Supplemental material for:

A discontinuous DNA glycosylase domain in a family of enzymes that excise 5-methylcytosine

María Isabel Ponferrada-Marín, Jara Teresa Parrilla-Doblas, Teresa Roldán-Arjona and Rafael R. Ariza

Department of Genetics, University of Córdoba, Spain

Table S1. Oligonucleotides used as substrates

Name	DNA Sequence	Strand	X=
FL-CGF	5´-TCACGGGATCAATGTGTTCTTTCAGCTC**X**GGTCACGCTGACCAGGAATACC-3´	Upper	C
FL-meCGF	5´-TCACGGGATCAATGTGTTCTTTCAGCTC**X**GGTCACGCTGACCAGGAATACC-3´	Upper	5-meC
FL-TGF	5´-TCACGGGATCAATGTGTTCTTTCAGCTC**X**GGTCACGCTGACCAGGAATACC-3´	Upper	T
FL-HUGF	5´-TCACGGGATCAATGTGTTCTTTCAGCTC**X**GGTCACGCTGACCAGGAATACC-3´	Upper	5-HU
FL-UGF	5´-TCACGGGATCAATGTGTTCTTTCAGCTC**X**GGTCACGCTGACCAGGAATACC-3´	Upper	U
FL-BrC	5´-TCACGGGATCAATGTGTTCTTTCAGCTC**X**GGTCACGCTGACCAGGAATACC-3´	Upper	5-BrC
FL-BrU	5´-TCACGGGATCAATGTGTTCTTTCAGCTC**X**GGTCACGCTGACCAGGAATACC-3´	Upper	5-BrU
Fl-MCGF	5´-TCACGGGATCAATGTGTTCTTTCAGCT**X**CGGTCACGCTGACCAGGAATACC-3´	Upper	5-meC
CGR	3´-AGTGCCCTAGTTACACAAGAAAGTCGAGG**X**CAGTGCGACTGGTCCTTATGG-5´	Lower	C
FL-MXF	5´-TCACGGGATCAATGTGTTCTTTCATATA**X**TATAACGCTGACCAGGAATACC-3´	Upper	5-meC
CXR	3´-AGTGCCCTAGTTACACAAGAAAGTCTATGATAGTGCGACTGGTCCTTATGG-5´	Lower	-
FL-MAGF	5´-TCACGGGATCAATGTGTTCTTTCAGCTC**X**AGGTCACGCTGACCAGGAATACC-3´	Upper	5-meC
CAGR	3´-AGTGCCCTAGTTACACAAGAAAGTCGAGGT**X**CAGTGCGACTGGTCCTTATGG-5´	Lower	C
Al-28P	5´-TCACGGGATCAATGTGTTCTTTCAGCTC-3´	Upper	-
P30_51	5´-GGTCACGCTGACCAGGAATACC-3´	Upper	-

[a] Relevant regions are boxed.

Table S2. Oligonucleotides used as primers for generation of ROS1 mutant variants

Name	**Sequence (5′- 3′)**[a]
ROS1Q584L-F	GGATGCACCTTGTACTAGGGGATCGACGTTTTACGCCTTG
ROS1Q584L-R	CAAGGCGTAAAACTGCGATCCCCTAGTACAAGGTGCATCC
ROS1F589A-F	AGGGGATCGACGTGCTACGCCTTGGAAGGG
ROS1F589A-R	CCCTTCCAAGGCGTAGCACGTCGATCCCCT
ROS1T606L-F	GTATTTCTCCTTCAAAATGTTTCAGACCATCTCTCAAGTTCGGC
ROS1T606L-R	GCCGAACTTGAGAGATGGTCTGAAACATTTTGAAGGAGAAATAC
ROS1Q607A-F	GGAGTATTTCTCACTGCAAATGTTTCAGACCATCTCTCAAG
ROS1Q607A-R	CTTGAGAGATGGTCTGAAACATTTGCAGTGAGAAATACTCC
ROS1N608A-F	GTATTTCTCACTCAAGCTGTTTCAGACCATCTCTCAAGTTCG
ROS1N608A-R	CGAACTTGAGAGATGGTCTGAAACAGCTTGAGTGAGAAATAC
ROS1D611V-F	TCTCACTCAAAATGTTTCAGTCCATCTCTCAAGTTCGG
ROS1D611V-R	CCGAACTTGAGAGATGGACTGAAACATTTTGAGTGAGA
ROS1W1012A-F	TGAATCTATTCAAAAGTATCTTGCGCCCCGTCTCTGC
ROS1W1012A -R	GCAGAGACGGGGCGCAAGATACTTTTGAATAGATTCA
ROS1Y1028S-F	CATTGTATGAGTTGCACTCCCAGATGATTACTTTTGGAAAGG
ROS1Y1028S -R	CCTTTCCAAAAGTAATCATCTGGGAGTGCAACTCATACAATG

[a]Underlining indicates the mutagenized codon.

