Plant Biotechnology 24, 339–344 (2007)

Short Communication

Identification of tobacco genes encoding proteins possessing removal activity of 5-methylcytosines from intact tobacco DNA

Chang-Sung Choi, Hiroshi Sano*

Research and Education Center for Genetic Information, Nara Institute of Science and Technology, Nara 630-0192, Japan
* E-mail: sano@gtc.naist.jp Fax: +81-743-72-5659

Received January 12, 2007; accepted February 27, 2007 (Edited by K. Yoshida)

Abstract Cytosines in eukaryotic DNA is often methylated to yield 5-methylcytosines (m^5C), which play an important role in controlling gene expression. This is referred as DNA methylation, and its status dynamically changes during plant growth by active methylation and demethylation. DNA glycosylases have been known to possess base excision DNA repair activity, and an Arabidopsis enzyme, ROS1 (Repressor of Silencing 1), was recently shown to exhibit m^5C DNA glycosylase activity. We isolated and characterized four genes encoding ROS1-like proteins from tobacco plants (*Nicotiana tabacum*), and designated as *NtROS1*, *NtROS2a*, *NtROS2b* and *NtROS3*, each respectively encoding a polypeptide with 1796, 1673, 1673 and 1662 amino acids. Purified NtROS1 and NtROS2a proteins expressed in Sf9 insect cells clearly exhibited activity of m^5C removal from tobacco genomic DNA *in vitro*. GFP fusion assay showed that NtROS1 and NtROS2a were localized in nucleus. Transcripts of *NtROS1*, *NtROS2a* and *NtROS3* were induced by abiotic stresses, including aluminium, salt and reactive oxygen species. These observations suggested that NtROS proteins function in demethylating process of genomic DNA during plant stress response, thereby maintaining the balance of gene expression in combination with DNA methyltransferases.

Key words: DNA glycosylase, DNA methylation, 5-methylcytosine, *Nicotiana tabcum*, Repressor of Silencing1.

Eukaryotic DNA contains 5-methylcytosine (m^5C) as a minor base, and it is referred as DNA methylation. DNA methylation has been shown to be involved in gene silencing at both transcriptional and posttranscriptional levels. Transcriptional gene silencing is associated with hypermethylation of promoter sequences, while post-transcriptional gene silencing is linked with hypermethylation of transcribed or coding sequences (Paszkowski and Whitham 2001). DNA methylation is also involved in epigenetics, defined as mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in the DNA sequence (Bird 2002). This typically occurs during somatic cell differentiation in animal cells (Reik et al. 2001), and in plants and filamentous fungi as well (Kakutani et al. 1996; Martienssen and Colot 2001).

Methylation status dynamically changes depending upon cell type, developmental stages and even environmental conditions (Demeulemeester et al. 1999; Steward et al. 2002; Chakrabarty et al. 2003; Wada et al. 2004). Methylation is a post-replication event, catalyzed by DNA methyltransferases which have been relatively well characterized (Bender 2004). In contrast, demethylation process has largely remained unclear. One of such mechanisms was recently shown to be catalyzed by DNA glycosylases, which excise m^5C from intact DNA and replace it with cytosines (Kapoor et al. 2005). DNA glycosylases recognize and remove a particular subset of base alterations, including demethylation, oxidation, deamination, and alkylation (Lindahl and Wood 1999; McCulloughet al.1999; Sedgwick 2004). A DNA glycosylase specific to m^5C was first identified in chicken embryos (Jost et al. 1995), and an active demethylation pathway initiated by thymine DNA glycosylase (TDG) and/or methyl-CpG binding domain (MBD) protein 4 (MBD4) has been proposed in animal cell (Zhu et al. 2000; Jost et al. 2001; Zhu et al. 2001). Recently, two such DNA glycosylases were identified from Arabidopsis, designated as Repressor of Silencing 1 (ROS1) and DEMETER (DME) (Gong et al. 2002; Choi et al. 2002). Both enzymes remove m^5C at high efficiency through bifunctional DNA glycosylase/lyase mechanism *in vitro* (Gehring et al. 2006; Agius et al. 2006; Morales-Ruiz et al. 2006). They also showed significantly high incision activity in symmetric dinucleotide CpG context, but low activity for G:T mismatch repair. ROS1 was not active against damaged DNA (Agius et al. 2006; Morales-Ruiz et al. 2006). Despite the apparently important function, studies on ROS1/DME from other plants species have so far been

