Classification: Biological Sciences. Biochemistry

DEMETER and *REPRESSOR OF SILENCING 1* encode 5-methylcytosine-DNA glycosylases

Teresa Morales-Ruiz, Ana Pilar Ortega-Galisteo, María Isabel Ponferrada-Marín, María Isabel Martínez-Macías, Rafael R. Ariza and Teresa Roldán-Arjona

Departamento de Genética. Universidad de Córdoba. 14071-Córdoba, SPAIN

Corresponding Author:
Teresa Roldán-Arjona
Departamento de Genética
Edificio Gregor Mendel
Campus de Rabanales s/n
Universidad de Córdoba
14071-Córdoba
SPAIN

Tel: +34 957 218 979
Fax: +34 957 212 072
e-mail: ge2roarm@uco.es

Manuscript information: 20 text pages, 6 figures and 1 Supplementary Table

Word and character count: 237 words in Abstract and 46,971 total number of characters in the paper

Abbreviations: 5-meC, 5-methylcytosine; 8-OG, 7,8-dihydro-8-oxoguanine; *DME*, *DEMETER*; *ROS1*, *REPRESSOR OF SILENCING 1*; TDG, thymine-DNA glycosylase; MBD4, methyl CpG binding protein 4; AtOGG1, *Arabidopsis thaliana* 8-OG-DNA glycosylase.

Data deposition: The sequence reported in this paper for *DME* cDNA has been deposited in the GenBank database (accession number DQ335243)

Abstract

Cytosine methylation is an epigenetic mark that promotes gene silencing and plays important roles in development and genome defense against transposons. Methylation patterns are established and maintained by DNA-methyltransferases that catalyze transfer of a methyl group from S-adenosyl-L-methionine to cytosine bases in DNA. Erasure of cytosine methylation occurs during development, but the enzymatic basis of active demethylation remains controversial. In *Arabidopsis thaliana*, *DEMETER (DME)* activates the maternal expression of two imprinted genes silenced by methylation, and *REPRESSOR OF SILENCING 1 (ROS1)* is required for release of transcriptional silencing of a hypermethylated transgene. *DME* and *ROS1* encode two closely related DNA glycosylase domain proteins, but it is unknown if they participate directly in a DNA demethylation process or counteract silencing through an indirect effect on chromatin structure. Here we show that DME and ROS1 catalyze the release of 5-meC from DNA by a glycosylase/lyase mechanism. Both enzymes also remove thymine, but not uracil, mismatched to guanine. DME and ROS1 show a preference for 5-meC over thymine in the symmetric dinucleotide CpG context, where most plant DNA methylation occurs. Nevertheless, they also have significant activity on both substrates at CpApG and asymmetric sequences, which are additional methylation targets in plant genomes. These findings suggest that a function of ROS1 and DME is to initiate erasure of 5-meC through a base excision repair process, and provide strong biochemical evidence for the existence of an active DNA demethylation pathway in plants.

Introduction

Methylation of cytosine at carbon 5 of the pyrimidine ring (5-meC) is an epigenetic modification that guides formation of transcriptionally silent chromatin and allows transmission of specific patterns of gene activity across cellular divisions (1, 2). In eukaryotes, DNA methylation is detected in protists, fungi, plants and animals (3), and plays important roles in the establishment of developmental programs (4, 5) and in genome defense against parasitic mobile elements (6). Most of mammalian and plant DNA methylation is restricted to symmetrical CpG sequences, but plants also have significant levels of cytosine methylation in the symmetric context CpNpG (where N is any nucleotide) and even in asymmetric contexts (2, 7). Similarly to other biochemical modifications such as protein phosphorylation and acetylation, DNA methylation is also reversible. Demethylation may take place as a passive process due to lack of maintenance methylation during several cycles of DNA replication, or as an active mechanism in the absence of replication (8). In mammalian preimplantation embryos the maternal genome is demethylated by a passive process along cleavage stages, whereas the paternal genome is demethylated by an active mechanism immediately after fertilization (9). In addition to global demethylation, site-specific local demethylation also occurs throughout development and tissue differentiation (10).

In contrast to the well studied genetics, biochemistry and biology of cytosine-DNA-methyltransferases, the enzymatic basis of active demethylation has remained elusive. Direct removal of the methyl group from 5-meC residues has been claimed (11) but this mechanism, which involves a thermodynamically unfavorable breakage of the carbon-carbon bond, has been questioned (12) and could not be independently reproduced (13, 14). Another proposed mechanism is disruption of the labile N-

glycosidic bond between the 5-meC base and the deoxyribose moiety in DNA, followed by replacement with an unmodified cytosine. A 5-meC-DNA glycosylase activity was first identified in chicken embryos (15) and found to copurify with a protein homologous to human thymine DNA glycosylase (TDG) (16, 17). It was latter reported that methyl CpG binding protein 4 (MBD4), another human DNA glycosylase with no sequence similarity to TDG, also has 5-methylcytosine-DNA glycosylase activity (18). Both TDG and MBD4 are monofunctional DNA glycosylases that show a preference for U·G and T·G mismatches within a CpG context (19, 20). However, they have been shown to have a very weak activity on 5meC·G pairs compared to their activities towards U·G and T·G mismatches (18, 20, 21), and hence their roles in DNA demethylation remain unclear. It has been also suggested that demethylation might be achieved indirectly though deamination of 5-meC by Activation induced cytidine deaminase (Aid) and repair of the ensuing T·G mismatch (22).