FIGURE S1

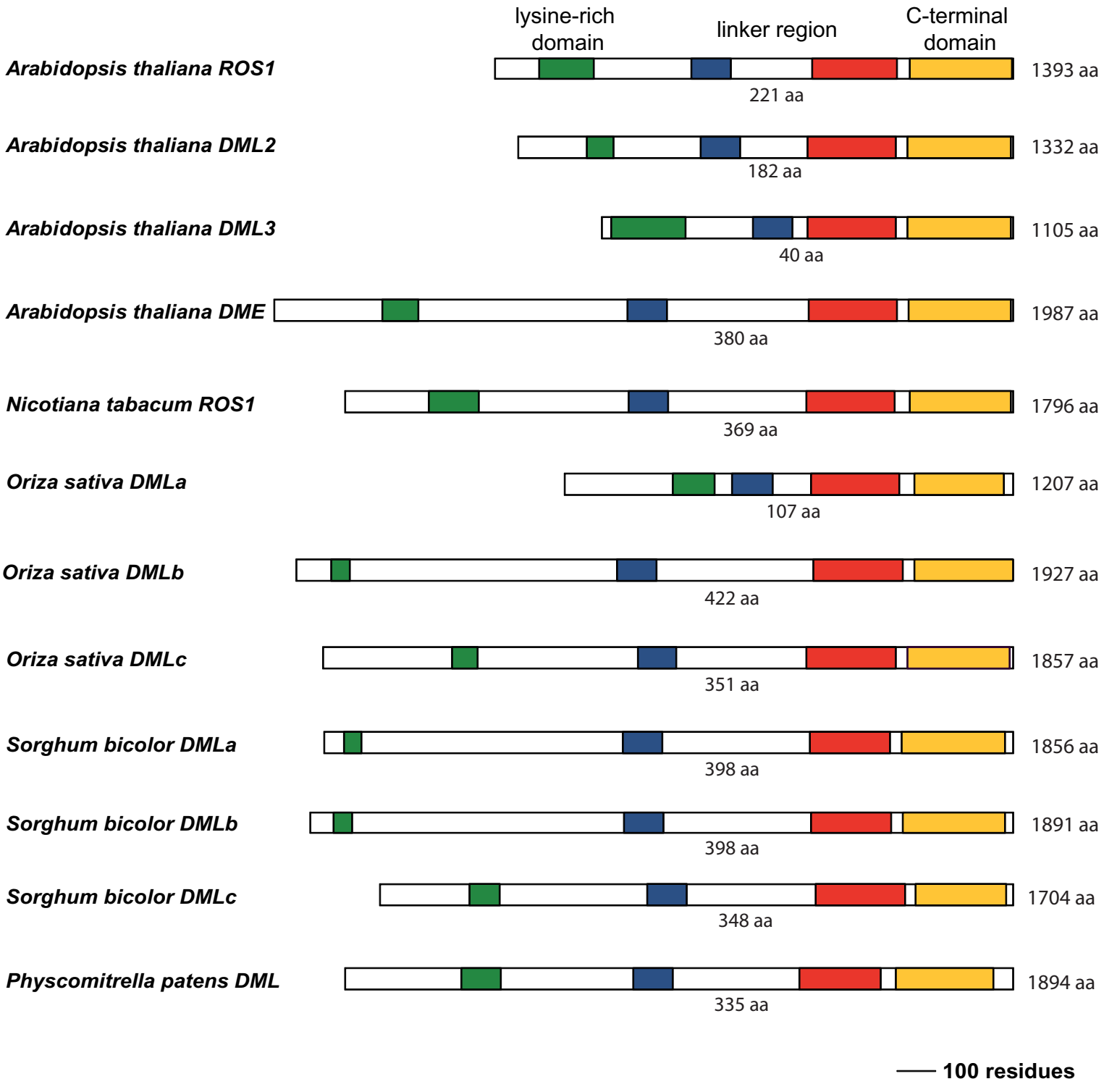

Figure S1. Schematic alignment of members of the DML family. Members of this subfamily of DNA glycosylases possess a lysine-rich domain (green) at the amino-terminus and a domain of unknown function unique to this class of proteins at the carboxy-terminus (yellow). Sequence similarity to HhH-GPD proteins is distributed over two non-contiguous segments (blue and red) connected by a non-conserved linker region.