This article can be found at http://www.jspcmb.jp/

limited. In this report, we describe isolation and characterization of four ROS1-like proteins from tobacco plants. Wild-type tobacco plants (*Nicotiana tabacum* cv. Xanthi nc) were grown in a greenhouse at 23°C under a 14/10 h light/dark photo cycle for 1 to 2 months. Stress treatments were given as described (Taura et al. 2003). Briefly, healthy leaves were detached from wild type mature tobacco plants, and subjected to stress treatments with aluminium chloride (100 μM), sodium chloride (300 mM) or paraquat (100 μM). RNA blot analysis was performed using total RNA extracted from samples by the AGPC method (Nakano et al. 2000). Approximately 10 μg of RNA were separated by electrophoresis, transferred to nylon membrane, and subjected to hybridization with radioactively labeled probes as described (Yoda et al. 2002). A cDNA library was constructed from 5 μg of mRNA using the ZAP cDNA synthesis kit (Stratagene, La Jolla, CA) and Uni-ZAP XR vector, which was packaged *in vitro* with Gigapack III Gold packaging extract (Stratagene). The amplified library contained approximately 2.9×10^6 recombinant plaques, and 5×10^4 were plated and transferred to nylon membrane (Hybond N^-, Amersham, Buckinghamshire, UK). Membranes were screened by differential hybridization using ^{32}P-labeled probes. Tobacco cDNA fragments encoding the m^5C DNA glycosylase domain were obtained by reverse transcriptase (RT)-PCR. PCR primers were designed based on the sequence of plant DNA glycosylase genes in GenBank (accession AAM77215 and AAP37178): DNG1 forward, 5′-CTTACACAGATTGTCTCGGA-3′; DNG1 reverse, 5′-GCTGTTTCACCTGGTGTCCATATA-3′; DNG2 forward, 5′-CTTACACAGATTGTCTCGGA-3′; DNG2 reverse, 5′-GCTGTTTCACCTGGGGTCCATATA-3′. The size of amplified fragments was 2.2 kb, which encoded the m^5C DNA glycosylase domain found in AtROS1 and AtDME. These fragments were used to screen the cDNA library as the probe to obtain the full-length cDNAs. Proteins were expressed *in vitro* using S*podoptera frugiperda* (Sf9) insect cell line as described previously (Wada et al. 2003). Cells from one culture dish were suspended in 1 ml of lysis buffer (100 mM Hepes, pH 8.0, 240 mM KCl, 1% glycerol, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride), and sonicated twice for 10 s each. GST-fused NtROS1 and NtROS2a were purified through a glutathione-Sepharose column (Amersham). Approximately 10 μg of genomic DNA were incubated at 37°C for 2 h in a reaction mixture containing 10 mM Bis-Tris propane-HCl (pH 7.0), 10 mM $MgCl_2$, 1 mM dithiothreitol, 0.1 mg mL^{-1} BSA, 1 mM dNTP, 50 pmol of protein in a total volume of 100 μl. Reactions were stopped by adding EDTA to 20 mM, and SDS to 0.6%, and the mixtures were incubated for 10 min at 70°C. DNA was extracted with phenol/chloroform/isoamyl alcohol and ethanol-precipitated. The methylation status was examined by digesting 0.5 μg DNA with *Hpa*II or *Msp*I restriction endonucleases, followed by agarose gel electrophoresis (Gong et al. 2002).

GFP fusions were constructed as described (Yap et al. 2005). The entire coding region of *NtROS1* and *NtROS2a* were amplified by PCR. The attB-PCR products were cloned into pGWB6(35S-sGFP-NOS) vector to fuse to green fluorescence protein (GFP) using the GATEWAY cloning system (Invtrogen, Carlsbad, CA). As for the control, a plasmid expressing only GFP was used. Each plasmid was introduced into onion epidermis cell by biolistic bombardment, and observed for localization of GFP with a confocal microscope.

