affected (data not shown). These results suggest
that the Set8 PIP domain mediates interactions with
both PCNA and the nucleosome. While it is possible
that Set8, PCNA, and the nucleosome can form a
ternary complex, such an interpretation would
require Set8 to not quench the labeled nucleosome
only in the context of the ternary complex. While we
were unable to rule this out, we believe that such a
scenario is unlikely given that the labeled position on
the nucleosome is proximal to the nucleosomal
binding site of Set8’s SET domain near the H4 tail
and distant from the acidic patch where the PIP
domain is expected to bind.

Discussion

We have combined crystal diffraction data and
biochemical characterization to determine a structural
model for how the Set8 HMTase enzyme interacts
with its nucleosome substrate. Our model indicates
that the relatively compact Set8 protein uses at least
three distinct regions to interact with the nucleosome.
A basic N-terminal extension is the primary determi-
nant of nucleosome binding and interacts with the
nucleosomal H2A/H2B histone dimer acidic patch and
likely nucleosomal DNA to anchor Set8 to the
nucleosome. This positions the SET domain for
interaction with the nucleosome face near the targeted
methylation site. Finally, the Set8 catalytic site
engages the H4 tail and methylates lysine 20. The
multivalent interactions with the nucleosome and particulary nucleosomal DNA observed or predicted
by our studies provide a molecular basis for why the
Set8 enzyme greatly prefers a nucleosome substrate
over histone proteins or peptides.

We have used a novel approach employing the
chromatin factor RCC1 as a co-crystallization protein
for crystallizing the Set8/nucleosome complex. Using
the diffraction data we collected from these crystals,
we obtained two crystallographic MR solutions using
Set8 and RCC1/nucleosome structures as search
models. We then challenged these models by
examining the consequences of mutating Set8
residues in key nucleosome interfaces. Our results
are consistent with Set8 using its i-SET and c-SET
regions to interact with nucleosomal DNA 1 to 1.5
DNA helical turns from the nucleosomal dyad.
Notably, the mutation of the i-SET region alone or in
combination with the c-SET impairs enzymatic activity
to a greater degree than nucleosome binding when

Fig. 7. PCNA competes with the
nucleosome for binding to Set8 via
the Set8 PIP box. Nucleosome core
particles (NCP) fluorescently la-
beled on H4 Q27C with Oregon
Green 488 were incubated with
buffer (black), Set8Δ1 (red), and
Set8Δ1 with the PIP F184A,Y185A
mutations shown to disrupt binding
to PCNA (blue) were titrated with
PCNA. A normalized fluorescence
change value of 1.0 represents full
binding of nucleosomes by Set8,
and a value of 0 signifies no binding.
The decrease in fluorescence signal
of Set8Δ1/NCP upon incubation
with PCNA demonstrates that
PCNA sequesters wild-type Set8
from the nucleosome, while Set8
defective for binding to PCNA con-
tinues to bind the nucleosome in the
presence of PCNA.
compared to truncations or mutations of Set8 regions adjacent to the SET domain. One possible explanation is that the SET domain nucleosome interactions do not function in Set8 recruitment, but rather orient Set8 on the nucleosome surface for optimal catalysis. Our preferred model is broadly consistent with previous crystal structures of the Set8 SET domain bound to its substrate histone H4 peptide, given the correct orientation of the Set8 catalytic site relative to the H4 histone fold core. However, the depth of the Set8 catalytic cleft requires at least four residues of the H4 tail that are typically structured in nucleosome crystal structures (residues 24–27) to unfold and extend to reach the Set8 active site. We note that the unraveling of the H4 tail is not a special requirement of our particular structural model. Given how embedded the H4 tail is within the Set8 catalytic cleft, it is difficult to explain how Set8, in any model, could bind to the H4 tail without unfolding some of the structured H4.

