Methylation-independent DNA Binding Modulates Specificity of Repressor of Silencing 1 (ROS1) and Facilitates Demethylation in Long Substrates

Received for publication, March 17, 2010, and in revised form, May 7, 2010 Published, JBC Papers in Press, May 19, 2010, DOI 10.1074/jbc.M110.124578

María Isabel Ponferrada-Marín, María Isabel Martínez-Macias, Teresa Morales-Ruiz, Teresa Roldán-Arjona, and Rafael R. Ariza

From the Department of Genetics, University of Córdoba, 14071 Córdoba, Spain

DNA methylation is an epigenetic mark that promotes gene silencing and performs critical roles during reproduction and development in both plants and animals. The genomic distribution of DNA methylation is the dynamic outcome of opposing methylation and demethylation processes. In plants, active demethylation occurs through a base excision repair pathway initiated by 5-methylcytosine (5-meC) DNA glycosylases of the REPRESSOR OF SILENCING 1 (ROS1)/DEMETER (DME) family. To gain insight into the mechanism by which Arabidopsis ROS1 recognizes and excises 5-meC, we have identified those protein regions that are required for efficient DNA binding and catalysis. We have found that a short N-terminal lysine-rich domain conserved in members of the ROS1/DME family mediates strong methylation-independent binding of ROS1 to DNA and is required for efficient activity on 5-meCG, but not for T-G processing. Removal of this domain does not significantly affect 5-meC excision from short molecules, but strongly decreases ROS1 activity on long DNA substrates. This region is not required for product binding and is not involved in the distributive behavior of the enzyme on substrates containing multiple 5-meC residues. Altogether, our results suggest that methylation-independent DNA binding allows ROS1 to perform a highly redundant search for efficient excision of a nondamaged, correctly paired base such as 5-meC in long stretches of DNA. These findings may have implications for understanding the evolution of structure and target specificity in DNA glycosylases.

DNA methylation is catalyzed by DNA methyltransferases that transfer a methyl group from S-adenosyl-L-methionine to the carbon 5 of cytosine to generate 5-methylcytosine (5-meC)³ (1). This epigenetic mark may promote gene silencing directly, by inhibiting binding of transcriptional activators, and indirectly through recruitment of methyl DNA-binding proteins that initiate formation of repressive chromatin (2). In higher eukaryotes, methylation of specific loci allows transmission of stable states of gene expression and plays essential roles in the establishment of developmental programs and in genome defense against parasitic mobile elements (3). Distortion of DNA methylation patterns may have important consequences for cells, and it is a crucial component in many forms of human disease, including cancer (4, 5).

Methylation profiles are the dynamic outcome of both methylation and demethylation processes (6, 7). Demethylation may occur passively due to lack of maintenance methylation during several DNA replication cycles, or as an active process in the absence of replication (6, 7). The nature of the enzymatic mechanism(s) responsible for active DNA demethylation in animal cells remains poorly understood (6, 8). However, in plants there is convincing genetic and biochemical evidence that proteins from a subfamily of DNA glycosylases, typified by Arabidopsis REPRESSOR OF SILENCING 1 (ROS1) and DEMETER (DME), excise 5-meC and initiate DNA demethylation through a base excision repair process (9–13).