FIGURE S2

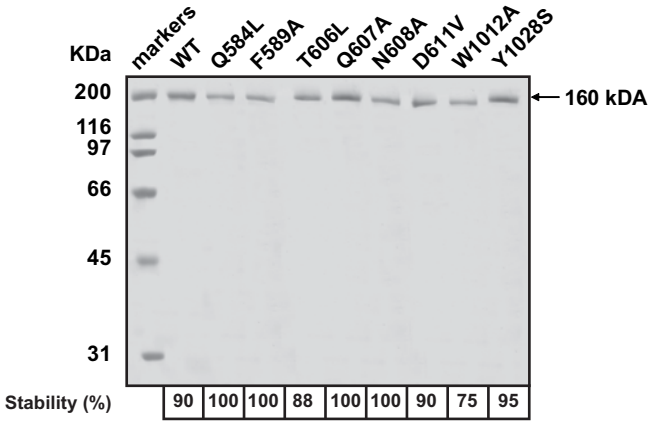

Figure S2. SDS-PAGE analysis of wild-type and mutant ROS1 proteins. The 10% SDS-PAGE gel was loaded with 0.4 µg of purified protein per lane and stained with Coomassie blue. The percentage of stable protein after 4 h incubation under assay conditions (see Materials and Methods) is indicated below each lane.

