Вавиловский журнал генетики и селекции. 2022;26(4):341-348
DOI 10.18699/VJGB-22-41

Original Russian text www.bionet.nsc.ru/vogis/

Characterization of demethylating DNA glycosylase ROS1 from *Nicotiana tabacum* L.

D.V. Petrova[1], N.V. Permyakova[2], I.R. Grin[1], D.O. Zharkov[1, 3]

[1] Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Novosibirsk State University, Novosibirsk, Russia
dzharkov@niboch.nsc.ru

Abstract. One of the main mechanisms of epigenetic regulation in higher eukaryotes is based on the methylation of cytosine at the C5 position with the formation of 5-methylcytosine (mC), which is further recognized by regulatory proteins. In mammals, methylation mainly occurs in CG dinucleotides, while in plants it targets CG, CHG, and CHH sequences (H is any base but G). Correct maintenance of the DNA methylation status is based on the balance of methylation, passive demethylation, and active demethylation. While in mammals active demethylation is based on targeted regulated damage to mC in DNA followed by the action of repair enzymes, demethylation in plants is performed by specialized DNA glycosylases that hydrolyze the *N*-glycosidic bond of mC nucleotides. The genome of the model plant *Arabidopsis thaliana* encodes four paralogous proteins, two of which, DEMETER (DME) and REPRESSOR OF SILENCING 1 (ROS1), possess 5-methylcytosine-DNA glycosylase activity and are necessary for the regulation of development, response to infections and abiotic stress and silencing of transgenes and mobile elements. Homologues of DME and ROS1 are present in all plant groups; however, outside *A. thaliana*, they are poorly studied. Here we report the properties of a recombinant fragment of the ROS1 protein from *Nicotiana tabacum* (NtROS1), which contains all main structural domains required for catalytic activity. Using homologous modeling, we have constructed a structural model of NtROS1, which revealed folding characteristic of DNA glycosylases of the helix–hairpin–helix structural superfamily. The recombinant NtROS1 protein was able to remove mC bases from DNA, and the enzyme activity was barely affected by the methylation status of CG dinucleotides in the opposite strand. The enzyme removed 5-hydroxymethylcytosine (hmC) from DNA with a lower efficiency, showing minimal activity in the presence of mC in the opposite strand. Expression of the NtROS1 gene in cultured human cells resulted in a global decrease in the level of genomic DNA methylation. In general, it can be said that the NtROS1 protein and other homologues of DME and ROS1 represent a promising scaffold for engineering enzymes to analyze the status of epigenetic methylation and to control gene activity.
Key words: epigenetic demethylation; 5-methylcytosine; 5-hydroxymethylcytosine; DNA glycosylases; REPRESSOR OF SILENCING 1; *Nicotiana tabacum*.

For citation: Petrova D.V., Permyakova N.V., Grin I.R., Zharkov D.O. Characterization of demethylating DNA glycosylase ROS1 from *Nicotiana tabacum* L. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2022;26(4):341-348. DOI 10.18699/VJGB-22-41

Характеристика деметилирующей ДНК-гликозилазы ROS1 из *Nicotiana tabacum* L.

Д.В. Петрова[1], Н.В. Пермякова[2], И.Р. Грин[1], Д.О. Жарков[1, 3]

[1] Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
dzharkov@niboch.nsc.ru

Аннотация. Один из главных механизмов эпигенетической регуляции у высших эукариот основан на метилировании цитозина по положению С5 с образованием 5-метилцитозина (mC), который далее узнается регуляторными белками. У млекопитающих метилирование преимущественно протекает в динуклеотидах CG, тогда как у растений его мишенью служат последовательности CG, CHG и CHH (H – любое основание, кроме G). Корректное поддержание статуса метилирования ДНК требует баланса процессов метилирования, пассивного и активного деметилирования. В то время как у млекопитающих активное деметилирование происходит за счет направленного регулируемого повреждения mC в ДНК с последующим действием ферментов репарации, у растений функции деметилирования выполняют специализированные ДНК-гликозилазы, гидролизующие *N*-гликозидную связь mC-нуклеотидов. Геном модельного растения *Arabidopsis thaliana* кодирует че-