Members of the ROS1/DME family are atypical DNA glycosylases containing a helix-hairpin-helix-Gly/Pro-rich loop/Asp (HhH-GPD) superfamily domain embedded within a much larger polypeptide (9–11) (supplemental Fig. S1). Both ROS1 and DME remove 5-meC as a free base from DNA using a glycosylase/abasic site (AP) lyase mechanism (11) and cleave the phosphodiester backbone at the 5-meC removal site by successive $\beta,\delta$-elimination, leaving a gap that has to be further processed to generate a 3'-OH terminus suitable for polymerization and ligation (11, 12). In addition to ROS1 and DME, the genome of Arabidopsis encodes two additional paralogs with 5-meC DNA glycosylase activity, referred to as DEMETER-LIKE proteins DML2 and DML3 (14, 15). The in vivo functions of plant 5-meC DNA glycosylases are not fully understood, but they seem to contribute to the stability and flexibility of the plant epigenome. DME is required for demethylation of several imprinted loci in female gametes (12, 16) and contributes to genome-wide demethylation during endosperm development (17, 18). ROS1, DML2, and DML3 may protect the genome from excess methylation, counteracting hypermethylation at hundreds of discrete regions across the plant genome (14, 15, 19). It is unclear how these 5-meC DNA glycosylases are targeted to specific genomic regions, but there is evidence indicating that demethylation by ROS1 may be guided by small RNAs bound to ROS3 (20).
A notable feature of all DNA repair glycosylases is their ability to single out damaged and/or mismatched bases among a vast excess of nontarget bases. However, 5-meC DNA glycosylases perform the most remarkable task of excising a nondamaged, correctly paired base. It remains unknown how the enzymes of the ROS1/DME family specifically recognize 5-meC in DNA. In addition to 5-meC, ROS1 and DME also excise with less efficiency its deamination product thymine (5-methyluracil) from T-G mispairs, but do not show detectable activity on either C-G pairs or U-G mispairs (11). We have recently reported that ROS1 activity is facilitated at mismatched 5-meC residues, but strongly inhibited by replacement of the C5 methyl group by halogen derivatives, even if these substituents decrease the strength of the scissile C1-N-glycosidic bond (21). These results suggest that ROS1 specificity arises by a combination of selective steric recognition of the methyl group at the active site and the intrinsic stability and dynamics of the target base pair.

We have undertaken the task of identifying critical regions necessary for ROS1 activity and specificity. In this report, we show that a lysine-rich N-terminal domain conserved across the ROS1/DME family allows ROS1 to bind DNA in a methyltransferation-independent manner, selectively favors 5-meC over T-G pairs or U-G pairs mispairs (11). We have recently reported that ROS1 activity is facilitated at mismatched 5-meC residues, but strongly inhibited by replacement of the C5 methyl group by halogen derivatives, even if these substituents decrease the strength of the scissile C1-N-glycosidic bond (21). These results suggest that ROS1 specificity arises by a combination of selective steric recognition of the methyl group at the active site and the intrinsic stability and dynamics of the target base pair.

We have undertaken the task of identifying critical regions necessary for ROS1 activity and specificity. In this report, we show that a lysine-rich N-terminal domain conserved across the ROS1/DME family allows ROS1 to bind DNA in a methyltransferation-independent manner, selectively favors 5-meC over T-G pairs or U-G pairs mispairs (11). We have recently reported that ROS1 activity is facilitated at mismatched 5-meC residues, but strongly inhibited by replacement of the C5 methyl group by halogen derivatives, even if these substituents decrease the strength of the scissile C1-N-glycosidic bond (21). These results suggest that ROS1 specificity arises by a combination of selective steric recognition of the methyl group at the active site and the intrinsic stability and dynamics of the target base pair.
Role of Lysine-rich Domain in DNA Demethylation by ROS1

A

B

FIGURE 1. ROS1 binds with similar affinity to both methylated and nonmethylated DNA. A, EMSA. Increasing concentrations (0, 35, and 70 nM) of ROS1 were incubated with 100 nM labeled nonmethylated (lanes 1–3 and 7–9) or methylated (lanes 4–6 and 10–12) oligonucleotide duplex containing either one (lanes 4–6) or four (lanes 10–12) dimethylated CG sites. B, ROS1 (105 nM) incubated with 100 nM labeled nonmethylated (upper panel) or methylated DNA containing a dimethylated CG site (lower panel), in the presence of increasing amounts (0, 0.1, 0.25, 0.5, 1.0, and 2.5 μM) of either nonmethylated (lanes 2–7) or dimethylated (lanes 8–13) unlabeled competitor DNA. After non-denaturing gel electrophoresis, protein-DNA complexes were identified by their retarded mobility compared with that of free DNA, as indicated. Graphs on the left show the percentage of remaining complex versus competitor molar excess ratios. M indicates 5-meC.

\[
B = \frac{B_{\text{max}}(D^* + K_d^*)}{D^* + K_d^* + D} \quad \text{(Eq. 2)}
\]

where \(K_d^*\) and \(K_d\) are the dissociation constants for the labeled probe and unlabeled competitor DNA, respectively, \(D^*\) and \(D\) are the concentration of labeled probe and competitor DNA, respectively, and \(B_{\text{max}}\) is the maximum concentration of the protein-DNA complex. When the labeled probe and competitor DNA are identical, \(K_d^* = K_d\), and Equation 1 simplifies to

\[
B = \frac{B_{\text{max}}(D^* + K_d^*)}{D^* + K_d^* + D} \quad \text{(Eq. 3)}
\]

\(K_d^*\) values for labeled probes were determined from the fit to Equation 3 and then used in Equation 2 to derive \(K_d\) values for the unlabeled competitors.