Recent work in plants has added genetic evidence for a role of DNA glycosylases in an active demethylation pathway. *DME* was identified in a search for mutations causing parent-of-origin effects on seed viability (23). *DME* is expressed primarily in the central cell of the female gametophyte, where it is required for the expression of the maternal alleles of the imprinted genes *MEA* and *FWA* (23, 24). At least in the case of *MEA* imprinting, mutations in the methyltransferase gene *MET1* suppress the requirement for *DME* (25). *ROS1* was identified in a screen for mutants with deregulated expression of the repetitive *RD29A-LUC* transgene (26). Whereas in wild plants the transgene and the homologous endogenous gene are expressed, *ros1* mutants display transcriptional silencing and hypermethylation of both loci (26). *ROS1* and *DME* are predicted to encode large proteins containing a DNA glycosylase domain with significant sequence similarity to base excision DNA repair proteins in the HhH-GPD

superfamily (27) (Fig.1). The presence of a conserved invariant lysine in the HhH motif suggests that DME and ROS1 belong to the subset of DNA glycosylases/lyases, able both to hydrolyze the N-glycosidic bond linking bases to DNA and to cleave the phosphodiester backbone at the site where a base has been removed. A recombinant truncated form of ROS1 was found to have strand-breaking activity of methylated plasmid DNA (26), and ectopic *DME* expression in plants results in nicks in the *MEA* promoter (23). Furthermore, mutation of an invariant aspartic acid in the HhH-GPD motif of DME suppresses its capacity to activate *MEA* gene transcription in the central cell (28).

The effects of ROS1 and DME in counteracting DNA methylation and gene silencing may be explained by distinct mechanisms. One possibility is that they exert an indirect effect due to interactions with transcription factors and or chromatin-modifying proteins, perhaps promoted by DNA incisions that facilitate local chromatin remodeling. A second possibility is that ROS1 and DME directly excise 5-meC from DNA, initiating replacement by an unmethylated cytosine, and leading to hypomethylation and gene activation. Our biochemical characterization of the enzymatic activity of ROS1 and DME supports the latter mechanism.

Results

DME and ROS1 excise 5-mC and mismatched thymine from DNA

We cloned full-length cDNAs corresponding to Arabidopsis *DME* and *ROS1* genes. The protein sequence encoded by the full-length *DME* cDNA is identical to that reported by Choi and colleagues (23) but includes 258 additional amino acids at the amino terminus (Fig. 1). Comparison of the *DME* cDNA sequence reported here (accession number DQ335243) and the sequence previously described (23) reveals that

the latter comprises two additional segments with consensus donor and acceptor splice sites that probably represent introns. Hence, both cDNA sequences may correspond to alternative splicing forms of the *DME* transcript.

We expressed recombinant full-length DME and ROS1 and the purified proteins were used in oligonucleotide incision assays on DNA substrates containing 5-meC in diverse sequence contexts. Both DME and ROS1 incised a radioactively labeled oligonucleotide duplex containing a single 5-meC at an internal position in a CpCpGpGp sequence (Fig 2*A*). Methylated cytosines located at the external position were also processed, but with a reduced efficiency. Incision was also observed in the symmetrical context CpApG and in an asymmetric context. In all symmetric contexts the incision activity on hemimethylated and bimethylated DNA was not significantly different (Fig. 2*A* and data not shown). Hydrolytic deamination of cytosine produces uracil mismatched to guanine, whereas deamination of 5-meC generates mismatched thymine (29). To determine if DME and ROS1 excise 5-meC as a free base or whether it is transformed in thymine before removal, we analyzed the excision products by two-dimensional thin-layer chromatography (See Methods). Incubation of DME or ROS1 with a DNA substrate containing ^{3}H-labelled 5-meC resulted in release of ethanol-soluble radioactive material that co-migrated with a non-radioactive 5-meC standard in both dimensions, whereas no radioactivity was detected co-migrating with thymine (Fig 2*B*). From these experiments we conclude that both DME and ROS1 are 5-meC-DNA glycosylases.

In addition to 5-meC paired to guanine, DME and ROS1 also removed thymine from a T·G mismatch locater either in CpG or non-CpG sequence contexts (Fig 3*A*). However, no activity was detected on uracil either mismatched to guanine (Fig. 3*A*) or paired to adenine (data not shown). To verify that the observed DNA incision activity is

intrinsic to DME and ROS1 we generated mutant proteins in which a conserved aspartic acid residue in the glycosylase domain (Fig. 1) was changed to alanine. Preparations of both mutant proteins, purified by the same procedure as the wild-type proteins described above, lacked any detectable incision activity (Fig 3*B*).

DME and ROS1 have a preference for 5-meC in a CpG sequence context.

We wished to determine the relative activities of DME and ROS1 towards substrates containing a 5-meC·G pair or a T·G mispair positioned in different sequence contexts. Time course experiments revealed that 5-meC and T·G mismatches were efficiently processed by both proteins when located in a CpG sequence context, but were removed at a notably slower rate when present at the external position of the sequence CpCpGpG (Fig. 4). The incision activity of DME and ROS1 was also significant in the symmetrical sequence CpApG and in an asymmetric context (Fig 4). For both proteins, base preference was dependent on the sequence context. Thus, DME processed 5-meC at a higher rate than thymine in all symmetrical sequence contexts, whereas in the asymmetrical context both bases were removed with similar efficiency (Fig 4). ROS1 showed a very similar behavior, with the difference that excised 5-meC and thymine from a CpApG context at comparable rates. We conclude that both DME and ROS1 show a preference for 5-meC over thymine in a CpG context, and have significant activity on both base substrates at CpApG and asymmetric sequences, which are additional but less frequent sites of cytosine methylation in plants.

DME and ROS1 process 5-meC though a DNA glycosylase/lyase mechanism

Repair reactions catalyzed by DNA glycosylases/lyases proceed through a transient imine intermediate (Schiff base) that can be reduced by borohydride, leading to the formation of an irreversibly cross-linked complex between enzyme and substrate.

This borohydride-dependent ‘trapping assay’ is a convincing test for catalytic processing of a substrate by a DNA glycosylase/lyase (30). To determine if the excision of 5-meC catalyzed by DME and ROS proceeds through a DNA glycosylase/lyase mechanism, reactions with a methylated DNA substrate were performed in the presence of $NaBH_4$ and products were analyzed by SDS-PAGE (Fig. 5). A DNA duplex with a single 7,8-dihydro-8-oxoguanine (8-OG) paired to cytosine located at the same position as the 5-mC·G pair was used as a control in reactions with 8-OG-DNA glycosylase AtOGG1 from Arabidopsis (31). The upper shifted bands indicate formation of the cross-linked complex between the labeled oligonucleotide containing 5-meC and DME (Fig 5, lane 4) or ROS1 (Fig 5, lane 5). As expected, the complex for AtOGG1 (molecular mass = 44 kDa) migrated faster than those for DME (227 kDa) or ROS1 (200 kDa). No shifted bands were observed when DME or ROS1 were incubated in the presence of nonmethylated DNA or the duplex containing 8-OG, indicating that the Schiff base formation was specific to the DNA substrate containing 5-meC. These results indicate that DME and ROS1 process 5-meC though a DNA glycosylase/lyase mechanism.