Although the Set8 catalytic SET domain interacts directly with the nucleosome, these interactions are not sufficient for detectable nucleosome binding. We have identified a positively charged N-terminal extension to the SET domain spanning residues 153–190 that is required to restore nucleosome binding and methylation by the SET domain to levels comparable to the full-length protein. Further dissection of this region implicates sequences between 181–190 and 153–174. Within the Set8(181–190) sequence, the Set8 R188,R189 basic cluster plays a significant role in nucleosome binding and activity. The nearby R179,K180 basic cluster contributes only minimally, suggesting that the interactions mediated by Set8 R188,R189 residues with the nucleosome are more than simply nonspecific electrostatic interactions. We provide evidence that Set8 R188,R189 likely interacts with the histone H2A/H2B dimer acidic patch on the nucleosome: mutation of the nucleosome acidic patch has a similar deleterious effect on Set8 binding as the Set8 R188A,R189A mutation, while combining these mutations (Set8 R188A,R189A binding to nucleosomes mutated in the acidic patch) weakens binding only slightly further. We would have expected an additive effect on nucleosome binding mutations if the Set8 and histone mutations affected separate binding surfaces. Interestingly, the mutation of the acidic patch results in a larger impairment of enzymatic activity than the mutation of the Set8 R188,R189 cluster. This suggests that the elimination of the acidic patch may result in an additional unanticipated effect on nucleosome methylation possibly through reorienting Set8 on the nucleosome surface. Previous studies reporting that Set8 does not rely on the acidic patch for nucleosomal methylation may have missed this effect due to the use of significantly higher Set8 concentrations or single-point mutation of the acidic patch [44]. Binding of the nucleosome acidic patch by a chromatin protein arginine residue (the “arginine anchor”) has been observed in every chromatin protein/nucleosome crystal structure determined to date (complexes of the nucleosome with RCC1, Sir3, and the PRC1 ubiquitylation module proteins, and with the LANA and CENP-C peptides). We propose here that Set8 also employs an arginine anchor to bind to the nucleosome.

The Set8 arginine residues that may act as an arginine anchor to bind to the nucleosome acidic patch are included within the PIP box of Set8. We show that PCNA and the nucleosome compete for binding to Set8, consistent with the extensive interactions made by the homologous PIP box of p21 with PCNA, including direct interactions between arginine residues equivalent to Set8 R188 and R189. Given that PCNA binding to Set8 is required for Set8 ubiquitylation and subsequent proteolytic degradation, this finding suggests an eloquent mechanism for ensuring that Set8 is protected from proteolytic degradation when it is engaged on the nucleosome. As long as Set8 is bound to the nucleosome, the signal to recruit PCNA is obscured and proteolysis is prevented. Our results also exemplify how a PIP box can serve multiple functions: it is utilized to bind to PCNA but it also contributes to nucleosome binding in the case of Set8.

Our deletion analysis suggests that Set8(153–174) also contributes to nucleosome binding as we detect a decrease in binding affinity from ~8 to 1800 nM when this region is deleted. This effect is comparable to that observed for the R188A,R189A basic cluster mutation. As shown in Fig. 6a, Set8(153–174) contains additional basic amino acids (8 lysine and 1 arginine residues). Since we believe that the PIP box that immediately follows this region binds to the histone dimer acidic patch situated next to nucleosomal DNA, it seems likely that the basic residues within Set8(153–174) may interact with nucleosomal DNA near the acidic patch (approximately five helical turns from the nucleosome dyad). Interestingly, unlike the mutation of the PIP box basic cluster, the truncation of Set8(153–174) did not appreciably affect HMTase activity on nucleosome substrates. Thus, this region may similarly enhance Set8–nucleosome affinity, but unlike the proposed PIP box-acidic patch, interaction plays a minimal role in orienting the SET domain for catalysis.