Enzyme Activity Assays—Double-stranded oligonucleotides (20 nM, unless otherwise stated) were incubated at 30 °C for the indicated times in a reaction mixture containing 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM dithiothreitol, 0.1 mg/ml bovine serum albumin, and the indicated amounts of full-length ROS1 or its deletion derivatives in a total volume of 50 μl. Reactions were stopped by adding 20 mM EDTA, 0.6% SDS, and 0.5 mg/ml proteinase K, and the mixtures were incubated at 37 °C for 30 min. DNA was extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and ethanol precipitated 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. Fluorescencelabeled DNA was visualized in a FLA-5100 imager and analyzed using Multigauge software.

RESULTS

ROS1 Binds with Similar Affinity to Both Methylated and Nonmethylated DNA—We used the EMSA to compare the capability of ROS1 to bind methylated and nonmethylated DNA (Fig. 1). We found that ROS1 formed a complex with a 51-bp double-stranded DNA probe that contained a single dimethylated CG site, but also bound with similar efficiency to an equivalent unmethylated probe (Fig. 1A). Analogous results were obtained when comparing ROS1 binding to a 50-bp probe containing four dimethylated CG sites and its unmethylated counterpart (Fig. 1A). We also compared ROS1 binding with unmethylated DNA duplexes containing either 3 CG sites or no CG sites at all and found that the absence of CG dinucleotide sequences does not affect binding (supplemental Fig. S2).

We next examined in detail the binding affinity of ROS1 through competition experiments with unlabeled oligonucleotides. ROS1 binding to a methylated DNA probe was reduced with equivalent intensity when incubated with increasing amounts of either unmethylated or methylated unlabeled competitor DNA (Fig. 1B). The relative affinity of ROS1 for nonmethylated and methylated competitor DNA was similar, with dissociation constants (\(K_d\)) of 39.7 nM and 34.5 nM, respectively. Comparable results were observed when a nonmethylated DNA was used as a probe (Fig. 1B), with estimated \(K_d\) values of 36.6 nM and 34.6 nM for the unmethylated and methylated unlabeled competitor, respectively. We therefore conclude that ROS1 displays a strong DNA binding capacity that is independent of the methylation state of DNA.

Methylation-independent DNA Binding Capacity Resides in a Lysine-rich Domain at the N Terminus of ROS1—We next asked whether the nonspecific DNA binding capacity of ROS1 could be assigned to a particular region of the protein sequence. ROS1 and its homologs are large polypeptides containing a DNA glycosylase domain with significant sequence similarity to base excision DNA repair proteins in the HhH-GPD super-
Role of Lysine-rich Domain in DNA Demethylation by ROS1

family (23) (supplemental Fig. S1). In addition to the DNA glycosylase domain, the proteins of the ROS1/DME family share two other conserved domains of unknown function and unique to this class of proteins (11). An additional conserved feature is a short amino-terminal domain significantly rich in lysine (supplemental Fig. S1 and Table S3).

