DNA Repair 115 (2022) 103341

Contents lists available at ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

The C-terminal domain of *Arabidopsis* ROS1 DNA demethylase interacts with histone H3 and is required for DNA binding and catalytic activity

J.T. Parrilla-Doblas [a,b,c], T. Morales-Ruiz [a,b,c], R.R. Ariza [a,b,c], M.I. Martínez-Macías [a,b,c,*], T. Roldán-Arjona [a,b,c,*]

[a] Department of Genetics, University of Córdoba, Spain
[b] Maimónides Biomedical Research Institute of Córdoba (IMIBIC), Spain
[c] Reina Sofía University Hospital, Spain

ARTICLE INFO

Keywords:
DNA glycosylases
DNA methylation
DNA demethylation
Histone modifications
Histone reader domains
Epigenetics

ABSTRACT

Active DNA demethylation plays an important role in controlling methylation patterns in eukaryotes. In plants, the DEMETER-LIKE (DML) family of 5-methylcytosine DNA glycosylases initiates DNA demethylation through a base excision repair pathway. However, it is poorly understood how these DNA demethylases are recruited to their target loci and the role that histone marks play in this process. *Arabidopsis* REPRESSOR OF SILENCING 1 (ROS1) is a representative enzyme of the DML family, whose members are uniquely characterized by a basic amino-terminal domain mediating nonspecific binding to DNA, a discontinuous catalytic domain, and a conserved carboxy-terminal domain of unknown function. Here, we show that ROS1 interacts with the N-terminal tail of H3 through its C-terminal domain. Importantly, phosphorylation at H3 Ser28, but not Ser10, abrogates ROS1 interaction with H3. Conserved residues at the C-terminal domain are not only required for H3 interaction, but also for efficient DNA binding and catalytic activity. Our findings suggest that the C-terminal domain of ROS1 may function as a histone reader module involved in recruitment of the DNA demethylase activity to specific genomic regions.

1. Introduction

DNA methylation at carbon 5 of cytosine (5-methylcytosine, 5-meC) is a stable but reversible modification, usually associated with gene silencing, that functions as an epigenetic mark in embryonic development, X-chromosome inactivation, imprinting, and control of transposon activity [1]. In mammals, DNA methylation mainly occurs in the symmetric CG context, although non-CG methylation has been reported in brain tissues and embryonic stem cells [2,3]. Plant DNA methylation occurs at any cytosine sequence context: CG, CHG, or CHH (H = A, T, C) [1]. DNA methylation patterns, which change during normal development, are controlled by the antagonistic actions of methylation and demethylation pathways. DNA demethylation can be either passive or active. Passive demethylation involves dilution of 5-meC after DNA replication cycles in the absence of maintenance methylation. Active demethylation occurs independently of replication and involves one or more enzymes [4]. Despite the knowledge gained on the mechanisms of DNA methylation, how active DNA demethylation is regulated and how the proteins involved in such process are recruited to target loci remain open questions.

Plants possess a distinctive active DNA demethylation mechanism involving DNA glycosylases that directly excise 5-meC and initiate its replacement with unmodified C through a base excision repair (BER) pathway [5]. Plant 5-meC DNA glycosylases are typified by *Arabidopsis* ROS1 (REPRESSOR OF SILENCING 1) [6–8] and its paralogs DME (DEMETER) [7,9], DML2, and DML3 (DEMETER-LIKE proteins 2 and 3) [5,7,10,11]. These proteins are grouped in the DEMETER-LIKE (DML) family, which belongs to the HhH-GPD superfamily, the largest and most functionally diverse group of DNA glycosylases [12].

Arabidopsis ROS1 is a multi-domain bifunctional DNA glycosylase with a bipartite catalytic domain divided by a large insert predicted to have an unstructured conformation, a short N-terminal domain significantly rich in lysine, and a large C-terminal domain exclusively conserved among DML family members. Sequence alignment and available structural data of HhH-GPD enzymes allowed generating a tridimensional model of the discontinuous catalytic domain of ROS1

* Corresponding authors at: Department of Genetics, University of Córdoba, Spain.
E-mail addresses: q92mamam@uco.es (M.I. Martínez-Macías), ge2roarm@uco.es (T. Roldán-Arjona).

https://doi.org/10.1016/j.dnarep.2022.103341
Received 1 March 2022; Received in revised form 13 April 2022; Accepted 3 May 2022
Available online 6 May 2022

[13]. Together with biochemical analysis, this model has been used to identify residues important for ROS1 function [13]. The N-terminal basic domain is not essential for catalytic activity [14,15], but in ROS1 mediates methylation-independent binding to DNA and endows the protein with the capacity to slide along DNA substrates in search of 5-meC [15,16]. The C-terminal domain of ROS1 lacks detectable enzymatic activity and binds DNA with very low affinity [15], whereas the isolated DNA glycosylase domain is inactive for 5-meC excision but retains partial AP lyase activity [17]. It has been shown that the addition of the C-terminal domain in trans to the DNA glycosylase domain restores base excision capacity, suggesting that the C-terminal region is essential for the 5-meC DNA glycosylase activity, perhaps by stabilizing the DNA glycosylase domain and/or stimulating its enzymatic activity [17].

ROS1 removes 5-meC at several hundred loci across the genome in vegetative tissues, apparently to counteract excessive methylation [10,18]. It is still poorly understood how ROS1 is directed to specific genomic regions, although the process likely involves the presence of specific chromatin modifications at target loci and/or the activity of recruiting factors. In chromatin, DNA is closely associated to core histones proteins (H2A, H2B, H3 and H4) whose N-terminal tails undergo different post-translational modifications [19]. Such modifications act as signal marks that can be read by different effector/reader proteins and chromatin-remodeling enzymes, thus influencing particular cellular processes [20]. It has been reported that loci targeted by ROS1 are enriched for H3K18Ac and H3K27me3 and depleted of H3K27me and H3K9me2 [21]. Furthermore, it has been identified a regulatory pathway initiated by the increased DNA methylation (IDM) complex in active DNA demethylation in *Arabidopsis* [22,23]. In this pathway, ROS1 is targeted to specific genomic loci through a complex that includes IDM1, IDM2, IDM3, the Methyl-CpG-Binding Domain-containing protein 7 (MBD7), and the Harbinger transposon-derived proteins HDP1 and HDP2. The IDM complex catalyzes histone acetylation at a subset of DNA demethylation target loci [23–25]. The histone acetylation marks created by the IDM complex are recognized by the SWR1 chromatin-remodeling complex, which mediates ROS1-mediated DNA demethylation at some loci [22]. However, whether ROS1 interacts with other histone variants or any histone modification is still unknown. Here, we report that ROS1 interacts with the N-terminal tail of H3 core histones through its C-terminal domain. We also show that ROS1 interaction with H3 is specifically abolished by phosphorylation of H3 Ser28 (H3S28) and requires conserved residues that are also important for DNA binding and catalytic activity.

2. Materials and methods

2.1. Protein expression and purification in E. coli

2.1.1. WT ROS1 and mutant versions (E1305Q, C1286A/R1287A and Y1300A/F1301A)

Site-directed mutagenesis was performed using the Quick-Change II XL kit (Stratagene). The mutations were introduced into the expression vector pET28a (Novagen) containing the full-length wild-type (WT) *ROS1* cDNA using specific oligonucleotides (Table 1, Supplementary Data). Mutational changes were confirmed by DNA sequencing and the constructs were used to transform *Escherichia coli* BL21 (DE3) *dcm*$^-$ Codon Plus cells (Stratagene). WT and mutant versions were expressed and purified as N-terminal His-tagged proteins, as follows: a fresh single transformed colony was inoculated into 10 ml of LB (Luria-Bertani) medium containing kanamycin (30 μg/ml) and chloramphenicol (34 μg/ml), and the culture was incubated at 37 °C overnight with shaking. A 2.5 ml aliquot of the overnight culture was inoculated into 250 ml of LB medium containing kanamycin (30 μg/ml) and chloramphenicol (34 μg/ml) and incubated at 37 °C, 250 rpm, until the A_{600} was 0.1. The culture was then placed at 23 °C, and incubation continued at 250 rpm for 90 min before adding 5 mM betaine, 5 mM Na-glutamate and 0.5 M NaCl. When the A_{600} reached 0.7, expression was induced by adding isopropyl-1-thio-β-D-galactopyranoside (IPTG) to 1 mM and incubating for 2 h. After induction, cells were collected by centrifugation at 13,000 *g* for 30 min, and the pellet was kept frozen at −80 °C. The stored pellet was thawed and resuspended in 3.5 ml of Sonication Buffer (SB: 20 mM Tris–HCl pH 8.0, 500 mM NaCl, 20% glycerol, 15 mM β-mercaptoethanol, 1% Tween-20) supplemented with 5 mM imidazole. Cells were disrupted by sonication and the lysate was clarified by centrifugation. The supernatant was loaded onto a Ni^{2+}-sepharose column (GE Healthcare) pre-equilibrated with SB buffer supplemented with 5 mM imidazole. The column was washed with 10 ml of SB supplemented with 5 mM imidazole, followed by 10 ml of SB supplemented with 100 mM imidazole. Proteins were eluted with a 30 ml gradient of imidazole (100 mM to 1 M) in SB and collected in 0.5 ml fractions. An aliquot of each fraction was analyzed by SDS–PAGE and those containing a single band of the overexpressed protein were pooled and dialyzed overnight against Dialysis Buffer (DB: 50 mM Tris–HCl pH 8.0, 1 mM DTT, 50% glycerol) containing 200 mM NaCl. The protein preparation was divided into aliquots, and stored at −80 °C. All steps were carried out at 4 °C or on ice. Protein concentration was determined by the Bradford assay [26]. Denatured proteins were analyzed by SDS–PAGE (10%) using broad-range molecular weight standards (Bio-Rad).

2.1.2. ROS1 truncated versions (NΔ294, NΔ88CΔ1075, NΔ1080 and NΔ519CΔ313)

ROS1 deletion constructs [15] were expressed and purified as N-terminal His-tagged proteins, as described above, except that expression was induced by adding IPTG to 1 mM and performed at 15 °C overnight. Pooled pure fractions were dialyzed against DB containing 200 mM NaCl except for NΔ88CΔ1075 and NΔ519CΔ313 where 500 mM NaCl was added.