Analysis of DNA termini generated upon excision of 5-meC

To identify the products of the 5-meC excision reaction, we compared the activity of DME on a single 5-meC·G pair in a CpG context to the activities of the 8-oxoguanine-DNA glycosylases Fpg (32) and AtOGG1 on an 8-OG·C pair placed at the same position. As expected, the incubation of the 8-OG substrate with Fpg exclusively resulted in a fragment (Fig 6*A*, lane 3) migrating slightly faster than the 28-mer marker. This fragment is the product of successive β- and δ-elimination reactions at the abasic site resulting from the excision of 8-OG, leaving a single nucleoside gap limited by 3´- and 5´-phosphate ends in DNA. Incubation with AtOGG1 generated a main product

with a mobility similar to that of the 29-mer marker (Fig 6*A*, lane 4), result of a β-elimination that leaves an α, β-unsaturated aldehyde at the 3´ terminus of the nucleoside gap. The additional minor bands are presumably due to isomerization of the 3´hydroxypentenal terminus (33). Incubation of the 5-meC substrate with DME during 60 min generated a fragment co-migrating with the AtOGG1 β-elimination product (Fig 6*A*, lane 5). However, we found that longer incubation times generated an additional fragment (Fig 6*B*) co-migrating with the Fpg β, δ-elimination product (Fig. 6*A*, lane 7). Upon addition of *E. coli* Endo IV, which has a 3´-phosphodiesterase activity, both products were converted to 3´-OH fragments co-migrating with the 29-mer marker. Similar results were obtained with ROS1 (data not shown). Therefore, the products generated by DME and ROS1 upon excision of 5-meC are likely to be a mixture of β- and β, δ-elimination products.

Discussion

There is accumulating evidence of multiple connections between DNA glycosylases and the processes that regulate gene expression. A primary mechanism involves interactions with proteins that regulate transcription. It has been reported that methyl purine DNA glycosylase associates with the transcriptional repressor methyl CpG binding protein 1 (MBD1) to cooperatively inhibit gene expression (34), and 3-methyladenine DNA glycosylase interacts with estrogen receptor alpha to modulate estrogen-mediated transcription (35). There is also evidence that TDG modulates transcription though interactions with hormone receptors (36, 37), transcription factors (38) and coactivators such as the CPB/p300 acetylase complex (39). MBD4 has the ability to repress transcription through an interaction with Sin3A and HDAC1 (40).

But, in addition to protein-protein interactions, DNA glycosylases may also modulate gene expression by direct modification of the methylation status of DNA. A role for DNA glycosylases in genome demethylation during cell differentiation in vertebrates has been previously suggested (15), and an active demethylation pathway initiated by TDG and/or MBD4 DNA glycosylases has been proposed in animal cells (16, 18, 41, 42). However, it has been argued that the main *in vivo* role for both proteins is to counteract the mutagenic potential of 5-meC and C deamination in CpG sequences (20, 43), given their high efficiency on U·G and T·G mismatches (17, 19, 20), compared to their weak activity on 5-meC·G base pairs (18, 21). DME and ROS1 are structurally unrelated to TDG, which belongs to a large family of uracil-DNA glycosylases different from the HhH-GPD family (44), but share with MBD4 a HhH-GPD DNA glycosylase domain located at the C-terminal half of the protein. However, unlike MBD4, neither DME nor ROS1 have a methyl-CpG binding domain (20). In contrast to the strong substrate specificity of TDG and MBD4 for T·G and U·G mismatches, DME and ROS1 show a preference for 5-meC over a T·G mismatch in a CpG sequence context, the most frequent DNA methylation target in plant and animal genomes, and they do not display detectable activity on U·G mispairs. Thus, the biochemical properties of DME and ROS1, together with the available genetic evidence, suggest that an important role for both enzymes *in vivo* is excision of 5-meC. We cannot rule out the possibility that DME and ROS1 also play a role in neutralizing the mutagenic deamination of 5-meC to thymine through their activity on T·G mismatches. The effect of *dme* and *ros1* mutations on mutagenesis *in vivo* has not been assessed, but could be compounded by their epigenetic effects on plant development.

We have found that DME and ROS1 cleave the phosphodiester backbone at the 5-meC removal site by successive beta-delta elimination, leaving an abasic site that has to

be further processed to generate a 3´- OH terminus suitable for polymerization and ligation. It has been questioned whether a base excision repair mechanism would explain the active global demethylation observed in early mammalian embryogenesis, since that would require a base excision/substitution at a global scale (45). Unlike vertebrates, in plants there is no clear evidence for a global demethylation during the life cycle (7). We propose that the base excision process initiated by DME and ROS1 may be rather involved in local demethylation of selected loci during development. In this scenario, DME would be required to specifically initiate erasure of 5-meC at *MEA* and *FWA*, and perhaps other unidentified loci, in female gametes before fertilization. ROS1 is needed to prevent transcriptional gene silencing and hypermethylation of a repetitive transgene, but the observation of developmental abnormalities in *ros1* mutants after inbreeding (26) suggests that it also regulates expression of endogenous loci. Thus, the available evidence suggests that the two enzymes may have specialized roles in counteracting methylation and gene silencing at specific loci.

This leads to the important question of how the 5-meC-DNA glycosylase activity of DME and ROS1 is directed to specific target genes. Both DME and ROS1 are unusually large proteins much bigger than typical DNA glycosylases (23, 26). In addition to the DNA glycosylase domain, DME and ROS1 show a high sequence similarity at the carboxy-terminal region (Fig. 1). However, they are mostly unrelated in the rest of the sequence, where the only recognizable feature is a region rich in basic residues at the amino-terminal region that displays a weak similarity to H1 histones. It is possible that the divergent domains of DME and ROS1 play a role in targeting the two proteins to different genome loci, either directly or through interactions with associated cofactors, such as chromatin components. The biochemical demonstration that DME

and ROS1 can initiate an active demethylation process through 5-meC excision will help to unravel the intricacies of this mechanism of epigenetic regulation.