We generated a series of N- and C-terminal ROS1 deletion mutants, and the purified proteins were used in EMSAs to analyze their capacity to bind a methylated DNA probe (Fig. 2). Deletion of the N-terminal 294 amino acids of ROS1 yielded a protein (NΔ294) with a strongly reduced DNA binding capacity ($K_d = 7.3$ nM) compared with the full-length protein ($K_d = 34.3$ nM). The further loss of 225 amino acids from the N terminus (NΔ519) slightly increased this diminished binding activity ($K_d = 1.3$ nM) (Fig. 2). A truncated protein (NΔ519CΔ313) that additionally lacked the C-terminal 313 amino acids but still conserved the DNA glycosylase domain, did not exhibit any detectable DNA binding capacity. The C-terminal 313 amino acids fragment (NΔ1080) was able by itself to bind the DNA probe, but with very low affinity ($K_d = 2.6$ μM). Finally, a small fragment of 230 amino acids (NΔ88CΔ1075) that contained the N-terminal lysine-rich region (positions 112–260) displayed a strong DNA binding capacity ($K_d = 17.9$ nM), even higher than that of the full-length protein. Competition experiments with unmethylated and methylated competitor DNA confirmed that the DNA binding activity of this N-terminal region was independent of the methylation state of DNA (supplemental Fig. S3). Altogether, these results indicate that the methylation-independent DNA binding activity of ROS1 largely resides in a basic, lysine-rich, N-terminal region of the protein.

Deletion of the N-terminal Domain Abolishes ROS1 preference for 5-meC over T—We next analyzed the enzymatic activity of the different ROS1 deletion mutants by performing DNA glycosylase assays on a 51-mer duplex oligonucleotide substrate that contained a single 5-meCG pair (Fig. 3). Removal of the N-terminal 294 amino acids significantly reduced the 5-meC DNA glycosylase activity of ROS1, yielding a mutant protein that processed about half the amount of product than the full-length (FL) protein after 4-h and 24-h incubation time. Further elimination of 225 amino acids from the N terminus almost entirely abolished the enzymatic activity of the protein, despite preservation of those residues corresponding to the DNA glycosylase domain and the highly conserved C-terminal region. Polypeptides with further deletions did not show any detectable enzymatic activity (Fig. 3).

The strongly reduced DNA glycosylase activity of NΔ294 compared with the full-length protein suggested that the non-specific DNA binding capacity afforded by the N-terminal region of ROS1 might be important for substrate recognition and specificity. We therefore performed a kinetic analysis to examine the relative activities of FL- and NΔ294-ROS1 toward a 51-mer duplex oligonucleotide substrate that contained either 5-meC or T opposite G at position 29 in a CG context (Fig. 4 and supplemental Table S4). A simple Michaelis-Menten model is inadequate for a correct kinetic analysis of ROS1 because this protein does not exhibit significant turnover in vitro due to strong product binding (21). Therefore, we have used a previously described method (24) successfully employed
to measure and compare single-turnover kinetics with different orthologs of thymine DNA glycosylase (25). For each enzyme and substrate, we determined the parameters \( P_{\text{max}} \) (maximum substrate processing within an unlimited period of time); \( T_{50} \) (the time required to reach 50% of the product plateau level, \( P_{\text{max}} \)), and the relative processing efficiency (\( E_{\text{rel}} = P_{\text{max}}/T_{50} \)). Consistent with our previously reported observations (11, 21) we found that FL-ROS1 processed 5-meC with higher relative processing efficiency than T (Fig. 4 and supplemental Table S4). In contrast, \( \Delta \)294-ROS1 excised both 5-meC and T with similar processing efficiencies (Fig. 4 and supplemental Table S4). We therefore conclude that removal of the N terminus abrogates ROS1 preference for 5-meC over T.

**N Terminus of ROS1 Facilitates 5-meC Excision from Long DNA Substrates**—The results described above suggested that nonspecific DNA binding mediated by the N terminus of ROS1 might facilitate 5-meC recognition on DNA. We tested this hypothesis by comparing the relative abilities of FL- and \( \Delta \)294-ROS1 to excise a single 5-meC residue from a centrally located CG site in various DNA duplexes of equivalent DNA sequences but different lengths (Fig. 5A). We found that a 12-bp DNA was a poor substrate for both FL- and \( \Delta \)294-ROS1, but increasing substrate length up to 20 bp gradually enhanced the efficiency of 5-meC excision by both proteins. Interestingly, however, FL- and \( \Delta \)294-ROS1 exhibited a markedly different pattern of activity on longer DNA substrates. Whereas the full-length protein displayed a significant level of activity on DNA substrates up to 52 bp long, the efficiency of 5-meC excision by \( \Delta \)294-ROS1 on DNA duplexes longer than 20 bp was significantly reduced.