FIGURE S3

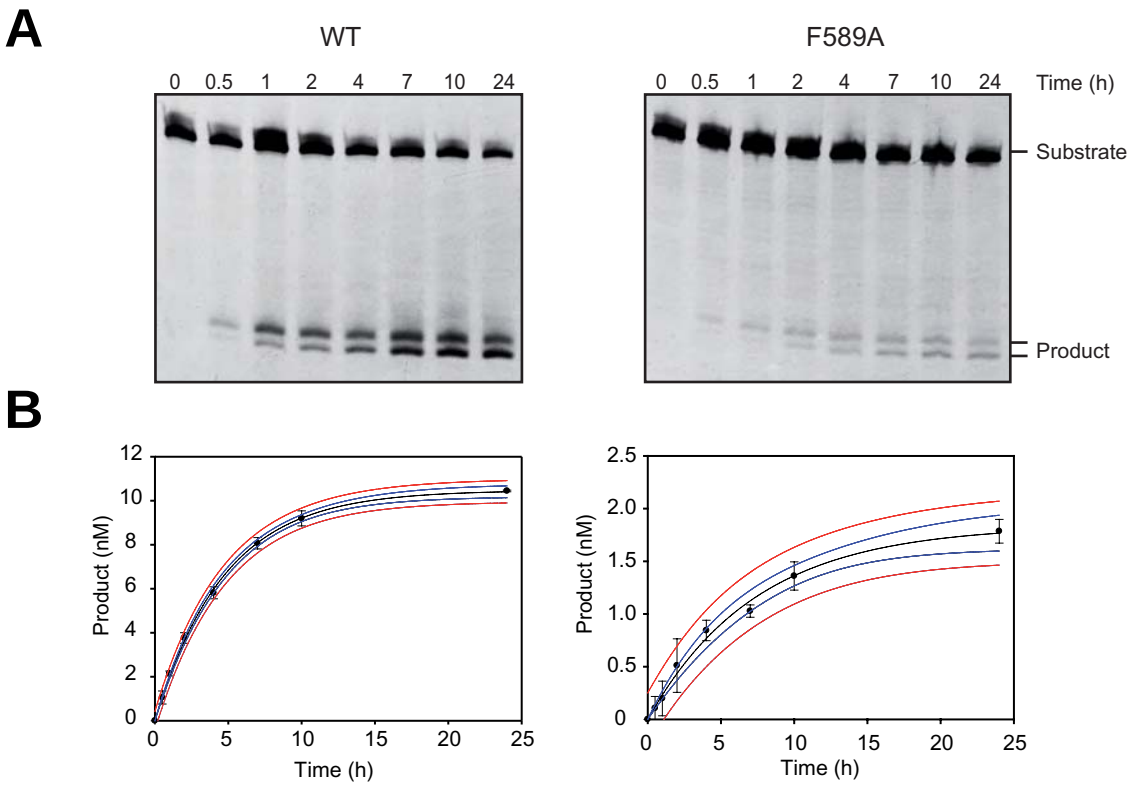

Figure S3. Representative examples of 5-meC DNA glycosylase assay and kinetic analysis. A. The time-dependent generation of incision products was measured by incubating purified WT ROS1 (left panel) or mutant variant F589A (right panel) (20 nM) at 30 °C with a 51-mer double-stranded oligonucleotide substrate (20 nM) containing a single 5-meC:G pair. Reactions were stopped at the indicated times, products were separated in a 12% denaturing polyacrylamide gel, and the amount of incised oligonucleotide was quantified by fluorescent scanning. **B.** Graphs show the generation of incision products *versus* time for WT ROS1 (left panel) and F589A mutant (right panel). Values are means ± S.E. (error bars) from two independent experiments. Data were fitted to the equation [Product] = $P_{max}(1\text{-}exp^{(-kt)})$ using non-linear regression analysis. Blue and red curves indicate 95% confidence and prediction intervals, respectively.