Set8’s multivalent interactions with the nucleosome may also explain why the electron density for Set8 in our Set8/RCC1/nucleosome crystallographic electron density maps was so poor, and why we did not observe electron density for the Set8 basic residues interacting with the nucleosome acidic patch. As noted in the Results section, electron density for RCC1, as well as for the histone and DNA components of the nucleosome, was far better than the electron density for Set8 particularly in MR model 1. Based on our studies suggesting that Set8 and
RCC1 both bind to the nucleosome acidic patch, we suspect that the occupancy of Set8 in the crystal may be less than the expected two Set8 molecules per nucleosome due to competition with RCC1 for binding to the nucleosome. Since Set8 is able to make multivalent interactions with the nucleosome via the Set8 basic N-terminal extension and its SET domain, Set8 could presumably form a triple complex with RCC1 and the nucleosome even if interactions via the acidic patch were blocked by RCC1. Another possibility is that Set8 is present in stoichiometric amounts in the crystal, but the competition between RCC1 and Set8 for the acidic patch leaves Set8 less structurally constrained when bound via the SET domain to the nucleosome. We have tried without success to substantiate these speculations. For example, comparison of SDS-PAGE band intensities of purified Set8/RCC1/nucleosome complex with washed Set8/RCC1/nucleosome did not provide definitive results for the stoichiometry of Set8 to nucleosome. The competition binding experiments show that RCC1 can compete for nucleosome binding with Set8 (Supplementary Fig. 1c). However, unlike PCNA, which neutralizes the PIP box, RCC1 is unable to completely remove Set8 from the nucleosome, likely due to the multivalent nature of the Set8–nucleosome interaction.

The Set8/nucleosome model adds to a small but growing number of structural descriptions of how chromatin factors and enzymes interact with the nucleosome. A common feature is that the chromatin factors and enzymes recognize not just individual elements of the nucleosome, but also the architecture of the nucleosome. In the case of RCC1, the PRC1 ubiquitylation enzyme and now Set8, interactions are made to both histone and DNA components of the nucleosome ensuring that binding occurs in a chromatin context. Despite these broad similarities and the apparent common use of an arginine anchor, the details of the interactions are different, reflecting the diverse functions including, in the case of PRC1 and Set8, the different histone tails (H2A C-terminal tail for PRC1 and the H4 N-terminal tail for Set8) targeted by the chromatin enzymes. Thus, although paradigms such as the arginine anchor may guide our understanding of how chromatin enzymes bind to the nucleosome, many more unique features are likely to govern how each of these enzymes targets catalysis to unique nucleosomal surfaces.

### Material and Methods

**Recombinant proteins and nucleosomes**

The human Set8 gene (UniProtKB Q9NQR1) and truncation or mutated variants were expressed as a combination of Strep tag–decahistidine-tagged protein using the pST50Tr expression vector [45] in BL21(DE3)-pLysS cells at 18 °C. Each fusion protein was purified by Talon metal affinity chromatography (Clontech), the fusion affinity tag was removed by tobacco etch virus (TEV) protease, and the untagged protein was further purified by SourceS cation-exchange chromatography (GE Healthcare). The yeast RCC1 gene, Smr1, was amplified from genomic DNA and similarly expressed using the pST50Tr expression vector in BL21(DE3)pLysS cells at 18 °C. Tagged yeast RCC1 was purified using Talon metal affinity chromatography, followed by TEV protease digestion to remove the affinity tag, and the untagged protein further purified by SourceQ anion-exchange chromatography (GE Healthcare). Human PCNA was expressed in pST50Tr with an N-terminal (Gly-Ser), linker between the Strep-decahistidine combination tag and PCNA to facilitate TEV protease cleavage. After Talon metal affinity purification and TEV treatment, the GS40PCNA protein was purified by SourceQ anion-exchange chromatography. Recombinant *Xenopus* histones and nucleosome core particles were prepared essentially as described previously, including SourceQ HPLC of the reconstituted nucleosome core particles. [46]. All proteins were analyzed by dynamic light scattering and determined to be non-aggregated. Selected Set8 proteins used for crystallization were N-terminally labeled with carboxyfluorescein as previously described [31–33]. Briefly, Set8 was reacted with two molar equivalents of 5-(and 6-carboxyfluorescein succinimidyl ester (Life Technologies) at pH 7.5 followed by purification using a Superdex 200 column. Typically, labeling percentages of approximately 3%–5% were achieved.