2.2. Binding of ROS1 to a histone peptide array

The MODified™ Histone Peptide Array (Active Motif) was used to study the interaction of ROS1 with different histones tails and their post-translational modifications. The array consists of a glass slide (26 ×76 mm) with a cellulose membrane containing 19-mer peptides from eight different regions of the N-terminal tails of H3, H4, H2A and H2B histones (H3 1–19, 7–26, 16–35 and 26–45, H4 1–19 and 11–30, H2A 1–19 and H2B 1–19). The peptides are covalently bound to cellulose via their C-terminus. The array contains 59 different post-translational modifications for histone acetylation, methylation, phosphorylation and/or citrullination. Each peptide may contain up to four modifications. In total, 384 different modification combinations are spotted in duplicate onto the array. Five control spots are included (locations P20-P24): biotin peptide, c-Myc tag, no histone peptide and two background spots containing a mixture of modifications that are present on the array. The array was pre-incubated in blocking solution [5% non-fat dried milk in TBS-T buffer (10 mM Tris–HCl pH 7.4, 150 mM NaCl and 0.05% Tween-20)] at 4 °C overnight, washed three times with TBS-T buffer, and incubated with purified His_6-tagged ROS1 (10 nM) for 1 h at 25 °C in binding buffer (50 mM Tris-HCl pH 8, 1 mM DTT, 100 μg/ml BSA, 1 mM EDTA, 100 mM NaCl and 0.1% glycerol) with gentle agitation. After washing three times with TBS-T buffer, the array was incubated with His-tag monoclonal antibody (Novagen) at a 1:4000 dilution in blocking solution for 1 h at room temperature. Then, it was washed three times with TBS-T buffer and incubated with horseradish peroxidase-conjugated anti-mouse IgG (GE healthcare) at a 1:5000 dilution in blocking solution for 1 h at room temperature. Finally, the array was incubated with ECL Western blotting Detection Reagent (GE Healthcare) and images were captured using the LAS-3000 analyzer (Fujifilm).

2.3. Pull-down assays

Histone peptides used for pull-down assays were obtained from

Biomatik, Millipore or Anaspec. All peptides were biotinylated at the C-terminus, purified by HPLC before use (purity > 90–95%) and dissolved in sterile distilled water before dilution in the desired buffer. Their sequence, modifications, and molecular weight are shown in Table 2 (Supplementary Data). For peptide loading, 30 μl of Dynabeads M-280 Streptavidin (Invitrogen) were washed once with PBS-T (Phosphate Buffer Saline pH 7.4 and 0.1% Tween-20) and then incubated with the appropriated biotinylated histone peptide (400 pmol) in PBS at 4 °C overnight, according to the manufacturer's instructions. After incubation, beads were washed three times with PBS-T and twice with Binding Buffer (50 mM Tris-HCl pH 8, 1 mM DTT, 1 mM EDTA, 100 mM NaCl, 0.1% Tween-20, 0.01% protease inhibitor cocktail p8849 (Sigma), 50 nM tautomycin and 50 nM okadaic acid). Once loaded with the appropriate peptide, the beads were incubated with 50 pmol of purified His_6-tagged ROS1 (WT, truncated or mutant versions) in 400 μl of Binding Buffer containing 0.09 μg/μl BSA for 2 h at 4 °C. After incubation, beads were washed three times with Binding Buffer. Proteins bound to beads were separated by SDS-PAGE (7% acrylamide/bisacrylamide, 37.5:1) and transferred to nitrocellulose membranes at 30 V at 4 °C overnight. After blotting, the membranes were blocked with Blocking Solution [1% BSA (bovine serum albumin) in TBS (10 mM Tris-HCl pH 7.4 and 150 mM NaCl)] for 1 h at room temperature and washed three times with TBS supplemented with 0.5% Tween-20 (TBS-T). Then, membranes were incubated with a His-tag monoclonal antibody (Novagen) at a 1:1000 dilution in Blocking Solution for 1 h at room temperature, washed three times with TBS-T, and incubated with horseradish peroxidase-conjugated anti-mouse (Cell Signaling) at a 1:3000 dilution in Blocking Solution for 1 h at room temperature. After three washes with TBS-T, membranes were incubated with ECL Prime Western Blotting Detection Reagent (GE Healthcare). Images were captured using the LAS-3000 analyzer (Fujifilm).

2.4. DNA substrates

Oligonucleotides used as DNA substrates in enzymatic assays (Table 3, Supplementary Data) were synthesized by Operon or Integrated DNA Technologies (IDT) and purified by polyacrylamide gel electrophoresis (PAGE) or dual HPLC before use. Double-stranded DNA substrates were prepared by mixing a 5 μM solution of a 5′-fluorescein-labeled oligonucleotide (upper-strand) with a 10 μM solution of an unlabeled oligomer (lower-strand), heating to 95 °C for 5 min and slowly cooling to room temperature. DNA containing a natural AP site opposite guanine was prepared by incubating a DNA duplex containing a U:G mispair (200 nM) with 2.5 U of *Escherichia coli* Uracil DNA glycosylase (NEB) at 30 °C for 5 min.

2.5. Enzyme activity assays

For DNA glycosylase/lyase activity assays, fluorescein-labeled duplex oligonucleotides containing 5-meC:G (20 nM) were incubated at 30 °C for 24 h in a reaction mixture containing 50 mM Tris–HCl pH 8.0, 1 mM EDTA, 1 mM DTT, 0.1 mg/ml bovine serum albumin (BSA) and 20 nM of WT ROS1 or mutant variants in a total volume of 50 μl. When measuring DNA glycosylase activity, after incubation NaOH (100 nM) was added and samples were immediately transferred to 90 °C for 10 min. For AP lyase activity assays, a fluorescein-labeled duplex oligonucleotide containing an AP site opposite G (20 nM) was used. After 2 h reaction at 30 °C, the mixture was incubated with 300 mM $NaBH_4$ at 0 °C for 30 min and neutralized with 100 mM Tris–HCl pH 7.4. All reactions were stopped by adding 20 mM EDTA, 0.6% sodium dodecyl sulphate (SDS) and 0.5 mg/ml proteinase K, and the mixtures were incubated at 37 °C for 30 min. DNA was extracted with 1 vol of phenol:chloroform:isoamyl alcohol (25:24:1, Sigma) and precipitated with 3 volumes of ethanol absolute at − 20 °C in the presence of 0.3 mM NaCl and 16 μg/ml glycogen. Samples were resuspended in 10 μl of 90% formamide and heated at 95 °C for 5 min. Reaction products were separated in a 12% denaturing polyacrylamide gel containing 7 M urea. Fluorescein-labeled DNA was visualized in a FLA-5100 imager and analyzed using Multigauge software (Fujifilm).

2.6. Electrophoretic mobility shift assay (EMSA)

WT ROS1 or ROS1 variants were incubated with 100 nM of a fluorescein-labeled duplex oligonucleotide. DNA binding reactions were carried out at 25 °C for 5 min in 10 mM Tris–HCl, pH 8.0, 1 mM DTT, 10 μg/ml BSA, 1 mM EDTA, in a final volume of 10 μl. Complexes were electrophoresed through 0.2% agarose gels in 1X TAE (40 mM Tris–HCl, pH 8.0, 20 mM acetic acid, 1 mM EDTA). Electrophoresis was carried out in 1X TAE for 40 min at 80 V at room temperature. Fluorescein-labeled DNA was visualized in a FLA-5100 imager and Multigauge software (Fujifilm).

3. Results

3.1. ROS1 binds all four core histones but specific modifications abrogate the interaction

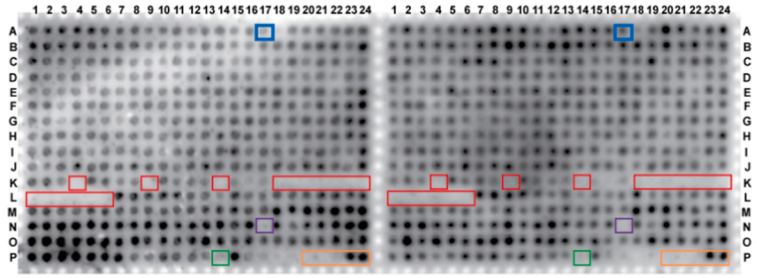
As a starting approach to determine whether ROS1 interacts with histones and specific histone marks, we performed a massive scrutiny using a histone peptide array (MODified Histone Peptide Array). The array comprises 384 spots containing 19-mer peptides of different regions of the N-terminal tails of core histones (H3 1–19, 7–26, 16–35 and 26–45, H4 1–19 and 11–30, H2A 1–19 and H2B 1–19). It features 59 post-translational modifications in different combinations including acetylation, methylation, phosphorylation, and citrullination. For quality control, all peptides included in the array are spotted in duplicate. The array was incubated with purified His-tagged ROS1, and bound protein was detected with an anti-His antibody. The results (Fig. 1A) showed that ROS1 remained bound to most peptides, as well as to two control spots (Fig. 1A, P23 and P24) containing a mixture of modifications that are present in the array. The apparent absence of ROS1 binding at spot A18 (H3T11P) was not consistently observed in independent experiments, and spots P17-P19 contain H2B modifications not detected in plants [27]. Importantly, no signal was detectable in three negative control spots (Fig. 1A, P20-P22) containing a biotin control peptide, a c-myc tag or a non-histone peptide, thus ruling out non-specific interactions.

Although ROS1 bound most histone peptides, it was consistently observed, in independent experiments and on both duplicates in the same array, that no signal was detectable at specific spots (Fig. 1A). The most relevant observation was that ROS1 binding to H3 was disrupted when Ser28 was phosphorylated, independently of the presence of other modification marks in the same peptide (Fig. 1A and B). In contrast, phosphorylation at H3S10 did not affect ROS1 binding (Fig. 1A and B, A17). The inhibition of ROS1 binding by specific modifications was also observed for other histones. For example, binding to H2B was inhibited when Ser14 was phosphorylated in combination with acetylation at Lys15 (Fig. 1A and B, P14). Additionally, binding to H4 was not detectable when acetylation at both Lys12 and Lys16, and dimethylation at Lys20 were combined in the same peptide (Fig. 1A and B, N17). Altogether, these results suggest that ROS1 interacts with the N-terminal tails of all four core histones, and that such interaction may be abolished by specific modifications.