To confirm this difference further, we performed time course experiments comparing the relative capacities of FL- and \( \Delta \)294-ROS1 to excise 5-meC either from a “short” (20-bp) or a “long” (52-bp) DNA substrate (Fig. 5B). We found that, although both proteins exhibited a similar level of activity on the 20-bp substrate, the efficiency of 5-meC excision by \( \Delta \)294-ROS1 from the 52-bp DNA duplex was significantly lower than that of FL-ROS1 (Fig. 5B). We therefore conclude that the N terminus of ROS1 facilitates 5-meC excision from long DNA substrates.

**N Terminus Is Not Required for ROS1 Binding to Its Reaction Product**—We next tested whether the positive effect of the N terminus on ROS1 activity on long DNA substrates might be alternatively explained by a higher catalytic turnover due to facilitated product release. Because ROS1 is a DNA glycosylase/AP lyase that cleaves the phosphodiester backbone at the 5-meC removal site by successive \( \beta,\gamma \)-elimination, its final reaction product is a single-nucleotide gap flanked by 3’-P and 5’-P ends (11). We therefore decided to investigate whether ROS1 binds to the further N terminus of the protein may influence such binding.

We performed EMSA analysis to compare the ability of FL-ROS1 and \( \Delta \)294-ROS1 to bind either a 51-bp duplex oligonucleotide with a C-G pair at position 29, or an analogous probe with a single-nucleotide gap flanked by 3’-P and 5’-P ends (Fig. 6). We found that both proteins bound with higher affinity the duplex oligonucleotide bearing the nucleotide gap than the homoduplex DNA (Fig. 6A). To compare the relative affinities of FL-ROS1 and \( \Delta \)294-ROS1 for the single-nucleotide gap, we carried out EMSA competition analysis with a gapped DNA probe in the presence of increasing amounts of either unlabelled competitor homoduplex or gapped DNA (Fig. 6B). For both FL-ROS1 and \( \Delta \)294-ROS1, competition was significantly more effective.
Role of Lysine-rich Domain in DNA Demethylation by ROS1

FIGURE 6. N terminus is not required for ROS1 binding to its reaction product. A, increasing concentrations of FL-ROS1 (0, 30, 60, 90, and 120 nM; upper panels) or NΔ294-ROS1 (0, 45, 90, 180, and 270 nM; lower panels) were incubated with a mixture of 100 nM fluorescein-labeled 51-bp homoduplex and 100 nM Alexa Fluor-labeled single-nucleotide gapped 51-bp duplex. After nondenaturing electrophoresis, the gel was scanned to detect fluorescein- or Alexa Fluor-labeled DNA. After nondenaturing gel electrophoresis, the gel was scanned to detect fluorescein- (left panels) or Alexa Fluor-labeled DNA (right panels). B, purified FL-ROS1 (upper panel, 120 nM) or NΔ294-ROS1 (lower panel, 270 nM) was incubated with 100 nM Alexa Fluor-labeled single-nucleotide gapped 51-bp duplex in the presence of increasing amounts (0, 0.1, 0.25, 0.5, 1.0, and 2.5 μM) of either homoduplex (lanes 2–7) or single-nucleotide gapped (lanes 8–13) unlabeled competitor DNA. After nondenaturing gel electrophoresis, the gel was scanned to detect Alexa Fluor-labeled DNA. Protein–DNA complexes were identified by their retarded mobility compared with that of free DNA, as indicated. Graphs on the left show the percentage of remaining complex versus competitor molar excess ratios.

with the unlabeled gapped duplex. Thus, the affinity of FL-ROS1 for the gapped DNA ($K_d = 17.6$ nM) was about 2-fold higher than for the homoduplex ($K_d = 38.2$ nM). A similar relative increase in affinity for the gapped substrate ($K_d = 2.3$ μM) compared with the homoduplex DNA ($K_d = 7.1$ μM) was observed for the truncated protein (Fig. 6B). Altogether, these results indicate that both FL-ROS1 and NΔ294-ROS1 bind specifically to the single-nucleotide gap left after 5-meC excision and strand incision. We therefore conclude that the N terminus of ROS1 is not required for product binding.