**Reconstitution and purification of Set8/RCC1/nucleosome complex**

To generate the Set8/RCC1/nucleosome complex, we mixed yeast RCC1(2–482) protein in 10 mM Tris–Cl (pH 7.6), 50 mM sodium chloride, 1 mM dithiothreitol (DTT) with recombinant nucleosome core particles containing Xenopus core histones, and 149 bp Widom 601 DNA in a 2.2:1 RCC1/nucleosome molar ratio followed by the addition of human Set8 protein (153–352) H347F in a 2.2:1 Set8/nucleosome molar ratio. The reconstituted complex was then purified on a Superdex 200 HR size-exclusion chromatography column (GE Healthcare) in 10 mM Tris–HCl (pH 7.5), 50 mM sodium acetate, 1 mM DTT, and 0.1 mM PMSF, and concentrated by filtration centrifugation in the same buffer. For complexes containing trace-labeled Set8, N-terminal carboxyfluorescein-labeled Set8 was doped into reconstitutions to give a 0.5% final labeling percentage.

**Crystallization and structure solution**

Crystals of the Set8/RCC1/nucleosome complex were grown in microbatch by mixing 1 μl of 10 to 15 mg/ml complex in 10 mM Tris–HCl (pH 7.5), 50 mM sodium acetate, 1 mM DTT, and 0.1 mM PMSF with 1 μl of 25 mM sodium acetate (pH 5.5), 25 to 50 mM sodium citrate, 1 mM DTT, and 2% to 8% PEG2000-MME, and then overlaid with 75 μl of Al2O3 oil (an equal-volume mixture of silicon and paraffin oils) at 4 °C in 96-well microtiter plates. Crystallization drops were visualized using a Nikon SMZ1000 stereo microscope equipped with a CoolPix 990 SMZ1000 stereoscope.
camera and the incorporation of trace fluorescently labeled Set8 was confirmed using a Nikon EpiFluorescence microscope equipped with a Nikon Intensilight C-HGFI light source, a DsRed filter (Ex 545/30 nm, DM 570 nm, BA 620/60 nm), and a CoolSnapEZ Turbo 1394 camera. To further validate crystal contents, single crystals were washed in stabilizing solution prior to dissolution and separation by SDS-PAGE, fluorescence scanning using a Typhoon 9410 variable mode imager ($\lambda_{ex}$ = 532 nm and $\lambda_{em}$ = 580 nm) from GE Healthcare, and Coomassie staining. Crystals were dehydrated following the soaking protocol previously described for Drosophila RCC1/nucleosome complex crystals [29,30]. The optimized soaking solution contained 25 mM sodium acetate (pH 5.5), 35 mM sodium citrate, 1 mM DTT, 5% ethanol, 8% PEG2000-MME, and a stepwise increase in PEG500-DME from 0–24% in 2% increments of 15 min before flash cooling in liquid nitrogen. Diffraction data were collected with an ADSC Quantum 315 CCD detector at Advanced Photon Source's NE-CAT beamline 24-ID-E and processed with the XDS program (Supplementary Table 1). MR was performed in PHASER [47] using polyalanine models of the yeast RCC1/nucleosome complex (unpublished) and human Set8 (PDB 1ZKK, chain A). The MR models were evaluated for packing clashes using COOT. After several rounds of rigid body refinement, MR models were restrained refined using REFMAC5 [35,36]. However, further attempts of model building and refinement were unsuccessful due to poor quality of the electron density map.

HMTase assay

HMTase assays were carried out in triplicate in a final volume of 10 μl, containing 2 μl of 20 μM recombinant nucleosome core particles and 1.2 μl of 3H-labeled S-adenosylmethionine (MP Biomedicals) in 60 mM Tris–HCl (pH 8.0), 50 mM NaCl, 20% glycerol, 1 mM DTT, and 0.1 mM PMSF. The reaction was started by the addition of 2 μl of 10 nM Set8 and incubated at 30 °C for 30 min before 8 μl of the reaction mixture was spotted onto P81 phosphocellulose filter paper (Whatman). Dried filter disks were washed 3 times in 50 mM sodium carbonate (pH 9.0) and dried before scintillation counting. The wild-type and mutant Set8 enzymes were diluted to ensure that the measured activity remained in the linear range of the assay.