3.2. Phosphorylation of H3S28, but not H3S10, specifically abrogates ROS1 interaction with histone H3

The results described above suggested that, although H3S28 and H3S10 are located at regions sharing the consensus ARKS sequence (Fig. 2A), only phosphorylation at Ser28 disrupts ROS1 binding to H3. To specifically analyze differences in ROS1 binding to H3S10 and H3S28, we performed pull-down assays with His-tagged ROS1 and

A



B

Peptide position	Peptide sequence	Name	Mod1	Mod2	Mod3	Mod4	N-terminus
A1	A R T K Q T A R K S T G G K A P R K Q	H3 1-19	unmod				free
A17	A R T K Q T A R K pS T G G K A P R K Q	H3 1-19	S10P				free
J20	P R K Q L A T K A A R K S A P A T G G	H3 16-35	unmod				acetylated
K4	P R K Q L A T K A A R K pS A P A T G G	H3 16-35	S28P				acetylated
K9	P R K Q L A T K A A Rme2s K pS A P A T G G	H3 16-35	R26me2s	S28P			acetylated
K14	P R K Q L A T K A A Rme2a K pS A P A T G G	H3 16-35	R26me2a	S28P			acetylated
K18	P R K Q L A T K A A Cit K pS A P A T G G	H3 16-35	R26Citr	S28P			acetylated
K19	P R K Q L A T K A A R Kme1 pS A P A T G G	H3 16-35	K27me1	S28P			acetylated
K20	P R K Q L A T K A A R Kme2 pS A P A T G G	H3 16-35	K27me2	S28P			acetylated
K21	P R K Q L A T K A A R Kme3 pS A P A T G G	H3 16-35	K27me3	S28P			acetylated
K22	P R K Q L A T K A A R Kac pS A P A T G G	H3 16-35	K27ac	S28P			acetylated
K23	P R K Q L A T K A A Rme2s Kme1 pS A P A T G G	H3 16-35	R26me2s	K27me1	S28P		acetylated
K24	P R K Q L A T K A A Rme2s Kme2 pS A P A T G G	H3 16-35	R26me2s	K27me2	S28P		acetylated
L1	P R K Q L A T K A A Rme2s Kme3 pS A P A T G G	H3 16-35	R26me2s	K27me3	S28P		acetylated
L2	P R K Q L A T K A A Rme2s Kac pS A P A T G G	H3 16-35	R26me2s	K27ac	S28P		acetylated
L3	P R K Q L A T K A A Rme2a Kme1 pS A P A T G G	H3 16-35	R26me2a	K27me1	S28P		acetylated
L4	P R K Q L A T K A A Rme2a Kme2 pS A P A T G G	H3 16-35	R26me2a	K27me2	S28P		acetylated
L5	P R K Q L A T K A A Rme2a Kme3 pS A P A T G G	H3 16-35	R26me2a	K27me3	S28P		acetylated
L6	P R K Q L A T K A A Rme2a Kac pS A P A T G G	H3 16-35	R26me2a	K27ac	S28P		acetylated
N17	G Kac G G A Kac R H R Kme2 V L R D N I Q G I T	H4 11-30	K12ac	K16ac	K20me2		acetylated
P14	P D P A K S A P A P K K G pS Kac K A V T	H2B 1-19	S14P	K15ac			free
P20	Bio A A N W S H P Q F E K A A	Biotin, control peptide					biotinylated
P21	E Q K L I S E E D L A	c-myc tag					free
P22	HAc	neg. contol					acetylated
P23	K Kme1 Kme2 Kme3 Kac R Rme2s R Rme2a R Cit K Kme1 Kac Kme3 R K	background 01					acetylated
P24	R Rme2s K Kme1 Kac R Rme2a Kme2 K Kme3 R Kme1 Rme2s K Kac R K	background 02					acetylated

Fig. 1. Massive scrutiny of ROS1 interaction with different histone peptides. (A) Binding of ROS1 to a histone peptide array (MODified Histone Peptide Array, Active Motif) including 384 peptides of different regions of the N-terminal tails of core histones containing different post-translational modification combinations. The array was incubated with purified His_6-tagged ROS1 (10 nM) for 1 h at 25 °C and ROS1 binding was detected by an anti-His_6-tag antibody. Colored boxes indicate selected spots described in panel B. (B) Description of selected spots of the histone array. The position, sequence, name, and type of modifications are shown. Coloring scheme is as in panel A. See http://www.active motif.com/catalog/668/modified histone-peptide-array for a detailed annotation of all spots.

biotinylated H3 peptides containing H3S10 (amino acids 1–21) or H3S28 (amino acids 21–43) in either unmodified or phosphorylated form (Fig. 2A). Consistent with the array results, ROS1 interacted with both regions of unmodified histone H3 (Fig. 2B, lanes 3 and 10), although the percentage of ROS1 bound to the region containing amino acids 1–21 was higher compared to the region containing amino acids 21–43 (Fig. 2D). However, ROS1 interaction was virtually abolished by H3S28 phosphorylation, but remained unaffected when H3S10 was phosphorylated (Fig. 2B lanes 5 and 12 and Fig. 2D). These results suggest that ROS1 interaction with H3 is sequence-specific and most likely involves residues preceding and/or following each ARKS consensus sequence.

3.3. The C-terminal domain of ROS1 is responsible for H3 binding

ROS1 is a large and complex protein composed of three major domains: a lysine-rich N-terminal domain, a discontinuous DNA glycosylase catalytic domain, and a C-terminal domain that is highly conserved among DML family proteins [5]. In order to identify the domain(s) involved in the interaction of ROS1 with histone H3, different His-tagged truncated versions of ROS1 (Fig. 2C) were purified and compared to the WT protein in their ability to bind H3 peptides with either unmodified or phosphorylated versions of Ser10 or Ser28 (Fig. 2B). A ROS1 variant lacking the N-terminal domain (NΔ294) behaved like the WT protein, since it bound H3 peptides unless Ser28 was phosphorylated (Fig. 2B, lane 3,5,10 and 12 and Fig. 2D). However, the percentage of bound protein was lower compared to the WT version, particularly when Ser10 was phosphorylated (Fig. 2D). Conversely, the interaction of a ROS1 version containing only the N-terminal domain (NΔ88CΔ1075) was virtually undetectable with any H3 peptide (Fig. 2B, lane 3, 5, 10 and 12 and Fig. 2D). These results indicate that the N-terminal domain of ROS1 is not essential for the interaction with histone H3. Furthermore, a truncated ROS1 version containing only the discontinuous DNA glycosylase domain (NΔ519CΔ313) also showed a significantly reduced capacity to bind the unmodified H3 peptide containing amino acids 1–21 and the interaction with the H3S10P peptide was barely detectable (Fig. 2B, lane 3 and 5 and Fig. 2D). Also, in contrast to the WT protein, NΔ519CΔ313 displayed a very weak interaction with the peptide containing phosphorylated Ser28, and the binding to the unmodified version was almost completely abolished

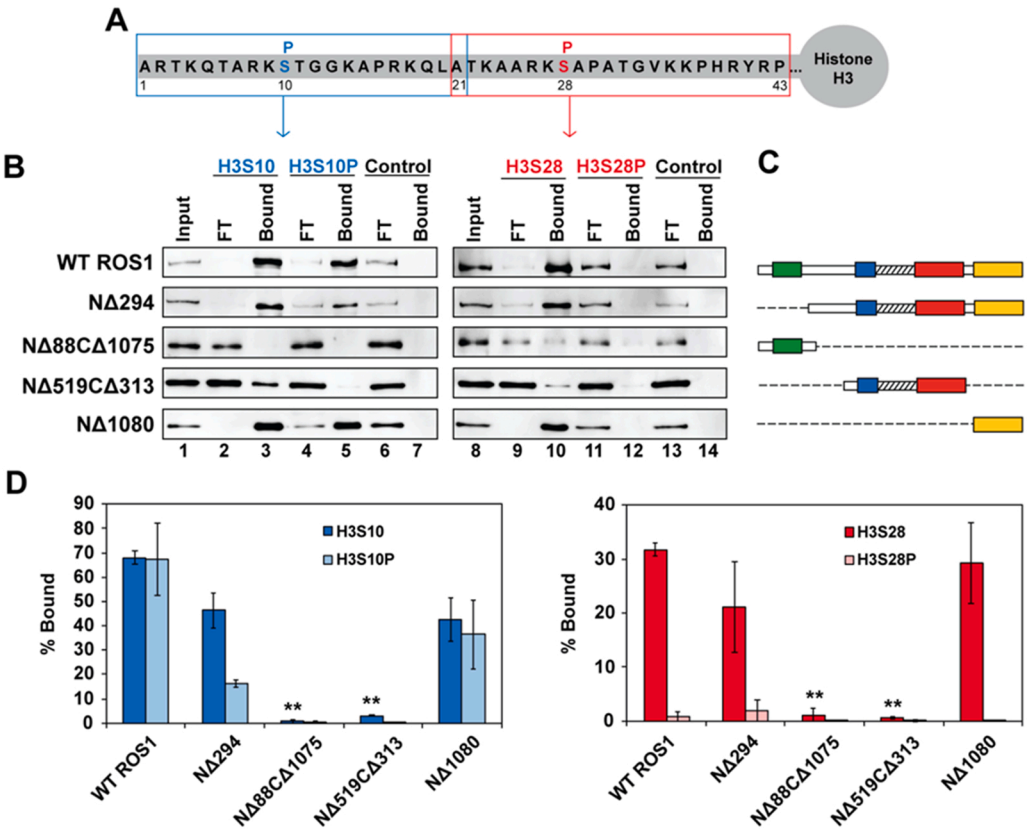


Fig. 2. Analysis of ROS1 interaction with H3 peptides containing either unmodified or phosphorylated versions of Ser10 or Ser28. (A) Schematic diagram of a region of the N-terminal tail of histone H3. Boxes indicate the sequence of histone H3 peptides (either unmodified or phosphorylated at Ser10, in blue, or either unmodified or phosphorylated at Ser28, in red) used in pull-down assays. (B) Pull-down assays performed with different ROS1 variants (schematized in panel C) and biotinylated H3 histone peptides (1–21 or 21–43 amino acids) phosphorylated or not at Ser10 or Ser28, respectively. A representative gel is shown for each protein. Input: 3% of each protein variant used for binding, FT: 3% of Flow-Through, bound: 25% bound proteins, beads with no peptide were used as the control for non-specific binding (C) Schematic diagrams showing the domains deleted in the different ROS1 variants. The N-terminal domain is shown in green, the discontinuous DNA glycosylase domain is distributed over two segments (blue and red) separated by a non-structured linker region (striped), and the C-terminal domain is colored in yellow. Discontinuous lines indicate deleted regions. (D) Graphs show the percentage of bound protein to each histone peptide normalized to the corresponding input. Values are the mean (± SEM) from two independent experiments. Asterisks indicate statistically significant differences compared to WT ROS1 (**= *P* < *0.01;* Student's unpaired t-test).

(Fig. 2B, lane 10 and 12 and Fig. 2D). Finally, a ROS1 variant containing only the C-terminal domain (NΔ1080) displayed a binding capacity similar to that of the full-length protein. Importantly, NΔ1080 bound the two unmodified H3 peptides and the one phosphorylated at Ser10 but was unable to bind the peptide containing phosphorylated Ser28 (Fig. 2B, lane 3, 5, 10 and 12 and Fig. 2D). Altogether, these results suggest that the main region responsible for ROS1 interaction with histone H3 is located at the C-terminal domain of the enzyme.