Distributive Behavior of ROS1 Is Not Affected by Removal of the N-terminal Domain—We have previously reported that ROS1 does not exhibit significant processivity in vitro (21). The above results suggested that a major factor explaining the lack of processivity of ROS1 is the strong binding of the enzyme to its reaction product, the single-nucleotide gap. We reasoned that if the N terminus of the enzyme is not involved in such product binding, then the distributive behavior of ROS1 should not be significantly affected by the loss of the N-terminal domain. To test this prediction we examined whether the distributive behavior of ROS1 is altered in the truncated NΔ294-ROS1 protein.

We incubated either FL-ROS1 or NΔ294-ROS1 with a double-stranded oligonucleotide substrate containing three 5-meC residues in the upper strand separated by nine nucleotides and located in the same sequence context (Fig. 7) (see supplemental Table S1). Whereas a processive mechanism will rapidly produce a completely processed substrate represented by a 16-nucleotide labeled fragment, a distributive mechanism will lead to the additional accumulation of partially processed reaction intermediates represented by 26- and 36-nucleotide labeled fragments (21). Consistent with our previously reported observations, we found that FL-ROS1 processed 5-meC in a highly distributive fashion, exhibiting a steady accumulation of the 16-nucleotide fragment but also of significant amounts of partially processed substrates, even after long incubation times (Fig. 7B). As expected, at equimolar concentration NΔ294-ROS1 displayed a lower activity than the full-length enzyme, but generated a very similar accumulation pattern of completely and partially processed substrates (Fig. 7B). Analogous results were obtained when the analysis was performed with lower concentrations of both enzymes (supplemental Fig. S4). We therefore conclude that the N terminus of ROS1 mediates nonspecific binding to DNA but does not influence the distributive behavior of the enzyme, which is rather caused by strong binding of the catalytic domain to the reaction product.

DISCUSSION

The potential of ROS1 and its homologs to remove a modified, but otherwise nondamaged, correctly paired base is rather exceptional among DNA glycosylases. In an effort to understand how 5-meC DNA glycosylases recognize and excise their target, we have identified ROS1 regions required for DNA binding and catalysis. We have found that a lysine-rich domain at the N terminus of ROS1 endows the enzyme with the capacity to bind DNA strongly in a methylation-independent manner and greatly facilitates 5-meC excision from long DNA substrates. Similar lysine-rich regions are present at the N-terminal extension of distant members of the ROS1/DME family of 5-meC DNA glycosylases (supplemental Fig. S1 and
Role of Lysine-rich Domain in DNA Demethylation by ROS1

supplemental Table S3), which suggests an evolutionarily conserved important function.

Previous studies have reported that noncatalytic N-terminal extensions of DNA glycosylases may regulate their enzymatic activity, usually through modulation of product dissociation rates (26, 27). Thus, the N termini of mouse and human endonuclease III homolog facilitate (27) and reduce (26) product dissociation rates, respectively. We have found, however, that the N-terminal extension of ROS1 is not required for the strong product binding capacity exhibited by this protein. In agreement with this observation, both the N-terminally truncated version and the full-length enzyme display a similar distributive behavior when acting on a DNA substrate containing multiple 5-meC targets. It might seem rather unexpected that the non-specific DNA binding afforded by the N terminus does not endow FL-ROS1 with higher processivity than NA294-ROS1. However, it should be kept in mind that nonspecific DNA binding merely increases the probability that the enzyme remains bound to the DNA after one step of random walk, thus effectively increasing the protein capacity to slide or scan the molecule (28). This capacity does not necessarily increase processivity, which specifically refers to the probability of the enzyme resuming the walk after a catalytic cycle is complete (28). Processivity is strongly dependent on the rate of product release. Therefore, a DNA glycosylase may be able to efficiently slide DNA in search for a target base, but nevertheless lack processivity because it remains strongly bound to the abasic site generated after base excision. Altogether, our results suggest that the N terminus of ROS1 facilitates target location through non-specific DNA binding but does not relieve the highly distributive behavior of the enzyme caused by robust binding to the reaction product through the catalytic domain.