FIGURE S4

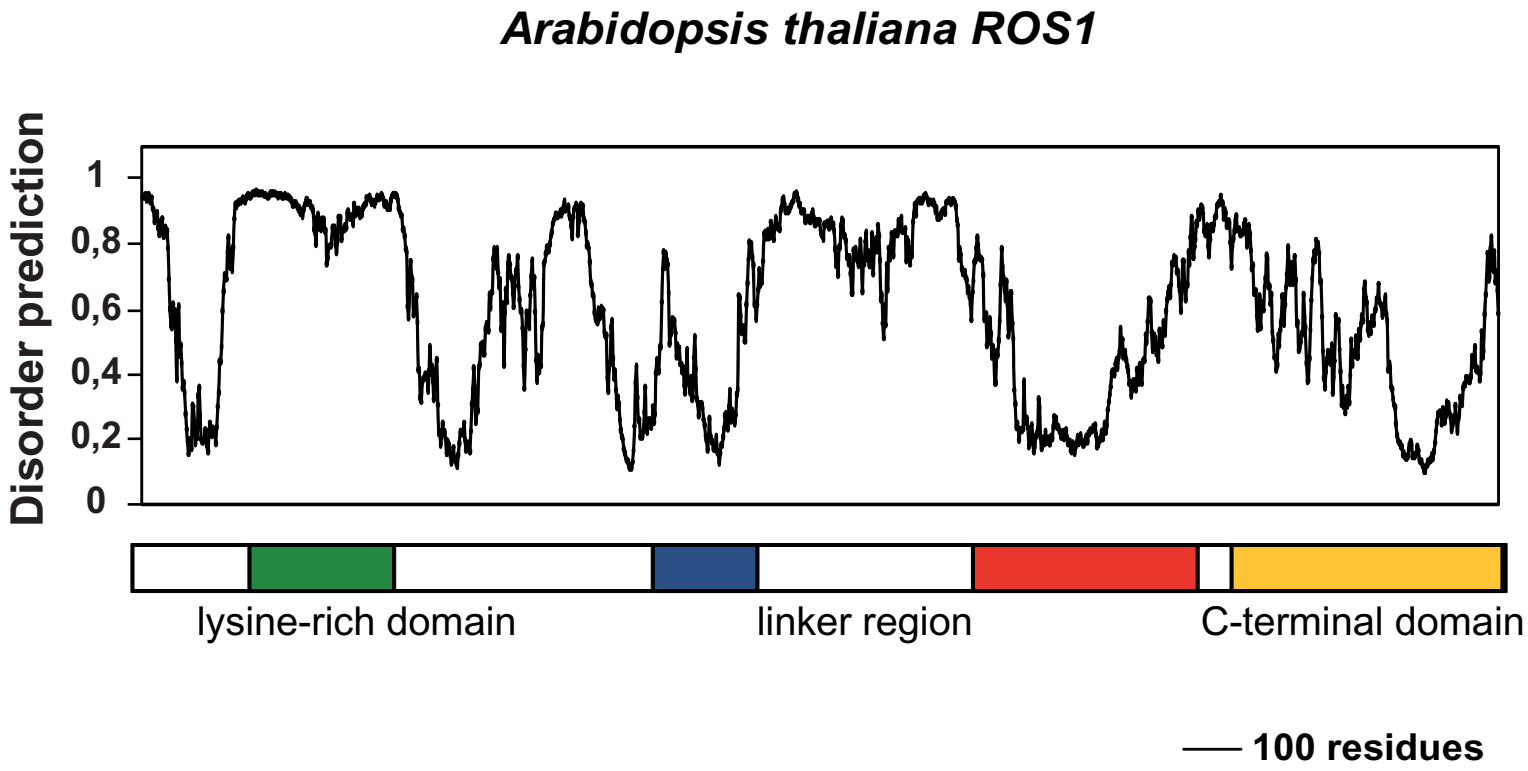

Figure S4. Disorder prediction analysis. Prediction of native disordered regions in the sequence of ROS1 by using the disorder-prediction program VL3H (http://www.ist.temple.edu/disprot/Predictors.html). Residues with a VL3H score > 0.5 are predicted to be disordered.

FIGURE S5

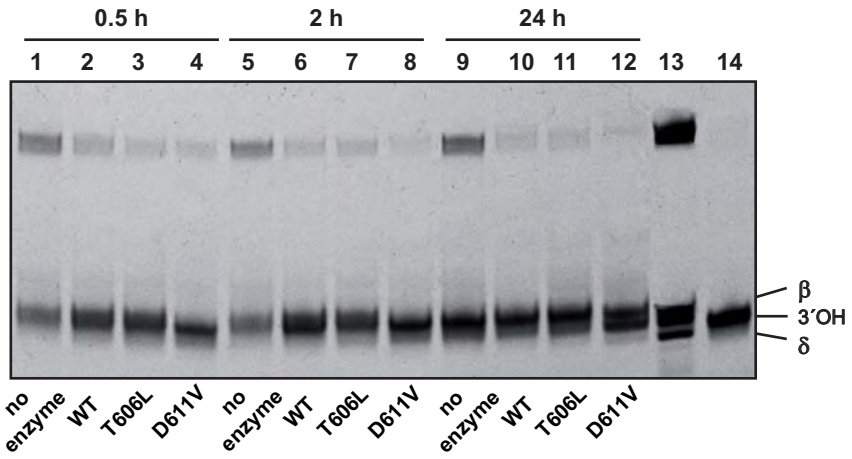

Figure S5. AP-lyase assay. A double-stranded oligonucleotide substrate containing an AP site opposite G (200 nM) was incubated at 30ºC either in the absence of enzyme (lanes 1, 5 and 9), or in the presence of 100 nM purified WT ROS1 (lanes 2, 6 and 10), T606L (lanes 3, 7 and 11) or D611V (lanes 4, 8 and 12). Reactions were stopped at the indicated times, products were separated in a 12% denaturing polyacrylamide gel and the amount of incised oligonucleotide was quantified by fluorescent scanning. Lane 13: control reaction of a 5-meC:G substrate (20 nM) incubated with WT ROS1 (20 nM) at 30 ºC for 3 h. Lane 14: control reaction of the AP-site substrate (100 nM) incubated with APE I (5 U) for 1 h. The β- and δ-elimination products and those carrying a 3´-OH terminus are indicated by arrows.