3.4. The methylation status of H3K27 does not affect ROS1 binding

Adjacent to H3S28 there is a lysine residue (H3K27) that can be mono-, di- or tri- methylated [28]. Since methylation at H3K27 is an important histone mark associated to transcriptional repression [28], we asked whether the methylation status of H3K27 affects ROS1 binding. To this end, the C-terminal domain of ROS1 (NΔ1080) was incubated with each of four versions of an H3 peptide (21–43 amino acids) containing unmodified, mono-, di- or tri-methylated H3K27 (Supplementary Figure 1A). We found that the amount of NΔ1080 bound to all four peptides was very similar (Supplementary Figure 1B and C). These findings agree with those observed in the peptide array experiments, where spots containing unmodified, mono-, di- or tri-methylated H3K27, retained ROS1 with similar affinity (Fig. 1A, spots A1, J24, K1 and K2, respectively). Altogether, these results indicate that the methylation status of H3K27 does not affect ROS1 binding.

3.5. Identification of ROS1 C-terminal residues important for interaction with histone H3

In order to identify candidate residues involved in the interaction of ROS1 with H3, we performed a multiple sequence alignment of the C-terminal domains of several DML family proteins (Supplementary Figure 2). The alignment confirmed that the C-terminus is highly conserved among members of this family of atypical DNA glycosylases. Two major conserved regions are observed, separated by a section containing 4 Cys residues that may be a permuted version of a single unit of a ZF-CXXC domain [29]. However, the two highly conserved regions separated by such domain are not related to any known protein. Sequence conservation is particularly high in the second region (residues 1279–1317 in *Arabidopsis* ROS1), suggesting that this area contains amino acids important for ROS1 function. Since there is no structural information available for the C-terminal domain of ROS1, we opted for residues located in such conserved region (residues 1279–1317) and representing different chemical classes: polar uncharged (C1286), positively charged (R1287), aromatic (Y1300 and F1301), and negatively charged (E1305). Therefore, we generated three different His-tagged ROS1 C-terminus mutant versions, one containing a single mutation (E1305Q), and two double-mutants (C1286A/R1287A and Y1300A/F1301A) (Fig. 3A and Supplementary Figure 2). All three mutant versions showed a drastically reduced capacity to interact with the unmodified version of H3S28 (21–43 region of H3) (Fig. 3B, lane 10 and Fig. 3C). As expected, none bound the H3S28P version (Fig. 3B, lane 12 and Fig. 3C). However, they retained at different degrees the capacity to interact with both the unmodified and phosphorylated versions of Ser10 (1–21 region of H3) (Fig. 3B, lanes 3 and 5 and Fig. 3C). These

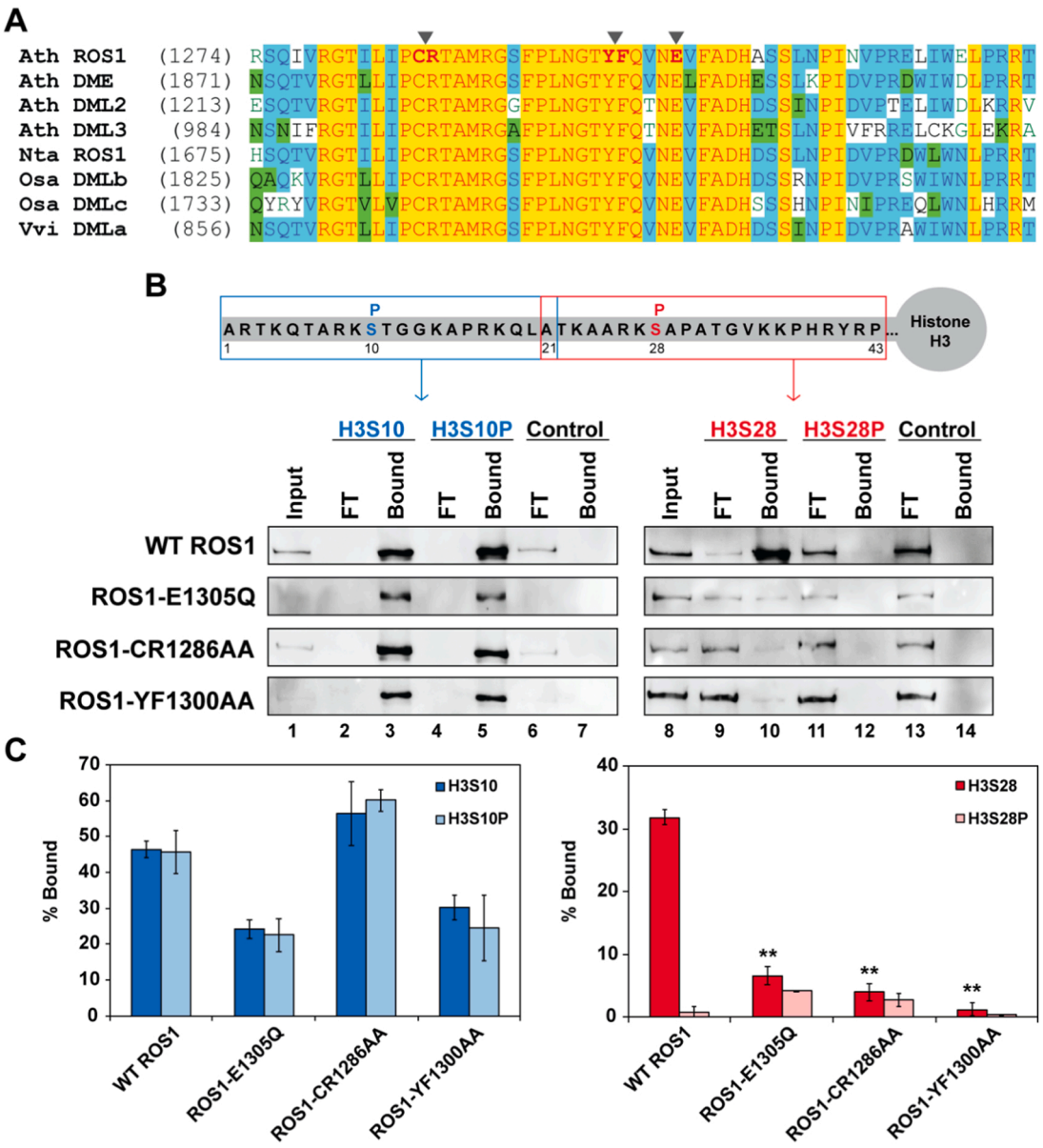


Fig. 3. Analysis of the interaction of ROS1 C-terminal mutant proteins with H3 peptides containing either unmodified or phosphorylated versions of Ser10 and Ser28. (A) Partial multiple sequence alignment of the C-terminal domain of several DML family proteins. Inverted grey triangles indicate mutated amino acids. Names of organisms are abbreviated as follows: Ath, *Arabidopsis thaliana*; Nta, *Nicotiana tabacum*; Osa, *Oryza sativa*; Vvi, *Vitis vinifera*. Genbank accession numbers are as follows: Ath ROS1: AAP37178; Ath DME: ABC61677; Ath DML2: NP_187612.5; Ath DML3: OAO99112; Nta ROS1: BAF52855; Osa DMLb: BAF04322; Osa DMLc: EEE63898; Vvi DMLa: CAO46558. (B) Pull-down assays performed with WT ROS1 or C-terminal mutant variants and biotinylated H3 histone peptides (1–21 or 21–43 amino acids) phosphorylated or not at Ser10 or Ser28, respectively. A representative gel is shown for each protein. Input: 3% of each protein variant used for binding, FT: 3% of Flow-Through, bound: 25% bound proteins, beads with no peptide were used as the control for non-specific binding (C) Graphs show the percentage of bound protein to each histone peptide, normalized to the corresponding input. Values are the mean (± SEM) from two independent experiments. Asterisks indicate statistically significant differences compared to WT ROS1 (**= $P < 0.01$; Student's unpaired t-test).

results suggest that the mutated residues (C1286 and/or R1287, Y1300 and/or F1301 and E1305) are specifically involved in the phosphorylation-sensitive recognition of the region surrounding H3S28 but are only marginally implicated in the phosphorylation-independent interaction with the region surrounding H3S10.

3.6. *The C-terminal mutant proteins E1305Q, C1286A/R1287A and Y1300A/F1301A are deficient in 5-meC excision*

It has been previously reported that the C-terminal domain of DML family proteins is necessary for 5-meC excision [14,17]. Therefore, we examined the catalytic activity of the C-terminal mutant proteins E1305Q, C1286A/R1287A and Y1300A/F1301A in order to study the role of these residues in ROS1 enzymatic function. We found that the amount of incision products generated by C1286A/R1287A was significantly reduced compared to that of the WT protein, whereas mutants E1305Q and Y1300A/F1301A displayed almost undetectable activity (Fig. 4A and Supplementary Figure 3A). Therefore, the results suggest that the mutated residues are required for efficient ROS1 catalytic function.

Since ROS1 is a bifunctional DNA glycosylase/lyase that catalyzes both the release of 5-meC and the cleavage of DNA at the resulting abasic site [7], we further analyzed the reduced capacity of the three mutant proteins to determine if it was due to a deficiency in DNA glycosylase