After intensive screening of the cDNA library, 13 positive plaques were initially obtained, among which 7 clones were finally selected and the sequence determined. They were classified into four polymorphic types, consisted of 5940, 5879, 5878 and 5651 bp fragments, and designated as Nicotiana tabacum Repressor of Silencing 1 (*NtROS1*), *NtROS2a*, *NtROS2b*, and *NtROS3*, respectively (Figure 1A) (accession numbers; AB281587 for *NtROS1*, AB281588 for *NtROS2a*, AB281589 for *NtROS2b*, and AB281590 for *NtROS3*). Each clone encoded a protein of 1796, 1673, 1673, and 1662 amino acids with predicted molecular mass of 201, 186, 186, and 185 kDa, respectively. A pair-wise comparison of their deduced amino acid sequences with AtROS1 showed the similarity of NtROS1 at 38.9%, NtROS2a at 32.2%, NtROS2b at 32.1% and NtROS3 at 31.3%. However, the similarity at the glycosylase domain with AtROS1 was 94.3% for NtROS1 (amino acid positions, 1291–1459), 94.3% for NtROS2a (amino acid positions, 1291–1459), 94.3% for NtROS2b (amino acid positions, 1291–1459) and 90.1% for NtROS3 (amino acid positions, 1305–1433). A pair-wise comparison with NtROS1 indicated the similarity of NtROS2a, NtROS2b and NtROS3 to be 51.1%, 51.2% and 49.3%, respectively. Notably, only 5 amino acids out of total 1673 amino acids (0.4%) differed between NtROS2a and NtROS2b, suggesting them to be isoforms. Conserved lysine and aspartic acid residues located in the helix-hairpin-helix domain in DNA glycosylases (Krokan et al. 1997) and four cysteine residues adjacent to the DNA glycosylase domain were also present in NtROS proteins. The latter cluster is thought to play a role in DNA binding. NtROS proteins also possess nuclear localization signal and histone deacetylase interacting domains (http://smart.embl-heidelberg.de/) (Figure 1A), suggesting them to localize in nucleus and to interact with DNA or chromatin (Gehring et al. 2006).

Full-length NtROS1 and NtROS2a as glutathione S-transferase (GST) fusion proteins were expressed in baculovirus-mediated expression system in *Spodoptera*