Materials and methods

Cloning of *DME and ROS1* full-length cDNAs. Truncated clones lacking the 5' portions of *DME* and *ROS1* cDNAs were isolated from an Arabidopsis cDNA library (46) (Arabidopsis Biological Resource Center, Ohio State University). The 5´ends of both cDNAs were obtained by RT-PCR on total RNA isolated from Arabidopsis plants (ecotype Columbia) using primers designed according to the sequence of a 3´-truncated *DME* cDNA clone and to *ROS1* genome sequence information. After sequencing, a PCR error at position 1330 of *DME* cDNA was corrected by site-directed mutagenesis (Stratagene QuickChange™ Site-Directed Mutagenesis Kit). Verified fragments were connected by digestion and ligation, and a full-length cDNA was assembled for each gene.

Protein expression and purification. The full-length *DME* cDNA was inserted into the pET30b expression vector (Novagen) to add a polyhistidine (His_6) Tag at the N-terminus of DME protein. Expression in *E. coli* BL21(DE3) *dcm*$^-$ Codon Plus cells (Stratagene) was induced by the addition of isopropyl-1-thio-β-D-galactopyranoside and the His_6-tagged DME protein was purified by affinity chromatography on a Ni^{2+}-NTA column (Amersham Pharmacia Biotech) using standard protocols. The full-length *ROS1* cDNA was cloned into the pMAL-c2X expression vector (New England Biolabs) to obtain a *malE-ROS1* in-frame fusion. Expression was induced in *E. coli* strain BL21(DE3) *dcm*$^-$ Codon Plus cells as indicated above and the MBP-ROS1 fusion protein was purified by amylose affinity chromatography using standard protocols.

Site-directed mutagenesis. Site-directed mutagenesis was performed using the QuickChange XL site-directed mutagenesis kit (Stratagene). The D1562A mutation was introduced into pET-DME using the oligonucleotides DMED1562A-F: 5´-CACAATCTTGCTTTCCCTGTTGCCACGAATGTTGGAAGG-3´, and DMED1562A-R: 5´-CCTTCCAACATTCGTGGCAACAGGGAAAGCAAGATTGTG-3'. The D971A mutation was introduced into pMAL-ROS1 using the primers ROSD971A-F: 5´-CACCATCTTGCCTTTCCAGTTGCTACAAATGTTGGGCGC-3´ and ROSD971A-R: 5´-GCGCCCAACATTTGTAGCAACTGGAAAGGCAAGATGGTG-3´. The mutant sequences were confirmed by DNA sequencing and the constructs were used to transform *E. coli* strain BL21(DE3) *dcm*$^-$ CodonPlus. Mutant proteins were overexpressed and purified as described above.

DNA substrates. Oligonucleotides used (see Supplementary Table 1) were synthesized by Operon and purified by polyacrylamide gel electrophoresis (PAGE). Upper-strand oligonucleotides were labelled at the 5'-end using T4 polynucleotide kinase (Roche) and [γ-^{32}P]ATP (Amersham Pharmacia Biotech). Double-stranded substrates were prepared by mixing a 0.05 μM solution of the labeled upper strand with a 0.1 μM solution of the complementary lower strand, heating to 95 °C for 5 min and slowly cooling to room temperature.

Enzymatic activity assay. Double-stranded oligodeoxynucleotides (50 fmol) were incubated at 30 °C for 1 hour (unless indicated otherwise) in a reaction mixture containing 50 mM Tris-HCl at pH 8.0, 1 mM EDTA, 1 mM DTT, 0.1 mg/mL BSA, and 25 pmol of protein in a total volume of 100 μL. Reactions were stopped by adding EDTA to 20 mM, sodium dodecyl sulfate to 0.6 %, and proteinase K to 200 μg/mL, and the mixtures were incubated for 30 min at 37ºC. DNA was extracted with

phenol:chloroform:isoamyl alcohol (25:24:1) and ethanol precipitated at -20ºC in the presence of 0.3 mM NaCl and 16 μg/mL glycogen. When required, the DNA was resuspended in 10 μL of EF buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, and 50 mM NaCl), and incubated with endonuclease IV (10 units) (New England Biolabs) at 37ºC for 30 min, and finally extracted and precipitated again as above. Samples were resuspended in 10 μL of formamide dye mix (80% formamide, 1 mg/mL bromophenol blue, 10 mM EDTA), and heated at 95ºC for 5 min. Reaction products were separated in a 12% denaturing polyacrylamide gel containing 7 M urea, visualized by autoradiography and quantitated using a PhosphorImager (Bio-Rad).

Thin-layer chromatography analysis. Plasmid pBluescript DNA (2 μg) (Stratagene) was methylated with SssI methylase (New England Biolabs) in the presence of 5 μCi S-adenosyl-L-[methyl-^{3}H]methionine (15 Ci/mmol) (Amersham Biosciences), precipitated in ethanol, and dissolved in 10mM Tris-HCl, pH 8.0, 1 mM EDTA. Purified DME or ROS1 (100 pmol) were incubated at 30 °C for 1 hours with methylated plasmid (1,200 c.p.m) in a reaction mixture containing 50 mM Tris-HCl at pH 8.0, 1 mM EDTA, 1 mM DTT, 0.1 mg/mL BSA, in a total volume of 100 μL. After precipitating in the presence of glycogen, the ethanol-soluble material was concentrated by centrifugation under vacuum and dissolved in 10 μL of a saturated equimolar solution of 5-methylcytosine and thymine in methanol. The mixture was spotted on a silica gel TLC plate (Silicagel 60 F254, 20 x 20 cm, Merck). The first dimension was run in chloroform/methanol/water (4:2:1, v/v/v). The plate was dried in air and developed in the second dimension with ethylacetate:propanol:water (75:16:9, v/v/v). Spots corresponding to 5-methylcytosine and thymine were visualized under UV light, scrapped off, and counted in 5 mL of scintillation fluid.