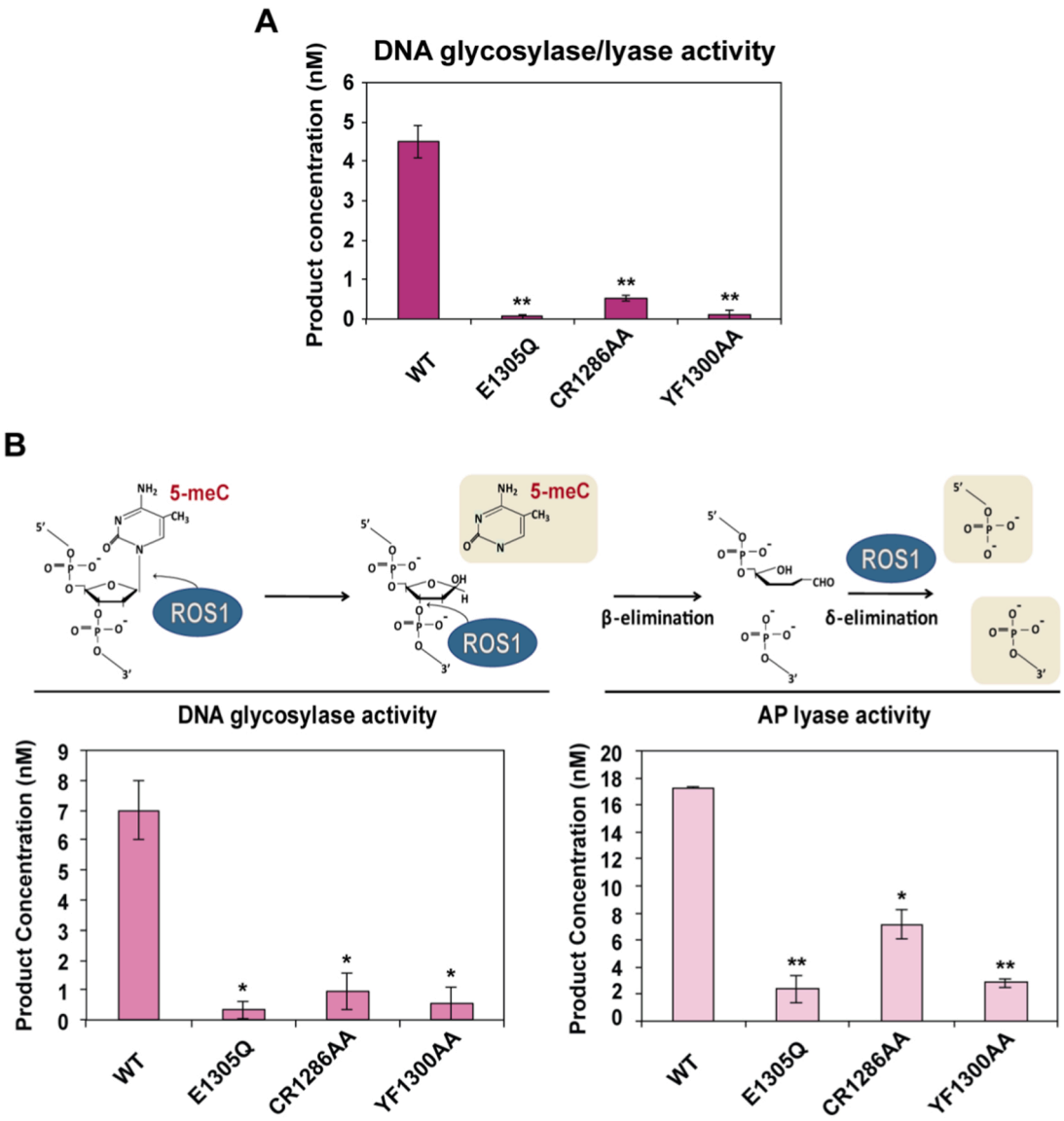


Fig. 4. Analysis of the catalytic activity of ROS1 C-terminal mutant proteins. (A) Combined DNA glycosylase/AP lyase activity of WT ROS1 and C-terminal mutant variants on a DNA substrate containing a 5-meC:G pair. Purified proteins (20 nM) were incubated at 30 °C for 24 h with a double-stranded oligonucleotide substrate (20 nM) containing 5-meC opposite G. Products were separated in a 12% urea denaturing polyacrylamide gel and the amount of incised oligonucleotide was quantified by fluorescent scanning. Values are the mean (± SEM) from two independent experiments. Asterisks indicate statistically significant differences compared to WT ROS1 (**= $P < 0.01$; Student's unpaired t-test). (B) Top, schematic diagram of ROS1 DNA glycosylase/AP lyase activity on 5-meC. ROS1 excises 5-meC as a free base and then cleaves the phosphodiester backbone at the 5-meC removal site by successive β,δ-elimination. Bottom left, DNA glycosylase assay. The generation of incision products was measured by incubating purified WT ROS1 or mutant variants (20 nM) at 30 °C for 24 h with a double-stranded oligonucleotide substrate (20 nM) containing a single 5-meC:G pair. After incubation, NaOH (100 nM) was added and samples were immediately transferred to 90 °C for 10 min. Bottom right, AP lyase assay. A double-stranded oligonucleotide substrate containing an AP site opposite G (20 nM) was incubated at 30 °C for 2 h in the presence of purified WT ROS1 or mutant variants (20 nM). Samples were treated with $NaBH_4$ (300 mM) at 0 °C for 30 min to stabilize non-processed AP sites and neutralized with 100 mM Tris–HCl, pH 7.4. Products were separated in a 12% urea denaturing polyacrylamide gel and the amount of incised oligonucleotide was quantified by fluorescent scanning. Values are the mean (± SEM) from two independent experiments (* = $P < 0.05$; **= $P < 0.01$; Student's unpaired t-test).

activity, AP lyase activity or both (Fig. 4B and Supplementary Figure 3B). To detect DNA glycosylase activity, the reaction products generated by the different ROS1 variants were analyzed after an additional alkaline treatment with NaOH, which cleaves all abasic sites generated by the enzyme and reflects 5-meC excision activity. All three mutant variants showed a significantly decreased DNA glycosylase activity (Fig. 4B, left and Supplementary Figure 3B, left). To determine whether the three variants retained AP lyase activity, the proteins were incubated with DNA substrate containing an AP site opposite G. The C1286A/R1287A mutant variant cleaved the abasic site less efficiently than WT ROS1, whereas E1305Q and Y1300A/F1301A mutant proteins lack significant AP lyase activity (Fig. 4B, right and Supplementary Figure 3B, right). These results indicate that E1305, C1286 (and/or R1287) and Y1300 (and/or F1301) are required for efficient 5-meC excision and the subsequent sugar-phosphate cleavage step.

3.7. Residue E1305 is required for DNA binding

Since the catalytic activity of C-terminal mutant proteins was null or drastically reduced, their DNA binding capacity was examined to determine whether the mutated residues are also important for binding methylated DNA. Two different concentrations of WT ROS1, E1305Q, C1286A/R1287A or Y1300A/F1301A proteins were incubated with a labeled DNA substrate containing a 5-meC:G pair in a gel-shift assay

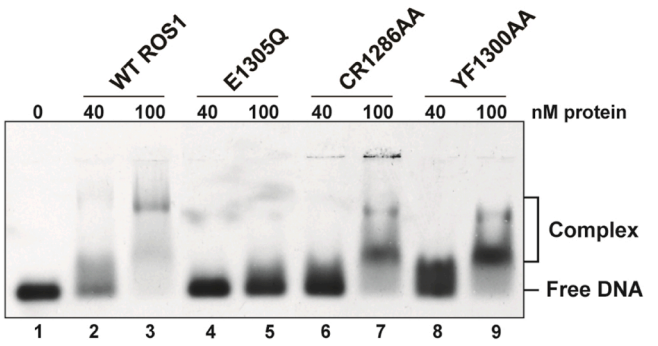


Fig. 5. DNA binding capacity of ROS1 C-terminal mutant proteins. WT ROS1 or C-terminal mutant variants (40 and 100 nM) were incubated at 25 °C for 5 min with a double-stranded oligonucleotide substrate (100 nM) containing a single 5-meC:G pair. After non-denaturing gel electrophoresis, protein–DNA complexes were identified by their retarded mobility compared with that of free DNA, as indicated.

(Fig. 5). In the case of WT ROS1, a major band with retarded mobility was observed at the highest protein concentration and a minor diffuse band was detectable at the lowest protein concentration (Fig. 5, lanes 2 and 3), suggesting the formation of protein–DNA complexes containing more than one ROS1 molecule, which agrees with previously reported data [30]. In contrast, no protein–DNA complexes were detected in binding reactions containing the mutant protein E1305Q (Fig. 5, lanes 4 and 5); only a barely detectable, diffuse band was observed at the highest protein concentration. On the other hand, both C1286A/R1287A and Y1300A/F1301A proteins bound DNA less efficiently than WT ROS1, since an intermediate and a major band with retarded mobility was observed at the highest concentration, whereas a minor diffuse band was detectable at the lowest protein concentration (Fig. 5, lanes 6–9). The intermediate band suggests the formation of protein–DNA complexes containing a smaller number of ROS1 molecules than the major band. Altogether, these results suggest that E1305 is essential for DNA binding, whereas C1286 (and/or R1287) and Y1300 (and/or F1301) contribute to stabilizing the DNA-protein complex.

4. Discussion

ROS1 is an atypical DNA glycosylase containing a C-terminal domain of unknown function that is highly conserved among DML family proteins [7]. In this study we show that ROS1 C-terminal domain binds the N-terminal tail of H3 histone and its post-translational modifications, pointing towards a mechanism for targeting ROS1 demethylation activity to specific loci. Several works have reported results supporting the existence of a targeting mechanism that regulates DNA demethylation in plants. For instance, the analysis of DNA methylation in the genome of *Arabidopsis ros1 dml2 dml3* triple mutant compared to WT plants showed no global changes but significant differences were observed in a small set of genes [10,18]. One of the existing models for ROS1 targeting to specific loci involves small RNAs bound to ROS3, a RNA-binding protein required for demethylation at a number of loci, some of which overlap with ROS1 targets [31]. Alternatively, since DNA demethylation occurs in a chromatin environment, nucleosome positioning, histone modifications and histone variants may contribute to removal of 5-meC. In this regard, it has been reported that IDM1 (Increased DNA methylation 1)-dependent histone acetylation regulates active DNA demethylation of a subset of loci targeted by ROS1 and its homologs in *Arabidopsis*. IDM1 binds methylated DNA at chromatin sites with unmethylated H3K4 and acetylates H3K18 and H3K23 to create a chromatin environment that allows 5-meC DNA glycosylases to function [23]. Interestingly, it has been recently shown that ROS1 is recruited to some of its genomic targets by interaction with the histone variant H2A.Z deposited by the SWR1 chromatin-remodeling complex, which in turn is recruited to chromatin through recognition of histone acetylation marks created by the IDM1 complex [22]. Our results further support the idea that ROS1 is able to interact with histones in the nucleosome core.

4.1. Phospho-sensitive interaction of ROS1 with histone H3

We have found that ROS1 interacts with the N-terminal tail of H3 through its C-terminal domain. Interestingly, phosphorylation of Ser28, but not Ser10, specifically inhibits ROS1 interaction with H3, independently of the presence of others histone modifications. Moreover, both Ser10 and Ser28 are part of the same ARKS sequence motif, therefore, the absence of ROS1 interaction with phosphorylated histone H3 at Ser28 is sequence-specific and most likely involves residues preceding and/or following the ARKS consensus motif. In agreement with this hypothesis, a study showed that the chromodomain of Polycomb (Pc) proteins, which specifically binds to H3K27me3, discriminates between the H3K27 and H3K9 methylation sites, both part of ARKS motifs, due to recognition of five additional residues preceding the ARKS consensus sequence [32].