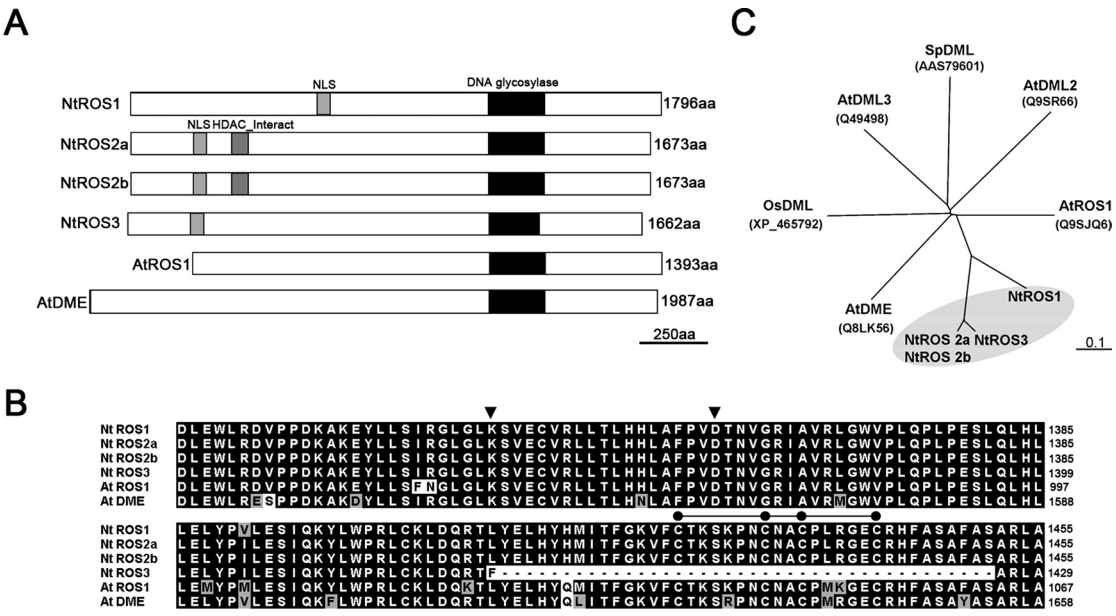

Figure 1. Properties of NtROS proteins. (A) Structural illustration of m^5C DNA glycosylase proteins. Protein size is shown in number of amino acids (aa), and conserved DNA glycosidase domain is indicated by closed box, and specific regions including histone deacetylase interacting (HDAC interact), nuclear localization signal (NLS) are indicated by shaded boxes. (B) Alignment of conserved m^5C DNA glycosylase domain of *NtROS1*, *NtROS2a*, *NtROS2b*, *NtROS3*, *AtROS1* and *AtDME*. Common amino acids among five proteins are indicated by closed box, and grey boxes indicate conservative changes. Conserved lysine and aspartic acid residues in the glycosylase/lyase active site (InterPRO search, IPRO11257) are indicated by closed arrowheads, and four cysteine residues predicted to function in DNA binding are indicated by closed circles. (C) Unrooted phylogenetic tree. Amino acid sequences were aligned using the ClustalW program (http://www.ddbj.nig.ac.jp/search/clustalw-j.html). Genebank accession numbers are: *NtROS1*, AB281587; *NtROS2a*; AB281588, *NtROS2b*, AB281589; *NtROS3*, AB281590; *AtROS1*, Q9SJQ6; *AtDME*, Q8LK56; *AtDML2*, Q9SR66; *AtDML3*, Q49498; *OsDML*, XP_475792 (*O. sativa*); *SpDML*, AAS79601 (*Ipomea trifida*).

frugiperda (Sf9) insect cell line. After 72 h post-infection, cell extracts were sampled, and protein was purified through a glutathione-Sepharose column and analyzed by SDS-PAGE, showing that full-length (ca. 300 kDa) proteins were successfully synthesized (Figure 2A). *In vitro* demethylation activity was then examined. Tobacco genomic DNA was incubated with or without purified NtROS1 or NtROS2a, purified with phenol extraction and ethanol precipitation, digested with either *Msp*I or *Hpa*II, and fractionated by agarose gel electrophoresis (Figure 2B). The untreated control sample was highly resistant to *Hpa*II but susceptible to *Msp*I, indicating hypermethylation of the second cytosine at CCGG sites. In contrast, samples treated with NtROS1 or NtROS2a were digested by *Hpa*II to the same extent as those by *Msp*I (Figure 2B). Since ROS-treated, but *Hpa*II/*Msp*I-untreated samples were not degraded at all, these results clearly indicated an efficient removal of m^5C in CCGG sites.

Cellular localization of NtROS1 and NtROS2a was examined using fusion proteins with a green fluorescence protein (GFP) reporter under the control of the cauliflower mosaic virus (CaMV) 35*S* promoter. Results showed that both proteins were exclusively localized in nucleus (Figure 3). This is consistent with a computational prediction by the Predotar (version 1.03 at http://urgi.infobiogen.fr/predotar/predotar.html), PSORT (Nakai and Kanehisa, 1992), and iPSORT (Bannai et al. 2002) programs, which suggested their nuclear localization.

Their expression profile was then examined. Transcripts for *NtROS1* and *NtROS2a*/*2b* in young leaves were scarcely detectable under non-stressed conditions (Figure 4). When subjected to aluminium stress, transcripts for *NtROS1* and *NtROS2a*/*2b* began to accumulate 24 h and 6 h later, respectively. Salt stress also induced their transcripts after 1 h, reaching maximal levels 3 h later and then gradually declining to the initial level. Similarly, paraquat treatment induced their transcripts within 2 h. These observations indicated that *NtROS1* and *NtROS2a*/*2b* specifically and temporarily respond to abiotic environmental stresses, and that such response might be mediated through reactive oxygen species. Transcript accumulation profile of *NtROS3* was essentially similar with that of *NtROS1*, except that the amount appeared to be lower (Figure 4).