Cross-link reaction with $NaBH_4$. Cross-link reactions between enzymes and substrates were performed by incubating 25 pmol of protein with double-stranded oligodeoxynucleotides (50 fmol) at 30 °C for 1 hour in a reaction mixture containing 50 mM Tris-HCl at pH 8.0, 1 mM EDTA, 1 mM DTT, 0.1 mg/mL BSA and 50 mM $NaBH_4$, in a total volume of 10 μL. After incubation, the samples were mixed with SDS-PAGE loading buffer, heated at 100ºC for 5 min and electrophoresed on a 12% SDS-polyacrylamide gel. The reaction products were visualized by autoradiography.

Note Added in Proof

While this work was under review, an article based on independent research reporting similar results for DEMETER was published (47).

Acknowledgements

We thank Emilio Fernández for generous support with radiation safety facilities. M.I.P.M. is the recipient of a Ph.D. Fellowship from Junta de Andalucía. This research was supported by grant BFU2004-05303 from Ministerio de Educación y Ciencia to T.R.A.

Author contributions

T.M.R., R.R.A and T.R.A designed research; T.M.R., A.P.O.G., M.I.P.M. and M.I.M.M. performed research; and R.R.A and T.R.A wrote the paper.

References

1. Bender, J. (2004) *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **55,** 41-68.
2. Bird, A. (2002) *Genes Dev.* **16,** 6-21.
3. Colot, V. & Rossignol, J. L. (1999) *Bioessays* **21,** 402-411.
4. Holliday, R. & Pugh, J. E. (1975) *Science* **187,** 226-232.

5. Riggs, A. D. (1975) *Cytogenet. Cell Genet.* **14,** 9-25.

6. Yoder, J. A., Walsh, C. P. & Bestor, T. H. (1997) *Trends Genet.* **13,** 335-340.

7. Finnegan, E. J., Genger, R. K., Peacock, W. J. & Dennis, E. S. (1998) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49,** 223-247.

8. Kress, C., Thomassin, H. & Grange, T. (2001) *FEBS Lett.* **494,** 135-140.

9. Reik, W., Dean, W. & Walter, J. (2001) *Science* **293,** 1089-1093.

10. Frank, D., Keshet, I., Shani, M., Levine, A., Razin, A. & Cedar, H. (1991) *Nature* **351,** 239-241.

11. Bhattacharya, S. K., Ramchandani, S., Cervoni, N. & Szyf, M. (1999) *Nature* **397,** 579-583.

12. Smith, S. S. (2000) *J. Mol. Biol.* **302,** 1-7.

13. Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D. & Bird, A. (1999) *Nat. Genet.* **23,** 58-61.

14. Wade, P. A., Gegonne, A., Jones, P. L., Ballestar, E., Aubry, F. & Wolffe, A. P. (1999) *Nat. Genet.* **23,** 62-66.

15. Jost, J. P., Siegmann, M., Sun, L. & Leung, R. (1995) *J. Biol. Chem.* **270,** 9734-9739.

16. Zhu, B., Zheng, Y., Hess, D., Angliker, H., Schwarz, S., Siegmann, M., Thiry, S. & Jost, J. P. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 5135-5139.

17. Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J. J., Wiebauer, K. & Jiricny, J. (1996) *J. Biol. Chem.* **271,** 12767-12774.

18. Zhu, B., Zheng, Y., Angliker, H., Schwarz, S., Thiry, S., Siegmann, M. & Jost, J. P. (2000) *Nucleic Acids Res.* **28,** 4157-4165.

19. Sibghat, U., Gallinari, P., Xu, Y. Z., Goodman, M. F., Bloom, L. B., Jiricny, J. & Day, R. S., 3rd (1996) *Biochemistry* **35,** 12926-12932.

20. Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J. & Bird, A. (1999) *Nature* **401,** 301-304.

21. Hardeland, U., Bentele, M., Jiricny, J. & Schar, P. (2003) *Nucleic Acids Res.* **31,** 2261-2271.

22. Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. (2004) *J. Biol. Chem.* **279,** 52353-52360.

23. Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., Jacobsen, S. E. & Fischer, R. L. (2002) *Cell* **110,** 33-42.

24. Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S. E., Fischer, R. L. & Kakutani, T. (2004) *Science* **303,** 521-523.

25. Xiao, W., Gehring, M., Choi, Y., Margossian, L., Pu, H., Harada, J. J., Goldberg, R. B., Pennell, R. I. & Fischer, R. L. (2003) *Dev. Cell* **5,** 891-901.

26. Gong, Z., Morales-Ruiz, T., Ariza, R. R., Roldan-Arjona, T., David, L. & Zhu, J. K. (2002) *Cell* **111,** 803-814.

27. Nash, H. M., Bruner, S. D., Scharer, O. D., Kawate, T., Addona, T. A., Spooner, E., Lane, W. S. & Verdine, G. L. (1996) *Curr. Biol.* **6,** 968-980.

28. Choi, Y., Harada, J. J., Goldberg, R. B. & Fischer, R. L. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 7481-7486.

29. Lindahl, T. (1993) *Nature* **362,** 709-715.

30. McCullough, A. K., Dodson, M. L. & Lloyd, R. S. (1999) *Annu. Rev. Biochem.* **68,** 255-285.

31. Garcia-Ortiz, M. V., Ariza, R. R. & Roldan-Arjona, T. (2001) *Plant Mol. Biol.* **47,** 795-804.

32. Boiteux, S., O'Connor, T. R. & Laval, J. (1987) *EMBO J.* **6,** 3177-3183.

33. Bailly, V. & Verly, W. G. (1987) *Biochem. J.* **242,** 565-572.

34. Watanabe, S., Ichimura, T., Fujita, N., Tsuruzoe, S., Ohki, I., Shirakawa, M., Kawasuji, M. & Nakao, M. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 12859-12864.

35. Likhite, V. S., Cass, E. I., Anderson, S. D., Yates, J. R. & Nardulli, A. M. (2004) *J. Biol. Chem.* **279,** 16875-16882.