High-throughput interactions by fluorescence intensity nucleosome-binding assay

Clear bottom 384-well microplates were prepared as published previously [48]. Xenopus histone H4Q27C protein was labeled with Oregon Green 488 maleimide and reconstituted into recombinant nucleosomes. Of these fluorescently labeled nucleosomes, 2 nM was titrated with 2 nM–20 μM Set8 protein in 40 μl reaction volumes containing 20 mM Tris–HCl (pH 7.5), 125 mM NaCl, 5% glycerol, 5 mM DTT, 0.01% of both CHAPS [3-(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate, and NP-40 (nonyl phenoxypolyethoxylethanol) detergents, and 0.1 mg/ml bovine serum albumin for 15 min at room temperature in the dark before scanning at the appropriate wavelength ($\lambda_{ex}$ = 488 nm and $\lambda_{em}$ = 526 nm) using a Typhoon 9410 variable mode fluorimagre. Fluorescence intensities were quantified using the program ImageQuant TL, and the data points were fitted with a nonlinear regression curve using PROFIT software. All experiments were performed in triplicates. Due to the use of 2 nM-labeled nucleosomes in binding experiments, $K_d$ values below 10 nM may slightly underestimate affinity.

Accession Numbers

Crystallographic structure factors and coordinates for MR model 1 have been deposited in the Protein Data Bank with accession number 5HQ2.

Acknowledgments

We would like to thank the staff of APS NE-CAT beamlines 24ID-E and 24ID-C for their assistance during synchrotron data collection; Hemant Yennawar and Neela Yennawar at the Penn State Huck Institute X-ray core facility; Allen Minns and Mike Moore for reagent preparation; Jiehuan Huang, Matt Jennings, and Sang-Ah Kim for assistance with crystallographic data collection; and members of the Tan laboratory and the Penn State Center for Eukaryotic Gene Regulation for helpful discussions. This work was supported by the National Institutes of Health grants GM088236 and GM111651 to S.T. and Damon Runyon Cancer Research Foundation grants DR-2107-12 and DFS-14-15 to R.K.M.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2016.02.025.

Received 4 September 2015; Received in revised form 3 February 2016; Accepted 27 February 2016
Available online 4 March 2016

Keywords:
chromatin biology; epigenetic; histone modifications; SET domain; X-ray crystallography

Abbreviations used:
PCNA, proliferating cell nuclear antigen; MR, molecular replacement; PIP, PCNA-interacting peptide; TEV, tobacco etch virus; HMTase, histone methyltransferase.
References


Supplementary Table 1: Crystallographic statistics

### Data collection and processing statistics

<table>
<thead>
<tr>
<th>Data collection and processing statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>C2221</td>
</tr>
<tr>
<td>Unit cell dimensions a, b, c (Å)</td>
<td>100.73, 300.75, 182.09</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90.0, 90.0, 90.0</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>46.19 – 4.50 (4.74 - 4.50)*</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.2</td>
</tr>
<tr>
<td>Reflections (total/unique)</td>
<td>122385/16854</td>
</tr>
<tr>
<td>Mean I/σ(I)</td>
<td>12.9 (2.4)</td>
</tr>
<tr>
<td>Completeness</td>
<td>99.8 (99.8)</td>
</tr>
<tr>
<td>Rmerge(%)</td>
<td>9.5 (85.2)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>7.3(7.4)</td>
</tr>
</tbody>
</table>

### Refinement statistics

<table>
<thead>
<tr>
<th>Refinement statistics</th>
<th>model 1</th>
<th>model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>4.50</td>
<td>4.50</td>
</tr>
<tr>
<td>Rwork/Rfree (%)</td>
<td>33.9/39.7</td>
<td>45.0/47.0</td>
</tr>
<tr>
<td>RMSD bond length (Å)</td>
<td>0.007</td>
<td>0.008</td>
</tr>
<tr>
<td>RMSD bond angle (°)</td>
<td>0.990</td>
<td>1.4</td>
</tr>
<tr>
<td># non-hydrogen protein atoms</td>
<td>4767</td>
<td>4785</td>
</tr>
<tr>
<td># non-hydrogen DNA atoms</td>
<td>2971</td>
<td>3091</td>
</tr>
</tbody>
</table>