DNA glycosylases are subdivided into four groups according to their substrate specificity: excision of uracil, uracil-containing mismatches, alkylated bases and oxidized bases (Fortini et al. 2003). NtROS proteins are structurally related to Arabidopsis ROS1 and DME, which belong to a large family of uracil-containing mismatch DNA glycosylases. However, apart from the m^5C DNA glycosylase domain and regions at the

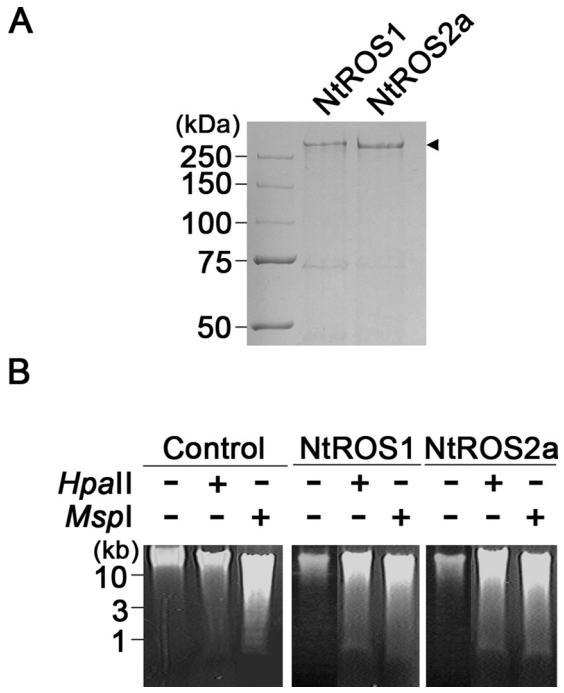

Figure 2. Enzymatic activity. (A) Expression of NtROS1 and NtROS2a. Corresponding full length cDNA fused to *GST* was expressed in Sf9 cells, and resulting proteins were purified through glutathione-Sepharose column. An appropriated aliquot was fractionated by SDS-PAGE (10%), and stained with Coomassie Brilliant Blue (CBB). The position of marker proteins is indicated at left. Closed arrowhead indicates expressed proteins with 300 kDa. (B) Removal of m^5C. Tobacco genomic DNA was treated with purified NtROS1 (indicated as NtROS1) or NtROS2a (NtROS2a) at 37°C for 2 h, extracted by phenol/chloroform, subjected to digestion with either *Hpa*II or *Msp*I as indicated by plus (+) or minus (−), fractionated on 1.2% agarose gel electrophoresis, and stained by ethidium bromide. Untreated DNA was used as the control (Control). The position of molecular marker is indicated at left.

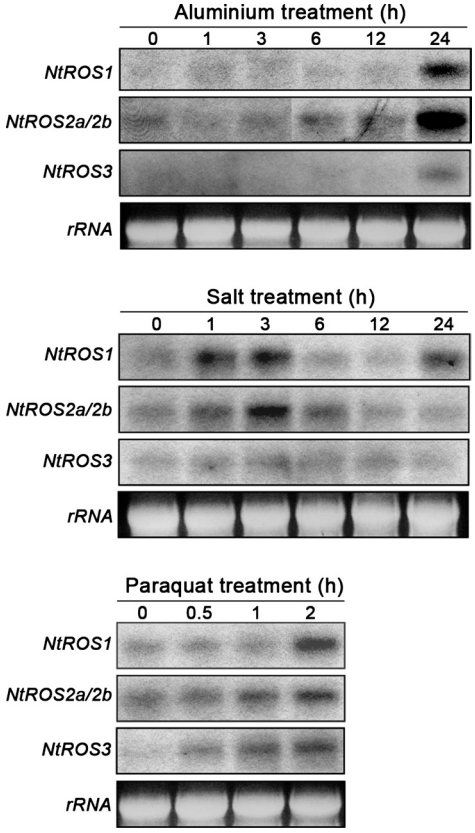

Figure 4. Transcript accumulation of *NtROS* upon stress treatment. Healthy leaves were detached from wild type mature tobacco plants, and petioles were soaked in solution containing aluminium chloride (100 μM), sodium chloride (300 mM) or paraquat (100 μM). After being incubated for indicated time period, total RNA was isolated and a 10-μg aliquot was fractionated on an agarose gel, and subjected to RNA blot hybridization with radioactively labeled probes indicated. Equal loading of RNA samples was confirmed by rRNA staining with ethidium bromide. Note that the probe used for *NtROS2a* also recognized *NtROS2b* due to their high similarity even in the 5′ UTR. Resulting RT-PCR thus indicates levels of both transcripts and indicated by *NtROS2a/2b*.