Histone H3 phosphorylation has been associated with chromosome condensation/segregation during mitosis and meiosis, but also with chromatin relaxation and transcription activation of specific genes during interphase in both animals and plants [33–36]. The role of both H3S10 and H3S28 phosphorylation is better understood in combination with other histone marks within a particular chromatin context [33]. Since many methylated lysine residues of histone H3 are adjacent to residues suitable to be phosphorylated, a “phospho-methyl switch” has been suggested as a mechanism to control the association of proteins with chromatin in various biological processes: during mitosis and during gene transcription regulation [37–41]. In mammals, the amino-terminal chromodomain of heterochromatin protein 1 (HP1) interacts with tri-methylated H3K9 during interphase, mediating heterochromatin formation. Phosphorylation of the adjacent H3S10 during mitosis abrogates HP1 binding to methylated H3K9 [42], and dephosphorylation of H3S10 at the end of mitosis re-establishes the association of HP1 with chromatin, suggesting that this binary “phospho-methyl switching” permits dynamic control of the HP1–H3K9me interaction [37]. Moreover, the mammalian Polycomb Repressive Complex 2 (PRC2) di- and tri-methylates H3K27 and binds such marks in order to repress Polycomb group (PcG) target genes [28]. Phosphorylation of the adjacent H3S28 by MSK1/2 kinases in response to mitogen, stress or differentiation signals displaces PcG proteins, leading to gene activation [41].

Although DNA and H3K27 methylation are traditionally considered as mutually exclusive chromatin marks, their relationship appears to be complex and there are reports supporting that both epigenetic modifications can act either synergistically or antagonistically [43]. In fact, the genomic regions targeted by ROS1 are generally enriched for H3K27me3 and depleted of H3K27me and H3K9me2 [21]. We have found that ROS1 binding to H3 is, at least *in vitro*, independent of the methylation status of H3K27. Interestingly, *Arabidopsis* PWWP-DOMAIN INTERACTOR OF POLYCOMBS1 (PWO1) protein also binds H3 independently of H3K27 methylation, and such interaction is inhibited by H3S28 (but not H3S10) phosphorylation [44]. It is possible that H3S28 phosphorylation serves as a general mechanism to displace activating and repressive modifiers from chromatin in response to mitotic, differentiation and/or stress signals.

4.2. The C-terminal domain of DML family proteins: a putative new histone reader motif

In this work, several pieces of evidence indicate that ROS1 C-terminal domain is the main responsible for mediating ROS1 interaction with histones and their specific post-translational modifications, thus suggesting that the C-terminus of ROS1 harbors a histone reader motif. First, the C-terminal domain is sufficient for specific interaction with

histone H3 and for discrimination between the regions surrounding phosphorylated Ser10 and phosphorylated Ser28. In comparison, the N-terminal domain of ROS1 does not interact with histone H3 peptides, either unmodified or phosphorylated at Ser10 or Ser28, and so far its function has been mainly related with non-specific DNA binding [15,16]. The DNA glycosylase domain, on the other hand, may modulate the C-terminal interaction with the 1–21 region of unmodified H3. Second, analysis of several ROS1versions mutated at different highly conserved residues of the C-terminal domain in DML family proteins identified C1286 and/or R1287, Y1300 and/or F1301 and E1305 ROS1 residues as strong candidates for mediating phosphorylation-sensitive recognition of the region surrounding H3S28. Interestingly, a region containing 4 Cys residues at ROS1C-terminus may be a permuted version of a single unit of a ZF-CXXC domain [29], which is found in a variety of chromatin-associated proteins (e.g. KMD2, DNMT1, MBD1, TET1/3) and binds non-methylated CpG. This putative ZF-CXXC separates the two C-terminal regions highly conserved among the members of DML family proteins, but they are not related to any known protein. Here, we show that the highest conserved subregion contains amino acids (C1286, R1287, Y1300, F1301 and E1305) are important for ROS1 function in histone interaction and in DNA demethylation.

Our data suggest that the C-terminal domain of ROS1 may harbor more than one histone-reader motif or unidentified versions of already known motifs. This is supported by the observation that ROS1 is able to bind unmodified histones and, at least, phosphorylated and methylated marks, and the fact that each histone mark is usually recognized by defined motifs [45,46]. There are several histone readers that interact with unmodified histone tails (PHD fingers, ADD and WD40 modules), phosphorylated marks (14–3–3 proteins, tandem BRCT and BIR domains) or methylated marks (the Royal superfamily, ADD, ankyrin, BAH, DCD PHD, WD40, Zf-CW and Tudor domains) [45,46]. A recent study revealed that plant and human histone readers share domain types and recognition mechanisms [47]. Furthermore, several proteins that interact with both DNA and histones or its epigenetics marks contain more than one histone reader domain. For example, UHRF1contains a SRA (SET and RING-associated) domain, a PHD finger and a TTD histone binding domain, whereas DNMT3A and DNMT3B contain a cysteine-rich ADD motif and PWWP domains [48].

4.3. Catalytic function of the ROS1 C-terminal domain

In addition to a role in histone interaction, several conserved C-terminus residues are required for efficient ROS1 catalytic activity, supporting the observation that, unlike the N-terminal domain, the C-terminal domain of DML family proteins is necessary for 5-meC excision [14,17]. The isolated DNA glycosylase domain of ROS1 is not able to excise 5-meC but partially cleaves the sugar phosphate backbone [17] and the isolated C-terminal domain is not able to process 5-meC [15]. The 5-meC excision activity is restored after addition of the C-terminal domain to the DNA glycosylase domain [17]. Furthermore, a small deletion at DME C-terminus (51 amino acids) was enough to completely abolish the DNA glycosylase activity of an otherwise active truncated version lacking 677 amino acids of the N-terminus [14]. Our results suggest that ROS1 C-terminus might play a role in ROS1 DNA binding, since E1305 is essential to form a DNA-protein complex and at least two of the other mutated residues, C1286 (or R1287) and Y1300 (or F1301), contribute to its stabilization. A previous study showed that the truncated ROS1 version containing only the C-terminal domain (NΔ1080) was able by itself to bind methylated DNA, but with very low affinity, and a truncated version containing only the discontinuous catalytic domain of ROS1 (therefore lacking both the C-terminal domain and the major responsible domain for ROS1 DNA binding, the N-terminal domain), did not exhibit any detectable DNA binding capacity [15], also suggesting that the C-terminal domain of ROS1 is somehow involved in DNA binding.

In conclusion, the present study provides new insights on the functional organization of plant DNA demethylases, revealing that their highly conserved C-terminal region contains a histone reader domain.

Funding

This research was funded by the Spanish Ministry of Science and Innovation (MICINN), as well as the European Regional Development Fund (FEDER), under grants numbers BFU2016-80728-P, PID2019-109967GB-I00, and by the Junta de Andalucía and the European Regional Development Fund under grant PY20_00051. J.T.P.D. was the recipient of a PhD FPU Fellowship from the Spanish Ministry of Education. Funding for open access charge: Universidad de Córdoba/CBUA.

CRediT authorship contribution statement

Parrilla-Doblas J.T.: Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Morales-Ruiz T.:** Software, Validation, Formal analysis, Investigation, Data curation, Writing – review & editing. **Ariza R.R.:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Supervision, Project administration. **Martínez-MacíasM.I.:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Supervision. **Roldán-ArjonaT.:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

We are grateful to members of our lab for helpful criticism and advice.

Conflicts of interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.dnarep.2022.103341.

References

[1] J.A. Law, S.E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals, Nat. Rev. Genet 11 (2010) 204–220.
[2] R. Lister, M. Pelizzola, R.H. Dowen, R.D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A.H. Millar, J.A. Thomson, B. Ren, J.R. Ecker, Human DNA methylomes at base resolution show widespread epigenomic differences, Nature 462 (2009) 315–322.
[3] K.E. Varley, J. Gertz, K.M. Bowling, S.L. Parker, T.E. Reddy, F. Pauli-Behn, M. K. Cross, B.A. Williams, J.A. Stamatoyannopoulos, G.E. Crawford, D.M. Absher, B. J. Wold, R.M. Myers, Dynamic DNA methylation across diverse human cell lines and tissues, Genome Res. 23 (2013) 555–567.
[4] C. Kress, H. Thomassin, T. Grange, Local DNA demethylation in vertebrates: how could it be performed and targeted? FEBS Lett. 494 (2001) 135–140.
[5] J.T. Parrilla-Doblas, T. Roldan-Arjona, R.R. Ariza, D. Cordoba-Canero, Active DNA demethylation in plants, Int. J. Mol. Sci. 20 (2019).
[6] Z. Gong, T. Morales-Ruiz, R.R. Ariza, T. Roldan-Arjona, L. David, J.K. Zhu, ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase, Cell 111 (2002) 803–814.
[7] T. Morales-Ruiz, A.P. Ortega-Galisteo, M.I. Ponferrada-Marin, M.I. Martinez-Macias, R.R. Ariza, T. Roldan-Arjona, DEMETER and repressor of silencing 1 encode 5-methylcytosine DNA glycosylases, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 6853–6858.

[8] F. Agius, A. Kapoor, J.K. Zhu, Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 11796–11801.

[9] Y. Choi, M. Gehring, L. Johnson, M. Hannon, J.J. Harada, R.B. Goldberg, S. E. Jacobsen, R.L. Fischer, DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis, Cell 110 (2002) 33–42.

[10] J. Penterman, D. Zilberman, J.H. Huh, T. Ballinger, S. Henikoff, R.L. Fischer, DNA demethylation in the Arabidopsis genome, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 6752–6757.

[11] A.P. Ortega-Galisteo, T. Morales-Ruiz, R.R. Ariza, T. Roldan-Arjona, Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks, Plant Mol. Biol. 67 (2008) 671–681.

[12] H.M. Nash, S.D. Bruner, O.D. Scharer, T. Kawate, T.A. Addona, E. Spooner, W. S. Lane, G.L. Verdine, Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily, Curr. Biol. 6 (1996) 968–980.

[13] M.I. Ponferrada-Marín, J.T. Parrilla-Doblas, T. Roldán-Arjona, R.R. Ariza, A discontinuous DNA glycosylase domain in a family of enzymes that excise 5-methylcytosine, Nucleic Acids Res. 39 (2011) 1473–1484.

[14] Y.G. Mok, R. Uzawa, J. Lee, G.M. Weiner, B.F. Eichman, R.L. Fischer, J.H. Huh, Domain structure of the DEMETER 5-methylcytosine DNA glycosylase, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 19225–19230.

[15] M.I. Ponferrada-Marín, M.I. Martínez-Macías, T. Morales-Ruiz, T. Roldán-Arjona, R.R. Ariza, Methylation-independent DNA binding modulates specificity of repressor of silencing 1 (ROS1) and facilitates demethylation in long substrates, J. Biol. Chem. 285 (2010) 23032–23039.

[16] M.I. Ponferrada-Marín, T. Roldán-Arjona, R.R. Ariza, Demethylation initiated by ROS1 glycosylase involves random sliding along DNA, Nucleic Acids Res. 40 (2012) 11554–11562.

[17] S. Hong, H. Hashimoto, Y.W. Kow, X. Zhang, X. Cheng, The carboxy-terminal domain of ROS1 is essential for 5-methylcytosine DNA glycosylase activity, J. Mol. Biol. 426 (2014) 3703–3712.