Importantly, the positive effect exerted by the N-terminal region on ROS1 activity is highly dependent on substrate length. We have found that removal of the N terminus does not significantly affect 5-meC excision from short molecules, but strongly decreases ROS1 activity on long DNA substrates. This length-dependent effect strongly suggests that 5-meC recognition involves initial nonspecific binding events to nontarget sites followed by sliding along DNA. The idea of DNA sliding as a diffusion process that effectively reduces search dimensions and facilitates target location was originally proposed for the E. coli lac repressor-operator interaction (29) and has found experimental support in a number of systems, including the highly processive T4 pyrimidine dimer glycosylase (30). Direct evidence of fast DNA glycosylase sliding along DNA has been obtained for human hOGG1 (31) and E. coli Ung (32).

Our data are consistent with the idea that ROS1 first binds DNA nonspecifically and then slides randomly toward its target base along the DNA molecule. We speculate that when 5-meC is present in short DNA molecules target detection by ROS1 is expected to approximate a three-dimensional search and, hence, be slow. This would explain why both the full-length and the N-terminally truncated enzyme display reduced activity on short (<20 bp) substrates. On the other hand, when 5-meC is located in long DNA molecules, target search may be accelerated by sliding. In this situation, a basic region affording non-specific DNA binding capacity would allow ROS1 to scan the DNA in a one-dimensional diffusion, which might provide the enzyme with ample opportunity to recognize and excise a base modification that minimally affects DNA structure.

To our knowledge, there are no published studies specifically addressing whether nonspecific DNA binding facilitates DNA glycosylase activity on long substrates. However, there is some evidence that basic domains are involved in efficient processing of specific target bases by other DNA glycosylases. Thus, deletion of a lysine-rich domain at the N terminus of human thymine DNA glycosylase selectively abolishes T-G processing, while decreasing only marginally the activity on U-G mispairs (33, 34). Interestingly, such a deletion is simultaneously accompanied by a drastic decrease in nonspecific DNA binding of the enzyme to homoduplex, undamaged DNA (34). It has been proposed that T-G is a suboptimal substrate for thymine DNA glycosylase, compared with U-G, and that its processing is somehow facilitated by nonspecific DNA binding (35). Likewise, our results suggest that strong binding of ROS1 to DNA increases excision of a challenging target (a correctly paired 5-meC) but do not affect processing of a less demanding substrate (its mispaired deamination derivative, T).

It is unclear how nonspecific DNA binding may influence substrate discrimination by DNA glycosylases. One possibility is that the structural and mechanistic features of target base binding are different for different substrates. A model suggested for E. coli Fpg, which excises 8-oxo-7,8-dihydroguanine and less efficiently 5,6-dihydrouracil, postulates two modes of
lesion binding. In the “run-on” mode the enzyme approaches the lesion while diffusing along DNA, whereas in the “jump-on” mode it binds in the immediate vicinity of the lesion (36). The model proposes that Fpg forms productive complexes in both modes with 8-oxo-7,8-dihydroguanine, but only in the latter mode with 5,6-dihydrouracil (36). In such a scenario, non-specific DNA binding may specifically facilitate excision of just a subset of the base modifications potentially targeted by an enzyme.

In summary, we found that methylation-independent DNA binding through a lysine-rich domain conserved across the ROS1/DME family selectively favors 5-meC over T excision and facilitates demethylation in long DNA substrates. It remains to be seen whether the accretion of basic domains that mediate nonspecific DNA binding has been a recurrent phenomenon in the evolution of DNA glycosylases able to excise suboptimal target bases.

Acknowledgments—We thank members of our laboratory for helpful discussions.