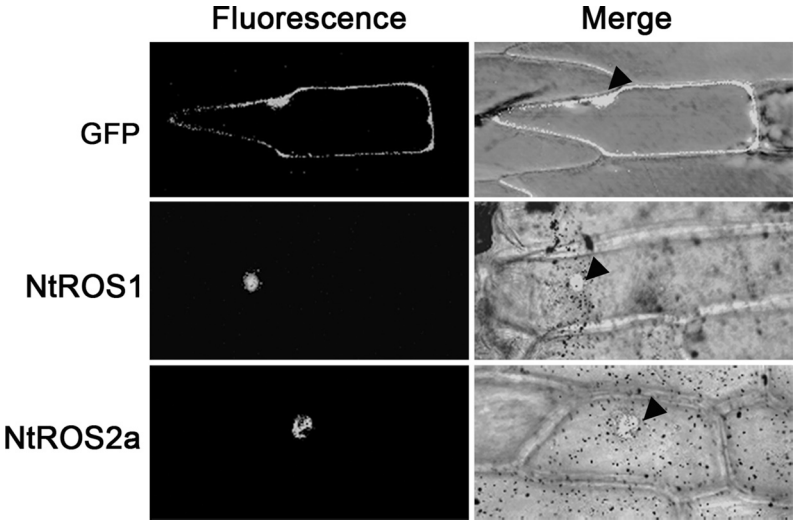

Figure 3. Cellular localization. Onion epidermal cell layers were bombarded with gold particles coated with 35S*NtROS1-GFP*, 35S*NtROS2a-GFP* or 35S*GFP* alone, and observed by epifluorescence for GFP. Bright field mages are merged to show positions of nucleus indicated by arrowheads.

carboxyl-terminus, their similarity was low, suggesting specific function of NtROS proteins. Since NtROS1 and NtROS2a were shown to possess m^5C removal activity, and to be localized in nucleus, it is conceivable that they are closely associated with DNA and engaged in DNA demethylation. In this context, their induction profile upon abiotic stresses is indicative, as DNA glycosylases are considered to deal with oxidative damages, which are commonly caused by reactive oxygen species (Dizdaroglu 2003; Fortini et al. 2003). To cope with such damages, a set of genes must be expressed to repair and/or strengthen defense reactions. If these genes are locked by methylation under non-stressed condition, demethylation is absolutely necessary to globally switch on them. It is tempting to speculate that NtROS proteins participate in this task, thereby finely adjusting gene expression through methylation/demethylation pathways.

Acknowledgements

The authors thank Dr. Yuko Wada (Nara Institute of Science and Technology) for valuable advice and discussion. This work was partly supported by a grant from the Research for the Future Program of the Japan Society for the Promotion of Science.

References

Agius F, Kapoor A, Zhu J-K (2006) Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc Natl Acad Sci USA* 103: 11796–11801

Bender J (2004) DNA methylation and epigenetics. *Annu Rev Plant Biol* 55: 41–68

Bird AP (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16: 6–21

Bannai H, Tamada Y, Maruyama O, Nakai K and Miyano S (2002) Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* 18: 298–305

Babiushin BF (2006) DNA methylation in plants. *Curr Top Microbiol Immunol* 301: 67–122

Chakrabarty D, Yu KW, Paek KY (2003) Detection of DNA methylation changes during somatic embryogenesis of Siberian ginseng (*Eleuterococcus senticosus*). *Plant Sci* 165: 61–68

Choi Y, Harada JJ, Goldberg RB, Fischer RL (2004) An invariant aspartic acid in the DNA glycosylase domain of DEMETER is necessary for transcriptional activation of the imprinted MEDEA gene. *Proc Natl Acad Sci USA* 101: 7481–7486

Demeulemeester MAC, Van Stallen N, De Proft MP (1999) Degree of DNA methylation in chicory (Cichorium intybus L.): influence of plant age and vernalisation. *Plant Sci* 142: 101–108