[18] J. Zhu, A. Kapoor, V.V. Sridhar, F. Agius, J.K. Zhu, The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in Arabidopsis, Curr. Biol. 17 (2007) 54–59.

[19] J.J. Hayes, J.C. Hansen, Nucleosomes and the chromatin fiber, Curr. Opin. Genet Dev. 11 (2001) 124–129.

[20] T. Kouzarides, Chromatin modifications and their function, Cell 128 (2007) 693–705.

[21] K. Tang, Z. Lang, H. Zhang, J.K. Zhu, The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications, Nat. Plants 2 (2016) 16169.

[22] W.F. Nie, M. Lei, M. Zhang, K. Tang, H. Huang, C. Zhang, D. Miki, P. Liu, Y. Yang, X. Wang, H. Zhang, Z. Lang, N. Liu, X. Xu, R. Yelagandula, H. Zhang, Z. Wang, X. Chai, A. Andreucci, J.Q. Yu, F. Berger, R. Lozano-Duran, J.K. Zhu, Histone acetylation recruits the SWR1complex to regulate active DNA demethylation in Arabidopsis, Proc. Natl. Acad. Sci. U.S.A. 116 (2019) 16641–16650.

[23] W. Qian, D. Miki, H. Zhang, Y. Liu, X. Zhang, K. Tang, Y. Kan, H. La, X. Li, S. Li, X. Zhu, X. Shi, K. Zhang, O. Pontes, X. Chen, R. Liu, Z. Gong, J.K. Zhu, A histone acetyltransferase regulates active DNA demethylation in Arabidopsis, Science 336 (2012) 1445–1448.

[24] Z. Lang, M. Lei, X. Wang, K. Tang, D. Miki, H. Zhang, S.K. Mangrauthia, W. Liu, W. Nie, G. Ma, J. Yan, C.G. Duan, C.C. Hsu, C. Wang, W.A. Tao, Z. Gong, J.K. Zhu, The methyl-CpG-binding protein MBD7 facilitates active DNA demethylation to limit DNA hyper-methylation and transcriptional gene silencing, Mol. Cell 57 (2015) 971–983.

[25] C.G. Duan, X. Wang, S. Xie, L. Pan, D. Miki, K. Tang, C.C. Hsu, M. Lei, Y. Zhong, Y. J. Hou, Z. Wang, Z. Zhang, S.K. Mangrauthia, H. Xu, H. Zhang, B. Dilkes, W.A. Tao, J.K. Zhu, A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation, Cell Res. 27 (2017) 226–240.

[26] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254.

[27] K. Zhang, V.V. Sridhar, J. Zhu, A. Kapoor, J.K. Zhu, Distinctive core histone post-translational modification patterns in Arabidopsis thaliana, PLoS One 2 (2007), e1210.

[28] R. Cao, L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R.S. Jones, Y. Zhang, Role of histone H3 lysine 27 methylation in Polycomb-group silencing, Science 298 (2002) 1039–1043.

[29] H.K. Long, N.P. Blackledge, R.J. Klose, ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection, Biochem. Soc. Trans. 41 (2013) 727–740.

[30] J.T. Parrilla-Doblas, M.I. Ponferrada-Marin, T. Roldan-Arjona, R.R. Ariza, Early steps of active DNA demethylation initiated by ROS1 glycosylase require three putative helix-invading residues, Nucleic Acids Res. 41 (2013) 8654–8664.

[31] X. Zheng, O. Pontes, J. Zhu, D. Miki, F. Zhang, W.X. Li, K. Iida, A. Kapoor, C. S. Pikaard, J.K. Zhu, ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis, Nature 455 (2008) 1259–1262.

[32] W. Fischle, Y. Wang, S.A. Jacobs, Y. Kim, C.D. Allis, S. Khorasanizadeh, Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1chromodomains, Genes Dev. 17 (2003) 1870–1881.

[33] H. Cerutti, J.A. Casas-Mollano, Histone H3 phosphorylation: universal code or lineage specific dialects? Epigenetics 4 (2009) 71–75.

[34] A. Houben, D. Demidov, A.D. Caperta, R. Karimi, F. Agueci, L. Vlasenko, Phosphorylation of histone H3 in plants–a dynamic affair, Biochim. Biophys. Acta 2007 (1769) 308–315.

[35] D. Rossetto, N. Avvakumov, J. Cote, Histone phosphorylation: a chromatin modification involved in diverse nuclear events, Epigenetics 7 (2012) 1098–1108.

[36] A. Sawicka, C. Seiser, Histone H3 phosphorylation - a versatile chromatin modification for different occasions, Biochimie 94 (2012) 2193–2201.

[37] W. Fischle, Y. Wang, C.D. Allis, Binary switches and modification cassettes in histone biology and beyond, Nature 425 (2003) 475–479.

[38] R.A. Varier, N.S. Outchkourov, P. de Graaf, F.M. van Schaik, H.J. Ensing, F. Wang, J.M. Higgins, G.J. Kops, H.T. Timmers, A phospho/methyl switch at histone H3 regulates TFIID association with mitotic chromosomes, EMBO J. 29 (2010) 3967–3978.

[39] J.Y. Kim, K.B. Kim, H.J. Son, Y.C. Chae, S.T. Oh, D.W. Kim, J.H. Pak, S.B. Seo, H3K27 methylation and H3S28 phosphorylation-dependent transcriptional regulation by INHAT subunit SET/TAF-Ibeta, FEBS Lett. 586 (2012) 3159–3165.

[40] K. Arita, S. Isogai, T. Oda, M. Unoki, K. Sugita, N. Sekiyama, K. Kuwata, R. Hamamoto, H. Tochio, M. Sato, M. Ariyoshi, M. Shirakawa, Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 12950–12955.

[41] S.S. Gehani, S. Agrawal-Singh, N. Dietrich, N.S. Christophersen, K. Helin, K. Hansen, Polycomb group protein displacement and gene activation through MSK-dependent H3K27me3S28 phosphorylation, Mol. Cell 39 (2010) 886–900.

[42] W. Fischle, B.S. Tseng, H.L. Dormann, B.M. Ueberheide, B.A. Garcia, J. Shabanowitz, D.F. Hunt, H. Funabiki, C.D. Allis, Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation, Nature 438 (2005) 1116–1122.

[43] A. Jeltsch, J. Broche, P. Bashtrykov, Molecular processes connecting DNA methylation patterns with dna methyltransferases and histone modifications in mammalian genomes, Genes 9 (2018).

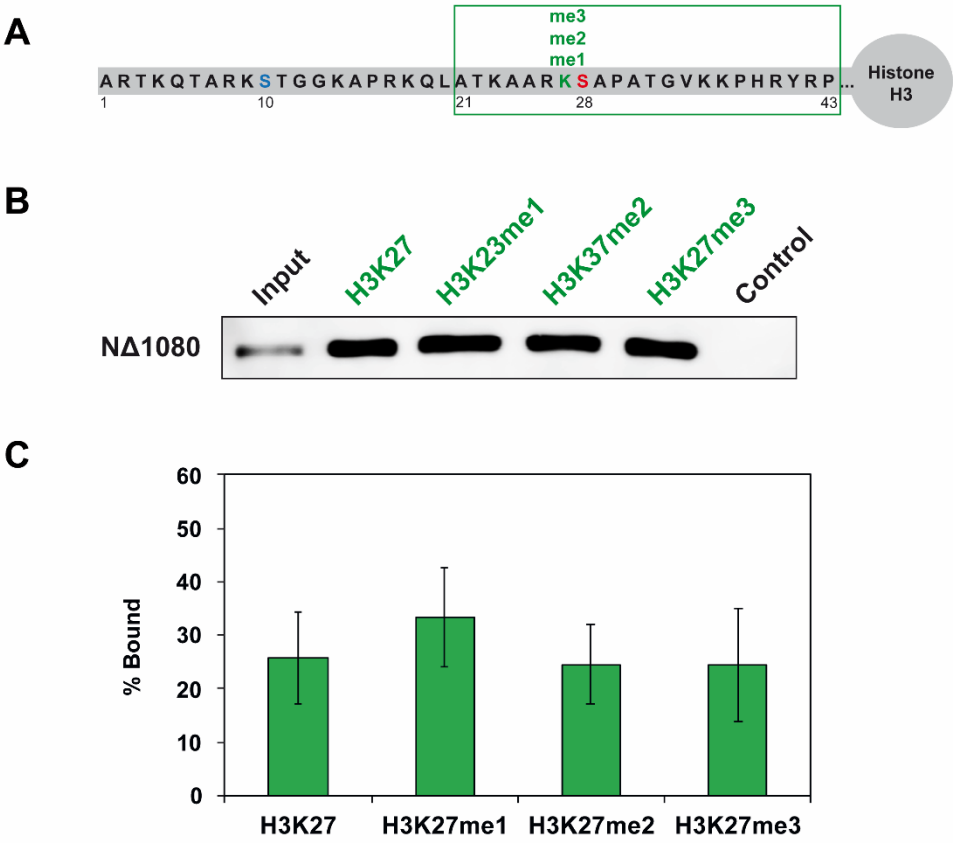
[44] M.L. Hohenstatt, P. Mikulski, O. Komarynets, C. Klose, I. Kycia, A. Jeltsch, S. Farrona, D. Schubert, PWWP-domain interactor of polycombs1 Interacts with POlycomb-group proteins and histones and regulates arabidopsis flowering and development, Plant Cell 30 (2018) 117–133.

[45] C.A. Musselman, M.E. Lalonde, J. Cote, T.G. Kutateladze, Perceiving the epigenetic landscape through histone readers, Nat. Struct. Mol. Biol. 19 (2012) 1218–1227.