36. Um, S., Harbers, M., Benecke, A., Pierrat, B., Losson, R. & Chambon, P. (1998) *J. Biol. Chem.* **273,** 20728-20736.

37. Chen, D., Lucey, M. J., Phoenix, F., Lopez-Garcia, J., Hart, S. M., Losson, R., Buluwela, L., Coombes, R. C., Chambon, P., Schar, P. & Ali, S. (2003) *J. Biol. Chem.* **278,** 38586-38592.

38. Missero, C., Pirro, M. T., Simeone, S., Pischetola, M. & Di Lauro, R. (2001) *J. Biol. Chem.* **276,** 33569-33575.

39. Tini, M., Benecke, A., Um, S. J., Torchia, J., Evans, R. M. & Chambon, P. (2002) *Mol. Cell* **9,** 265-277.

40. Kondo, E., Gu, Z., Horii, A. & Fukushige, S. (2005) *Mol. Cell. Biol.* **25,** 4388-4396.

41. Jost, J. P., Oakeley, E. J., Zhu, B., Benjamin, D., Thiry, S., Siegmann, M. & Jost, Y. C. (2001) *Nucleic Acids Res.* **29,** 4452-4461.

42. Zhu, B., Benjamin, D., Zheng, Y., Angliker, H., Thiry, S., Siegmann, M. & Jost, J. P. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 5031-5036.

43. Barnes, D. E. & Lindahl, T. (2004) *Annu. Rev. Genet.* **38,** 445-476.

44. Aravind, L. & Koonin, E. V. (2000) *Genome Biol.* **1,** research0007.0001-0007.0008.

45. Szyf, M. (2005) *Biochemistry (Mosc)* **70,** 533-549.

46. Kieber, J. J., Rothenberg, M., Roman, G., Feldmann, K. A. & Ecker, J. R. (1993) *Cell* **72,** 427-441.

47. Gehring, M., Huh, J. H., Hsieh, T. F., Penterman, J., Choi, Y., Harada, J. J., Goldberg, R. B. & Fischer, R. L. (2006) *Cell* **124,** 495-506.

Figure legends

Fig.1. DME and ROS1 contain a DNA glycosylase domain and are closely related. *(A)* diagram of DME and ROS1 showing the conserved domains as colored sections. Percentage of identical residues between both proteins is shown above each region. *(B)* Amino-acid sequence alignment of DME (amino acids 1519-1659) and ROS1 (amino acids 928-1068) with *A. thaliana* Nth1, *E. coli* Nth, and *H. sapiens* MutY and Ogg1. Accession numbers are ABC61677, AAP37178, CAC16135, P20625, Q9UIF7, and O15527, respectively. Asterisk marks the lysine residue that is diagnostic of a glycosylase/lyase activity; triangle indicates the conserved aspartic acid residue in the active site; diamonds label the cysteine residues that in *E. coli* Nth1 ligate a [4Fe-4S] cluster.

Fig. 2. DME and ROS1 are 5-meC DNA glycosylases. *(A)* Double-stranded oligonucleotide substrates containing 5-meC (M) in CpG (CXGG) and non-CpG (XCGG, XAG and AXT) sequence contexts, or the corresponding unmethylated controls were incubated with purified DME (upper panels) or ROS1 (lower panels) as described in Methods. Reaction products were separated in a 12 % denaturing polyacrylamide gel and visualized by autoradiography. *(B)* Release of 5-meC from methylated DNA. Plasmid DNA methylated with S-adenosyl-L-[methyl-^{3}H]methionine (1,200 c.p.m) was incubated with purified DME or ROS1 and the released ethanol-soluble material was analyzed by two-dimensional TLC (see Methods). Distribution of radioactivity between thymine (closed bars) and 5-methylcytosine (open bars) was assayed by scintillation counting. The mean of duplicate experiments and their standard errors are shown.

Fig. 3. DME and ROS1 excise 5-meC and thymine, but not uracil. *(A)* Double-stranded oligonucleotide substrates containing 5-mec (M), T·G or U·G mismatches in CpG (CXGG) and non-CpG (XCGG, XAG and AXT) sequence contexts were incubated with purified DME (upper panels) or ROS1 (lower panels) as described in Methods. Reaction products were separated in a 12 % denaturing polyacrylamide gel and visualized by autoradiography. *(B)* Double-stranded oligonucleotide substrates containing 5-meC (M), T·G or U·G mismatches in a CpG sequence context were incubated for 24 h with purified wild-type DME or mutant DME D1562A (upper panel) or wild-type ROS1 or mutant ROS1 D971A (lower panel). Reaction products were separated in a 12 % denaturing polyacrylamide gel and visualized by autoradiography.

Fig. 4. Kinetics of DME and ROS1 action on 5-meC and thymine in CpG and non-CpG sequence contexts. The time-dependent generation of incised oligonucleotides was measured by incubating purified DME or ROS1 with double stranded oligonucleotide substrates containing a 5-meC·G pair (closed symbols) or a T·G mismatch (open symbols) in CpG (CXGG) and non-CpG (XCGG, XAG and AXT) sequence contexts. Reactions were stopped at the indicated times, products were separated in a 12 % denaturing polyacrylamide gel and the relative amount of incised oligonucleotide was quantitated by phosphor imaging.

Fig 5. DME and ROS1 process 5-meC by a DNA glycosylase/lyase mechanism proceeding through a transient imine intermediate. Double stranded oligonucleotide substrates containing a C·G, 5-meC·G or 8-oxoG·C pair in a CpG context were incubated with purified DME (lanes 1, 4 and 7), ROS1 (lanes 2, 5 and 8), or AtOGG1 (lanes 3, 6, and 9) in the presence of $NaBH_4$. After incubation, the reaction mixture was subjected to SDS-PAGE and products were visualized by autoradiography. Free

substrates and cross-linked enzyme-substrate complexes are indicated by arrows. M= 5-meC; O= 8-oxoG.