* Values in parentheses are for highest-resolution shell
<table>
<thead>
<tr>
<th></th>
<th>$K_d$(H4 Q27C)</th>
<th>vs wt</th>
<th>$K_d$(H4 G56C)</th>
<th>vs wt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set8Δ1</td>
<td>7.9 ± 0.6 nM</td>
<td>1.0</td>
<td>16.5 ± 3.9 nM</td>
<td>1.0</td>
</tr>
<tr>
<td>Set8Δ1(R179A,K180A)</td>
<td>20 ± 2 nM</td>
<td>2.6</td>
<td>54 ± 26 nM</td>
<td>3.3</td>
</tr>
<tr>
<td>Set8Δ1(R188A,R188A)</td>
<td>2500 ± 400 nM</td>
<td>320</td>
<td>5600 ± 2900 nM</td>
<td>340</td>
</tr>
<tr>
<td>Set8Δ1(K256A,K257A,R258A,K341A,H347F)</td>
<td>330 ± 50 nM</td>
<td>42</td>
<td>81 ± 24 nM</td>
<td>4.9</td>
</tr>
</tbody>
</table>
Suppl. Fig. 1: Nucleosome binding data for (a) Set8 domain mutations, (b) Set8 basic region mutations, and (c) titration of Set8/nucleosome with RCC1.

(a) and (b) Nucleosome fluorescently labeled with Oregon Green 488 on H4Q27C were titrated with Set8 variants in triplicate, and the normalized fluorescence change plotted as a function of Set8 concentration. The calculated dissociation constants are shown above the binding curves. (c) Nucleosome binding data for Set8Δ1/nucleosome or nucleosome titrated with RCC1 show RCC1 can partially compete with Set8Δ1 for binding to nucleosomes.
Suppl. Fig. 2: Set8, RCC1 and the nucleosome core particle (NCP) can form a ternary complex.
(a) The Set8/RCC1/NCP (red), RCC1/NCP (blue), Set8/NCP (green) and NCP only (black) samples were separated over a Superdex 200 HR size exclusion chromatography column. The elution fractions and times are shown above and below the x-axis respectively. (b) SDS-PAGE gel of select fractions from the chromatography runs shown in part (a). Molecular weights of standards are shown on the right.
Suppl. Fig. 3: Analysis of Set8/RCC1/nucleosome crystals.
(a) and (b) Crystals in Set8/RCC1/nucleosome trials visualized by (a) light microscopy or (b) fluorescence microscopy. The Set8 protein was trace-labeled with carboxyrhodamine red, (c) Washed Set8/RCC1/nucleosome crystal analyzed by Coomassie staining (left) and fluorescence (right).
(d) and (e) Positive control showing RCC1/nucleosome crystals grown using RCC1 trace-labeled with carboxyrhodamine, visualized by (d) light microscopy or (e) fluorescence microscopy. Similar unlabeled crystals were used to determine the crystal structure of the RCC1/nucleosome complex (refs. 28 & 29).
(f) and (g) Negative control showing RCC1/nucleosome crystals grown using unlabeled RCC1, visualized by (f) light microscopy or (g) fluorescence microscopy.
Suppl. Fig. 4: Stereo 2F_o-F_c electron density maps for Set8/RCC1/nucleosome molecular replacement (MR) models 1 and 2 contoured at 1σ cutoff.

(a) In MR model 1, clear and continuous electron density is visible for RCC1 and for the histone polypeptide backbone and DNA components of the nucleosome, including histone H4 α2 helix, but poor and discontinuous density was observed for Set8 including the pre-SET, i-SET and post-SET regions. (b) In MR model 2, discontinuous density is visible for RCC1 and the nucleosome, and even poorer discontinuous density is visible for Set8. In contrast to model 1, the electron density for the histone H4 α2 helix is not visible at this contour.