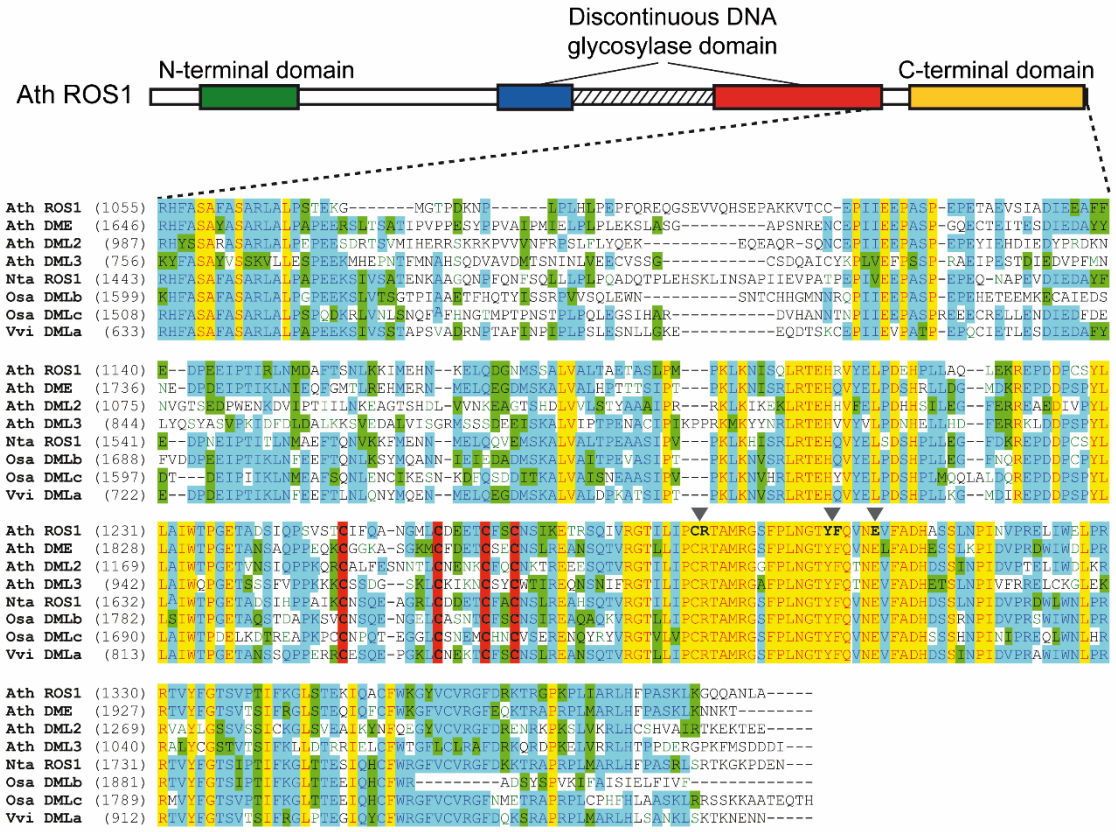
[46] D.J. Patel, Z. Wang, Readout of epigenetic modifications, Annu Rev. Biochem. 82 (2013) 81–118.

[47] S. Zhao, B. Zhang, M. Yang, J. Zhu, H. Li, Systematic profiling of histone readers in Arabidopsis thaliana, Cell Rep. 22 (2018) 1090–1102.

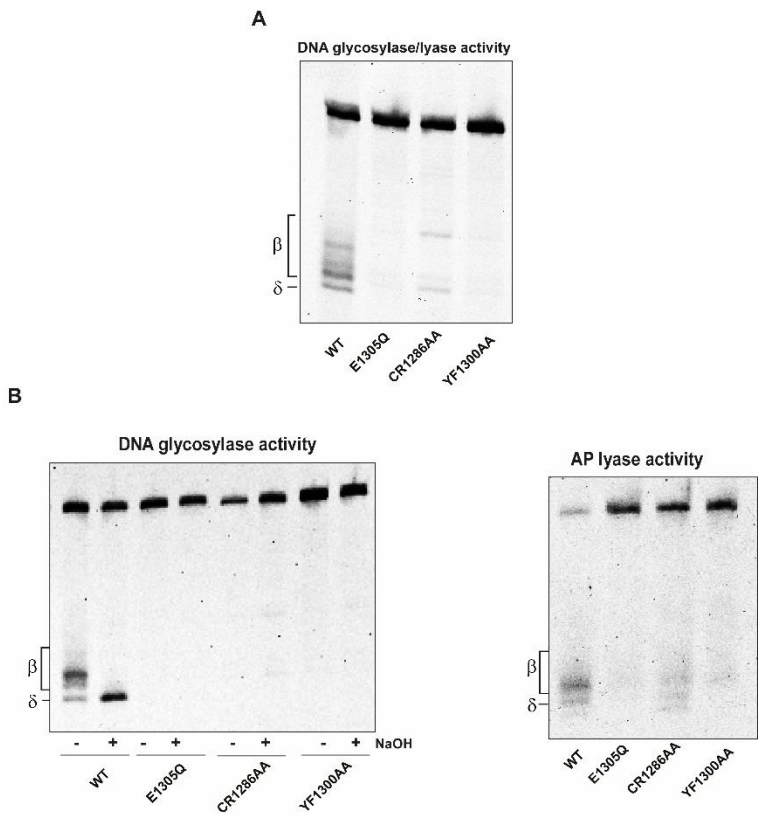
[48] S.B. Rothbart, B.D. Strahl, Interpreting the language of histone and DNA modifications, Biochim. Biophys. Acta 2014 (1839) 627–643.



Supplementary Figure 1. Analysis of ROS1 C-terminal domain interaction with H3 peptides containing different methylation states of Lys27. (A) Schematic diagram of a region of the N-terminal tail of histone H3. Box indicates the sequence of histone H3 peptides (either unmodified or methylated at Lys27, in green) used in pull-down assays. (B) Pull-down assay of ROS1NΔ1080 with biotinylated H3 histone peptides (21-43 amino acids) unmodified, mono-, di- or tri-methylated at Lys27. Peptides were fixed to streptavidin dynabeads and incubated with His_6-tagged ROS1NΔ1080. Proteins associated to beads were separated by SDS-PAGE, transferred to a nitrocellulose membrane and detected with an anti-His_6-tag antibody. Beads with no peptide were used as control for non-specific binding. (C) Graph shows the percentage of bound protein to each histone peptide, normalized to input. Values are the mean from two independent experiments.



Supplementary Figure 2. Multiple sequence alignment of the C-terminal domain of several DML family proteins. The upper schematic diagram shows the conserved regions among members of the DML family: a N-terminal lysine-rich region (green), a non-contiguous DNA glycosylase domain distributed over two segments (blue and red) separated by a non-structured linker region (striped), and a highly conserved C-terminal domain (yellow) that is not found in any other protein family. The C-terminal domain contains two conserved sub-regions separated by 4 invariant Cys residues (highlighted in red). Inverted grey triangles indicate mutated amino acids. Names of organisms are abbreviated as follows: Ath, *Arabidopsis thaliana*; Nta, *Nicotiana tabacum*; Osa, *Oryza sativa*; Vvi, *Vitis vinifera*. Genbank accession numbers are as follows: Ath ROS1: AAP37178; Ath DME: ABC61677; Nta ROS1: BAF52855; Osa DMLB: BAF04322; Osa DMLC: EEE63898; Osa DMLA: BAD23025; Vvi DMLA: CAO46558. The alignment was performed with the Clustal W algorithm of Align X (Vector NTI Suite, version 11.0, Invitrogen).



Supplementary Figure 3. Representative gel images of ROS1 C-terminal mutant proteins catalytic activity analysis. (A) Combined DNA glycosylase/AP lyase activity of WT ROS1 and C-terminal mutant variants on a DNA substrate containing a 5-meC: G pair. Purified proteins (20 nM) were incubated at 30°C for 24 h with a double-stranded oligonucleotide substrate (20 nM) containing 5-meC opposite G. Products were separated in a 12% urea denaturing polyacrylamide gel and visualized by fluorescent scanning by a FLA-5100 imager (Fujifilm). (B) Left, DNA glycosylase assay. The generation of incision products was measured by incubating purified WT ROS1 or mutant variants (20 nM) at 30°C for 24 h with a double-stranded oligonucleotide substrate (20 nM) containing a single 5-meC:G pair. After incubation, NaOH (100 nM) was either added or not, as indicated, and samples were immediately transferred to 90°C for 10 min. Right, AP lyase assay. A double-stranded oligonucleotide substrate containing an AP site opposite G (20 nM) was incubated at 30°C for 2 h in the presence of purified WT ROS1 or mutant variants (20 nM). Samples were treated with $NaBH_4$ (300 mM) at 0°C for 30 min to stabilize non-processed AP sites and neutralized with 100 mM Tris–HCl, pH 7.4. Products were separated in a 12% urea denaturing polyacrylamide gel and the amount of incised oligonucleotide was visualized by fluorescent scanning by a FLA-5100 imager (Fujifilm). Indicated are the β,δ-elimination products generated by ROS1 after excision of the 5-meC as a free base and cleavage of the phosphodiester backbone at 5-meC removal site.

Table 1. Oligonucleotides used as primers for generation of ROS1 mutant variants.

Name	Sequence (5'-3')[a]
ROS1CR1286AA-F	GAGGGACAATTTTGATTCCTGCTGCAACAGCGATGAGGGGTAG
ROS1CR1286AA-R	CTACCCCTCATCGCTGTTGCAGCAGGAATCAAAATTGTCCCTC
ROS1YF1300AA-F	CCTCTAAATGGAACGGCCGCTCAAGTAAATGAGGTGTTTGCG
ROS1YF1300AA-R	CGCAAACACCTCATTTACTTGAGCGGCCGTTCCATTTAGAGG
ROS1E1305Q-F	CTTTCAAGTAAATCAGGTGTTTGCGGATCATGCATCCAG
ROS1E1305Q-R	CTGGATGCATGATCCGCAAACACCTGATTTACTTGAAAG

[a]Underlined is indicated the mutagenized codon.

Table 2. Biotinylated histone peptides used in pull-down assays.

Name	Sequence (N-terminal to C-terminal)[a]	MW (Da)
H3S10	ARTKQTARKSTGGKAPRKQLA-GGK-biotin	2724
H3S10P	ARTKQTARK**[pS]**TGGKAPRKQLA-GGK-biotin	2801.7
H3S28	Ac-ATKAARKSAPATGGVKKPHRYRP-GGK-biotin	2959.51
H3S28P	Ac-ATKAARK**[pS]**APATGGVKKPHRYRP-GGK-biotin	3039.49
H3K27me1	ATKAAR**[me1K]**SAPATGGVKKPHRYRP-GGK-biotin	2932.9
H3K27me2	ATKAAR**[me2K]**SAPATGGVKKPHRYRP-GGK-biotin	2947.6
H3K27me3	ATKAAR**[me3K]**SAPATGGVKKPHRYRP-GGK-biotin	2961

[a] Ac = Acetylation; post-translational histone modifications are colored. p: phosphorylation; me: methylation

Table 3. Oligonucleotides used as substrates.

Name	DNA Sequence	Strand	X=	Length (bp)
FL-CGF	5´TCACGGGATCAATGTGTTCTTTCAGCTCX GGTCACGCTGACCAGGAATACC 3´	Upper	C	51
FL-meCGF	5'TCACGGGATCAATGTGTTCTTTCAGCTCX GGTCACGCTGACCAGGAATACC 3´	Upper	5-meC	51
FL-UGF	5'TCACGGGATCAATGTGTTCTTTCAGCTCX GGTCACGCTGACCAGGAATACC 3	Upper	U	51
CGR	3´AGTGCCCTAGTTACACAAGAAAGTCGAGX CCAGTGCGACTGGTCCTTATGG 5´	Lower	G	51