Fig 6. Products formed by DME upon excision of 5-meC. *(A)* Double stranded oligonucleotide substrates containing a 5-meC·G (lanes 5-8) or a 8-oxoG·C pair (lanes 3 and 4) in a CpG context were incubated with purified Fpg (lane 3), AtOGG1 (lane 4) or DME (lanes 5-8) and reaction mixtures were separated in a denaturing polyacrylamide sequencing gel (40 x 20 cm). The products formed by DME were further treated with Endo IV (lanes 6 and 8) to analyze the nature of 3´-termini. Substrates and enzymes used are indicated at the top of the gel. Oligonucleotide markers of 28 and 29 nucleotides were loaded in lanes 1 and 2, respectively. The β and δ-elimination products and those carrying 3´-OH termini are indicated by arrows. *(B)* a double-stranded oligonucleotide substrate containing a 5-meC·G pair in a CpG sequence context was incubated with purified DME and reactions were stopped at different times. Lanes 1- 6 correspond to reaction times of 0.25, 0.5, 1, 2, 10 and 24 h, respectively. Products were separated in a 12 % denaturing polyacrylamide gel and visualized by autoradiography.

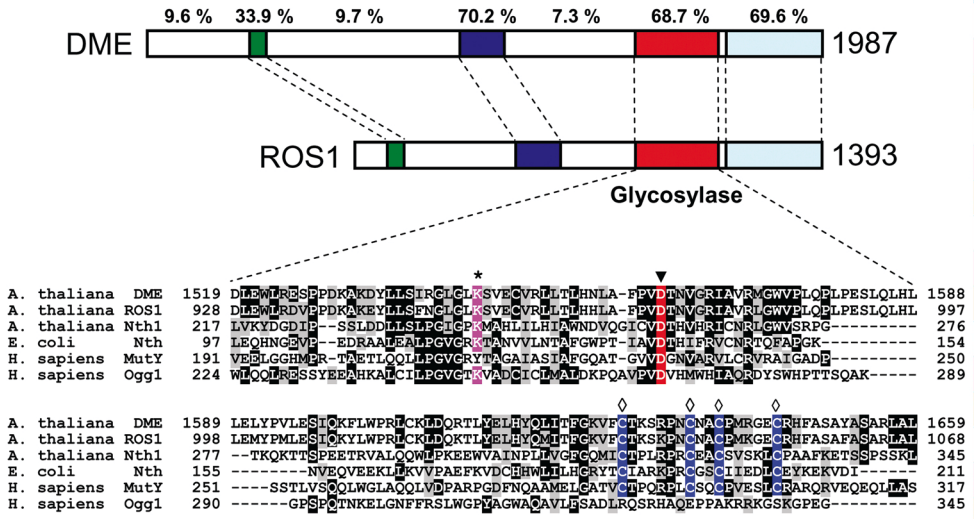

9.6 %
33.9 %
9.7 %
70.2 %
7.3 %
68.7 %
69.6 %
DME
1987
ROS1
1393
Glycosylase
A. thaliana DME 1519 DLEWLRESPPDKAKDYLLSIRGLCLKSVECVRLLTLHNLA-FPVDTNVCRIAVRMGWVPLQPLPESLQLHL 1588
A. thaliana ROS1 928 DLEWLRDVPPDKAKEYLLSFNGLCLKSVECVRLLTLHHLA-FPVDTNVCRIAVRLGWVPLQPLPESLQLHL 997
A. thaliana Nth1 217 LVKYDGDIP--SSLDDLLSLPGICPKMAHLILHIAWNDVQGICVDTHVHRICNRLGWVSRPG--------- 276
E. coli Nth 97 LEQHNGEVP--EDRAAFEALPGVCRKTANVVLNTAFGWPT-IAVDTHIFRVCNRTQFAEGK---------- 154
H. sapiens MutY 191 VEELGGHMER-TAETLQQLLPGVGRYTAGAIASIAFGQAT-GVVDGNVARVLCRVRAIGADP--------- 250
H. sapiens Ogg1 224 WLQQLRESSYEEAHKALCILPGVGTKVADCICLMALDKPQAVPVDVHMWHIAQRDYSWHPTTSQAK----- 289
A. thaliana DME 1589 LELYPVLESICKFLWPRLCKLDQRTLYELHYQLITFGKVFCTKSRPNCNACPMRGECRHFASAYASARLAL 1659
A. thaliana ROS1 998 LEMYPMLESICKYLWPRLCKLDQKTLYELHYQMITFGKVFCTKSKPNCNACPMKGECRHFASAFASARLAL 1068
A. thaliana Nth1 277 --TKQKTTSPEETRVALQQWLPKEEWVAINPLHVGFGQMICTPLRPRCPACSVSKLCPAAFKETSSPSSKL 345
E. coli Nth 155 --------NVEQVEEKLLKVVPAEFKVDCHHWLILHGRYTCIARKPRCGSCIIEDLCEYKEKVDI------- 211
H. sapiens MutY 251 ----SSTLVSQQLWGLAQQLVDPARPGDFNQAAMELGATVCTPQRPLCSQCPVESLCRARQRVEQEQLLAS 317
H. sapiens Ogg1 290 ------GPSPCTNKELGNFFRSLWGPYAGWAQAVLFSADLRQSRHAQEPPAKRRKGSKGPEG--------- 345

Figure 1

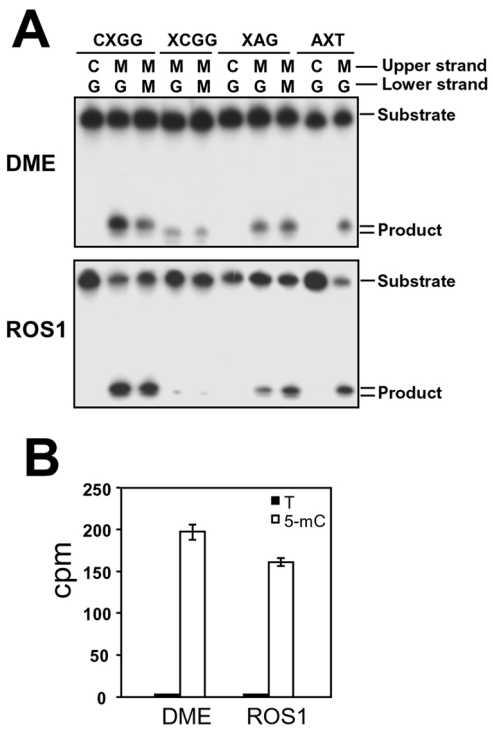

A
CXGG
XCGG
XAG
AXT
C M M M M C M M C M
G G M G M G G M G G
Upper strand
Lower strand
DME
Substrate
Product
ROS1
Substrate
Product
B
250
200
150
100
50
0
cpm
T
5-mC
DME
ROS1