тыре паралогичных белка, два из которых – DEMETER (DME) и REPRESSOR OF SILENCING 1 (ROS1) – обладают 5-метилцитозин-ДНК-гликозилазной активностью и необходимы для регуляции развития, ответа на инфекции и абиотический стресс и сайленсинга трансгенов и мобильных элементов. Гомологи DME и ROS1 присутствуют во всех группах растений, однако за пределами *A. thaliana* исследованы крайне слабо. В статье приведены результаты изучения свойств рекомбинантного фрагмента белка ROS1 из *Nicotiana tabacum* (NtROS1), содержащего основные структурные домены, необходимые для каталитической активности. Методами гомологичного моделирования была построена структурная модель NtROS1, в которой выявлена укладка, характерная для ДНК-гликозилаз структурного суперсемейства «спираль–шпилька–спираль». Рекомбинантный белок NtROS1 был способен удалять из ДНК основания mC, причем активность фермента слабо зависела от статуса метилирования CG-динуклеотидов в противоположной цепи. С меньшей эффективностью фермент удалял из ДНК 5-гидроксиметилцитозин (hmC), проявляя минимальную активность при наличии mC в противоположной цепи. При экспрессии гена *NtROS1* в клетках человека в культуре происходило глобальное снижение уровня метилирования геномной ДНК. В целом можно сказать, что белок NtROS1 и другие гомологи DME и ROS1 представляют собой многообещающую основу для инженерии ферментов с целью анализа статуса эпигенетического метилирования и управления активностью генов.
Ключевые слова: эпигенетическое деметилирование; 5-метилцитозин; 5-гидроксиметилцитозин; ДНК-гликозилазы; REPRESSOR OF SILENCING 1; *Nicotiana tabacum*.

Introduction

DNA methylation is a dedicated mechanism of gene regulation, especially developed in higher eukaryotes. The 5-methylcytosine (mC) nucleobase, formed by cytosine methylation at the C5 position, serves as a reversible epigenetic mark that plays an important role in the control of gene expression and the protection of the genome from mobile elements. DNA methylation occurs in a wide range of multicellular eukaryotes; however, its significance and functions in these organisms differ greatly (Lee et al., 2010; Zemach, Zilberman, 2010). For example, in mammals, methylation most often occurs in CpG dinucleotides, while in plants, a significant proportion of mC occurs in trinucleotides CHG and CHH (where H is "not G"). The consequences of mC for gene activity are mainly mediated by proteins containing methyl-binding domains that form complexes with histone deacetylases or themselves have the activity of histone-specific methyltransferases, chromatin remodeling factors, etc., which leads to chromatin condensation and transcription suppression (Ballestar, Wolffe, 2001; Baubec et al., 2013). Recent studies have shown that an oxidized derivative of 5-methylcytosine, 5-hydroxymethylcytosine (hmC), also plays an epigenetic role in the mammalian genome (Branco et al., 2011). Unlike mC, hmC is enriched in promoters and bodies of actively expressed genes and is considered an activating epigenetic marker (Pastor et al., 2011; Yu et al., 2012).

Correct methylation of various sites in the genome is extremely important, since the transcriptional activity of genes depends on it. Errors in DNA methylation can have grave consequences. In particular, in humans, global DNA demethylation or hypermethylation of tumor suppressor genes serve as cancer markers. Maintaining the status of DNA methylation in cells requires a balance of methylation and active and passive demethylation. The mechanisms of active demethylation of the genomic DNA in higher eukaryotes have only been discovered in the last decade. In mammals, active demethylation is initiated by regulated damage to mC, which can occur in two ways: either deamination of mC to T by AID/APOBEC enzymes, or oxidation of mC to hmC and further derivatives (5-formylcytosine and 5-carboxycytosine) by TET family dioxygenases (Pastor et al., 2013; Bochtler et al., 2017; Wu, Zhang, 2017). The modified bases are further perceived by cellular repair systems as damaged and are removed by the DNA base excision repair pathway.

Unlike mammals, plants have unique enzymes that directly hydrolyze *N*-glycosidic bonds of mC nucleotides. These DNA glycosylases – DEMETER (DME) and REPRESSOR OF SILENCING 1 (ROS1, also known as DEMETER-LIKE 1 or DML1) (Choi Y. et al., 2002; Gong et al., 2002; Agius et al., 2006; Morales-Ruiz et al., 2006) – are involved in the regulation of the methylation status of the sites in the plant genome that determine gene imprinting during paternal or maternal inheritance and silence or activate specific promoters during plant development and stress response (Li Y. et al., 2018; Parrilla-Doblas et al., 2019; Roldán-Arjona et al., 2019). After removal of mC, the resulting apurinic-apyrimidinic site (AP site) is cleaved either by the enzyme's own AP lyase activity or by the AP endonucleases APE1L or ARP, then a normal nucleotide is incorporated by one of the DNA polymerases, and the nick is ligated by the LIG1 DNA ligase. Interestingly, DME and ROS1 can also excise hmC, which is not regarded as an epigenetic base in plants (Jang et al., 2014). The localization of demethylation by DME/ROS1 is regulated by small RNAs that bind to the enzyme itself or to the protein complex that contains it (Penterman et al., 2007; Li X. et al., 2012).