FIGURE S6

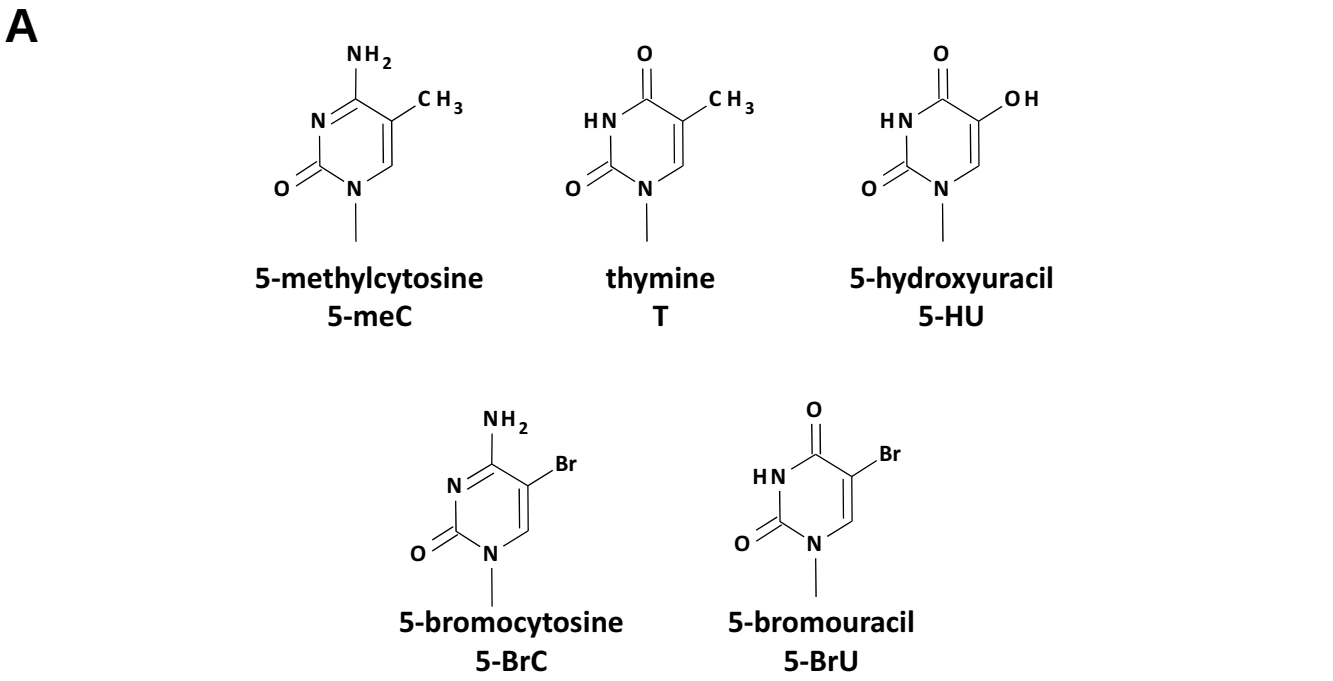

Figure S6. Incision activity of wild type ROS1 and mutant variants on duplex DNA substrates containing different modified bases. (A) Chemical structures of substrate DNA bases tested. (B) Purified WT ROS1 or mutant variants (20 nM) were incubated at 30ºC for 4 h with 51-mer double-stranded oligonucleotide substrates (20 nM) containing at position 29 of the labeled upper-strand different target DNA bases paired with G. Products were separated in a 12% denaturing polyacrylamide gel and the amount of incised oligonucleotide was quantified by fluorescent scanning.

FIGURE S7

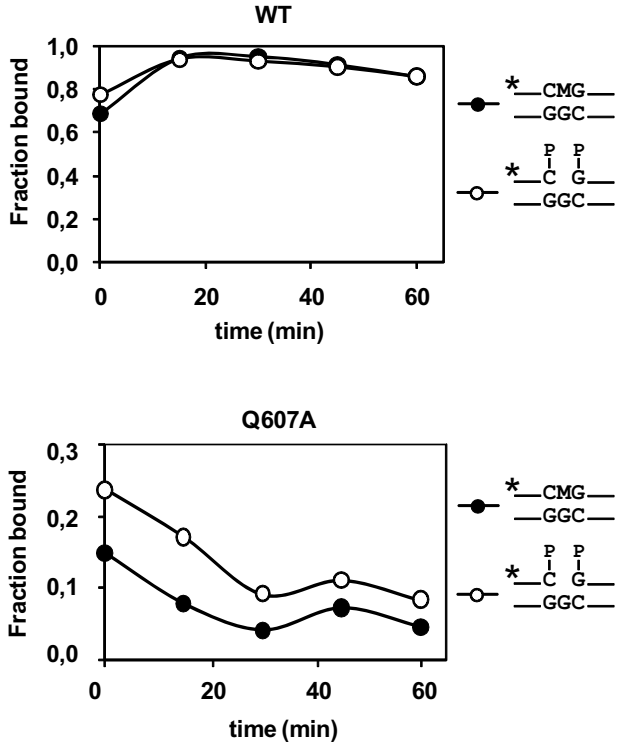

Figure S7. Q607 is required for stable ROS1 binding to substrate and product DNA. Fraction of 5-meC:G substrate or 1-nt gapped product bound by WT ROS1 or Q607A proteins at different incubation times. Data are from the gel shown in Figure 5.