Figure 2

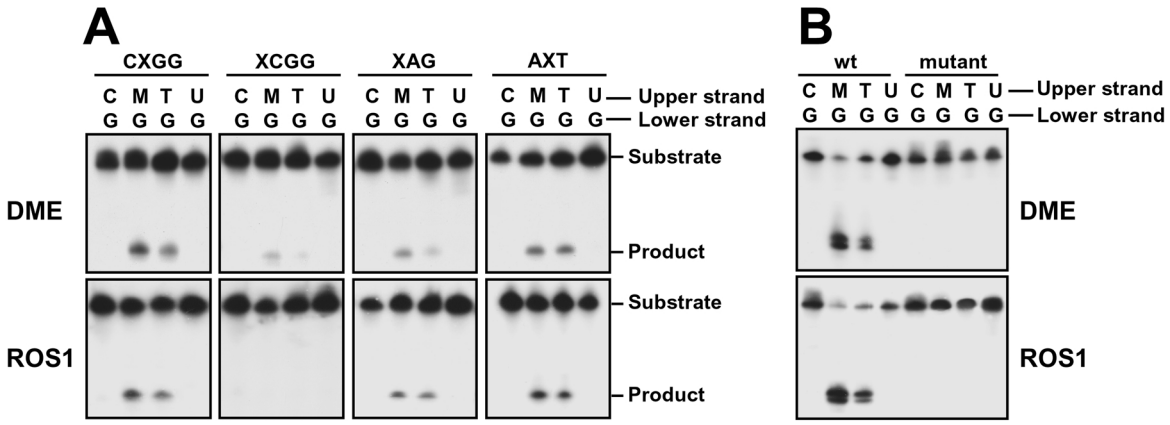

A
CXGG
XCGG
XAG
AXT
C M T U
G G G G
Upper strand
Lower strand
Substrate
Product
DME
ROS1
B
wt
mutant

Figure 3

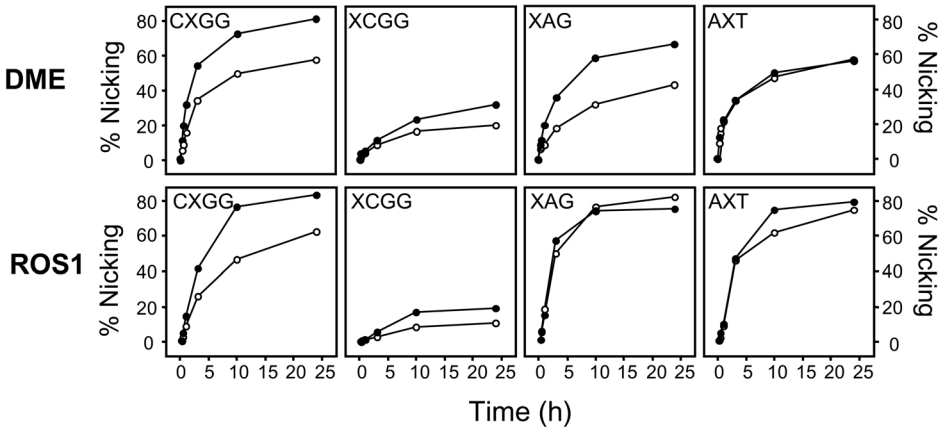

DME
ROS1
CXGG
XCGG
XAG
AXT
% Nicking
Time (h)
0
20
40
60
80
0 5 10 15 20 25

Figure 4

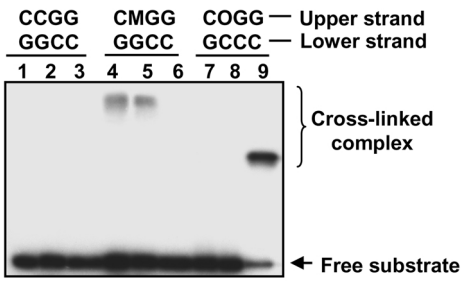
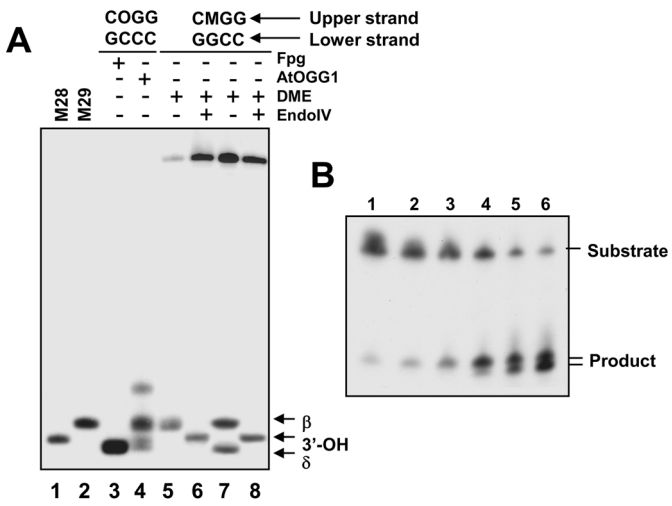

CCGG
GGCC
CMGG
GGCC
COGG
GCCC
Upper strand
Lower strand
1 2 3 4 5 6 7 8 9
Cross-linked complex
Free substrate

Figure 5


A
COGG
GCCC
CMGG
GGCC
Upper strand
Lower strand
Fpg
AtOGG1
DME
EndoIV
M28
M29
β
3'-OH
δ
1 2 3 4 5 6 7 8
B
1 2 3 4 5 6
Substrate
Product