Gong Z, Morales-Ruiz T, Ariza RR, Roldan-Arjona T, David L, Zhu JK (2002) ROS1, a repressor of transcriptional gene silencing in arabidopsis, encodes a DNA glycosylase/lyase. *Cell* 111: 803–814

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi YH, Harada JJ, Goldberg BR, Jost JP, Siegmann M, Sun L, Leung R (1995) Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase. *J Biol Chem* 270: 9734–9739

Jost JP, Oakeley EJ, Zhu B, Benjamin D, Thiry S, Siegmann M, Jost YC (2001) 5-Methylcytosine DNA glycosylase participates in the genome-wide loss of DNA methylation occurring during mouse myoblast differentiation. *Nucleic Acids Res* 29: 4452–4461

Kakutani T, Jeddeloh JA, Flowers SK, Munakata K, Richards EJ (1996) Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proc Natl Acad Sci USA* 93: 12406–12411

Kapoor A, Agius F, Zhu JK (2005) Prevent transcriptional gene silencing by DNA demethylation. *FEBS Lett* 579: 5889–5898

Krokan HE, Standal, R, Slupphaug G (1997) DNA glycosylases in the base excision repair of DNA. *Biochem J* 325: 1–16

Lindahl T, Wood RD (1999) Quality control by DNA repair. *Science* 286: 1897–1905

McCullough A K, Dodson M L, Lloyd RS (1999) Initiation of base excision repair: glycosylase mechanisms and structures. *Annu Rev Biochem* 68: 255–285.

Morales-Ruiz T, Ortega-Galisteo AP, Ponferrada-Marin MI, Martinez-Macias MI, Ariza RR, Roldan-Arjona T (2006) *DEMETER* and *REPRESSOR OF SILENCING 1* encode 5-methylcytosine DNA glycosylases. *Proc Natl Acad Sci USA* 103: 6853–6858

Nakano Y, Steward N, Koizumi N, Sano H (2000) A tobacco NtMET1 encoding a DNA methyltransferase: molecular characterization and abnormal phenotypes of antisence transgenic tobacco plants. *Plant Cell Physiol* 41: 448–457

Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* 14: 897–911

Paszkowski J, Whitham SA (2001) Gene silencing and DNA methylation processes. *Curr Opin Plant Biol* 4: 123–129

Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. *Science* 293: 1089–1093

Sedgwick B (2004) Repairing DNA methylation damage. *Nat Rev Mol Cell Biol* 5: 148–157

Steward N, Ito M, Yamakuchi Y, Koizumi N, Sano H (2002) Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. *J Biol Chem* 277: 37741–37746

Wada Y, Ohya H, Yamakuchi Y, Koizumi N, Sano H (2003) Preferential *de novo* methylation of cytosine residues in non-CpG sequences by a domains rearranged DNA methyltransferase from tobacco plants. *J Biol Chem* 278: 42386–42393

Wada Y, Miyamoto K, Kusano T, Sano H (2004) Association between up-regulation of stress-responsive genes and hypomethylation of genomic DNA in to tobacco plants. *Mol Genet Genom* 271: 658–666

Yap YK, Kodama Y, Waller F, Chung KM, Ueda H, Nakamura K, Oldsen M, Yoda H, Yamakuchi Y, Sano H (2005) Activation of a novel Transcription factor through phosphorylation by WIPK, a wound-induced mitogen-activated protein kinase in tobacco plants. *Plant Physiol* 139: 127–137

Yoda H, Ogawa M, Yamaguchi Y, Koizumi N, Sano H (2002) Identification of early responsive genes associated with the hypersensitive response to tobacco mosaic virus and characterization of a WRKY-type transcription factor in tobacco plants. *Mol Genet Genom* 267: 154–161

Zhu B, Zheng Y, Angliker H, Schwarz S, Thiry S, Siegmann M, Jost, JP (2000) 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucleic Acids Res* 28: 4157–

4165
Zhu B, Benjamin D, Zheng Y, Angliker H, Thiry S, Siegmann M, Jost JP (2001) Overexpression of 5-methylcytosine DNA glycosylase in human embryonic kidney cells EcR293 demethylates the promoter of a hormone-regulated reporter gene. *Proc Natl Acad Sci USA* 98: 5031–5036