REFERENCES
1. Goll, M. G., and Bestor, T. H. (2005) Annu. Rev. Biochem. 74, 481–514
2. Klose, R. J., and Bird, A. P. (2006) Trends Biochem. Sci. 31, 89–97
3. Weber, M., and Schübeler, D. (2007) Curr. Opin. Cell Biol. 19, 273–280
4. Robertson, K. D. (2005) Nat. Rev. Genet. 6, 597–610
5. Esteller, M. (2007) Nat. Rev. Genet. 8, 286–298
6. Roldan-Arjona, T., and Ariza, R. R. (2009) in DNA and RNA Modification Enzymes: Comparative Structure, Mechanism, Functions, Cellular Interactions and Evolution (Grosjean, H., ed), pp. 149–161, Landes Bioscience, Austin, TX
7. Zhu, J. K. (2009) Annu. Rev. Genet. 43, 143–166
8. Ooi, S. K., and Bestor, T. H. (2008) Cell 133, 1145–1148
9. Gong, Z., Morales-Ruiz, T., Ariza, R. R., Roldán-Arjona, T., David, L., and Zhu, J. K. (2002) Cell 111, 803–814
10. Choi, Y., Gehring, M., Johnson, H., Hannon, M., Harada, J. J., Goldberg, R. B., Jacobsen, S. E., and Fischer, R. L. (2002) Cell 110, 33–42
11. Morales-Ruiz, T., Ortega-Galisteo, A. P., Ponferrada-Marin, M. I., Martinez-Macías, M. I., Ariza, R. R., and Roldán-Arjona, T. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 6853–6858
12. Gehring, M., Huh, J. H., Hsieh, T. F., Penterman, J., Choi, Y., Harada, J. J., Goldberg, R. B., and Fischer, R. L. (2006) Cell 124, 495–506
13. Agius, F., Kapoor, A., and Zhu, J. K. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 11796–11801
14. Penterman, J., Zilberman, D., Huh, J. H., Ballinger, T., Henikoff, S., and Fischer, R. L. (2007) Proc. Natl. Acad. Sci. U.S.A. 104, 6752–6757
15. Ortega-Galisteo, A. P., Morales-Ruiz, T., Ariza, R. R., and Roldán-Arjona, T. (2008) Plant Mol. Biol. 67, 671–681
16. Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S. E., Fischer, R. L., and Kakutani, T. (2004) Science 303, 521–523
17. Hsieh, T. F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L., and Zilberman, D. (2009) Science 324, 1451–1454
18. Gehring, M., Bubb, K. L., and Henikoff, S. (2009) Science 324, 1447–1451
19. Zhu, J., Kapoor, A., Sridhar, V. V., Agius, F., and Zhu, J. K. (2007) Curr. Biol. 17, 54–59
20. Zheng, X., Pontes, O., Zhu, J., Miki, D., Zhang, F., Li, W. X., Iida, K., Kapoor, A., Pikaard, C. S., and Zhu, J. K. (2008) Nature 455, 1259–1262
21. Ponferrada-Marín, M. I., Roldán-Arjona, T., and Ariza, R. R. (2009) Nucleic Acids Res. 37, 4264–4274
22. Tan, T., and Chu, G. (2002) Mol. Cell. Biol. 22, 3247–3254
23. Nash, H. M., Bruner, S. D., Schärer, O. D., Kawate, T., Addona, T. A., Spooner, E., Lane, W. S., and Verdine, G. L. (1996) Curr. Biol. 6, 968–980
24. Hardeland, U., Bentele, M., Jiricny, J., and Schär, P. (2000) J. Biol. Chem. 275, 33449–33456
25. Hardeland, U., Bentele, M., Jiricny, J., and Schär, P. (2003) Nucleic Acids Res. 31, 2261–2271
26. Liu, X., and Roy, R. (2002) J. Mol. Biol. 321, 265–276
27. Adhikari, S., Uren, A., and Roy, R. (2007) J. Biol. Chem. 282, 30078–30084
28. Zharkov, D. O., and Grollman, A. P. (2005) Mutat. Res. 577, 24–54
29. von Hippel, P. H., and Berg, O. G. (1989) J. Mol. Biol. 220, 675–678
30. Lloyd, R. S. (2005) Mutat. Res. 577, 77–91
31. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L., and Xie, X. S. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 5752–5757
32. Porecha, R. H., and Stivers, J. T. (2008) Proc. Natl. Acad. Sci. U.S.A. 105, 10791–10796
33. Gallnini, P., and Jiricny, J. (1996) Nature 383, 735–738
34. Steinacher, R., and Schär, P. (2005) Curr. Biol. 15, 616–623
35. Cortázar, D., Kunz, C., Saito, Y., Steinacher, R., and Schär, P. (2007) DNA Repair 6, 489–504
36. Zaika, E. I., Perlow, R. A., Matz, E., Broyde, S., Gilboa, R., Grollman, A. P., and Zharkov, D. O. (2004) J. Biol. Chem. 279, 4849–4861