In addition to DME and ROS1, the genome of the model plant *Arabidopsis thaliana* contains three more genes that are homologous to DME and ROS1: DEMETER-LIKE 2 (DML2), DEMETER-LIKE 3 (DML3), and AT3G47830, which has not been characterized thus far except for participation of DML2 and DML3 in maintaining correct DNA methylation (Ortega-Galisteo et al., 2008; Le et al., 2014). All these proteins belong to the DNA helix–hairpin–helix (HhH) structural superfamily of DNA glycosylases. Other members of this superfamily are involved in the removal of oxidized, alkylated and deaminated nucleobases from the genome (Zharkov, 2008; Fedorova et al., 2010). DME/ROS1 enzymes attract attention as potential tools for targeted regulation of gene activity: for example, the possibility of targeted DNA demethylation in human cells by *A. thaliana* ROS1 (AtROS1) fused to the RNA-guided Cas9 protein has been shown (Devesa-Guerra et al., 2020), and

A. thaliana DME (AtDME) was used to analyze the level of mC in genomic DNA (Choi W.L. et al., 2021).

In plants other than *A. thaliana*, there were few studies on the role of DME-like proteins in active epigenetic demethylation; some data exist for rice, wheat, barley, and tomato (Ono et al., 2012; Wen et al., 2012; Kapazoglou et al., 2013; Liu et al., 2015). In 2007, a ROS1 homolog from *Nicotiana tabacum* (NtROS1) was cloned, and the recombinant protein produced in insect cell culture was shown to cleave methylated tobacco genomic DNA (Choi C.-S., Sano, 2007). None of these studies included a detailed biochemical characterization of the protein. We have previously shown that the NtROS1 fragment corresponding to the minimal catalytically active AtROS1 fragment has the activity of 5-methylcytosine-DNA glycosylase (Gruber et al., 2018).

Here, in view of the potential value of plant demethylation enzymes as tools for genetic technologies, we characterize the substrate specificity of the recombinant NtROS1 catalytic fragment on mC and hmC in different contexts of methylated CpG dinucleotides and show that the expression of NtROS1 in human cells causes a global decrease in DNA methylation.

Materials and methods

ProtoScript II reverse transcriptase, Q5 Hot Start High-Fidelity DNA polymerase, *Escherichia coli* uracil DNA glycosylase, *Cla*I and *Sac*I restriction endonucleases were purchased from New England Biolabs (USA), and bacteriophage T4 polynucleotide kinase, from Biosan (Novosibirsk, Russia). Oligonucleotides listed in the Table were synthesized at the SB RAS ICBFM Laboratory of Biomedical Chemistry using commercially available prosphoramidites (Glen Research, USA). If necessary, the oligonucleotides were ^{32}P-labeled at the 5′-end using γ[^{32}P]ATP (SB RAS ICBFM Laboratory of Biotechnology) and T4 polynucleotide kinase.

To build a model of the NtROS1 catalytic domain in the Swiss-Model program (Waterhouse et al., 2018), the AtDME (AF-Q8LK56-F1-model_v1) and AtROS1 (AF-Q9SJQ6-F1-model_v1) templates from the AlphaFold database (Jumper et al., 2021) were used.

To obtain a catalytically inactive NtROS1 with the Asp1359Asn substitution, a pLATE31 plasmid with an insert encoding the catalytically active NtROS1 fragment (amino acid residues 754–1796) (Gruber et al., 2018) was mutagenized using primers D1359Nfwd and D1359Nrev (see the Table) and the Q5 Site-Directed Mutagenesis Kit (New England Biolabs). The mutation was confirmed by Sanger sequencing. Wild-type NtROS1 and NtROS1 D1359N were overproduced and purified as described previously (Gruber et al., 2018).

To study the activity of NtROS1, double-stranded substrates were obtained by annealing oligonucleotides C1, C2, M1, M2, H1, and H2 (see the Table). The reaction mixture contained 50 nM substrate, 50 mM Tris–HCl (pH 8.0), 1 mM EDTA, 1 mM DTT, 0.1 % bovine serum albumin, and 100 nM NtROS1. The mixture was incubated at 37 °C; aliquots were withdrawn at various times (2–300 min) and mixed with an equal volume of the stop solution (80 % formamide, 20 mM EDTA, 0.1 % xylene cyanol, 0.1 % bromophenol blue). If necessary, the aliquots were preheated for 2 min at 95 °C in the presence of 0.1 M NaOH and neutralized with an equimolar amount of HCl. The reaction products were resolved by electrophoresis in a 20 % polyacrylamide gel containing 7.2 M urea, visualized by prosphorimaging using the Typhoon FLA 9500 system (GE Healthcare, USA), and quantified using the Quantity One v4.6.3 software (Bio-Rad Laboratories, USA). The apparent rate constants were determined using the SigmaPlot v11.0 software (Systat Software, USA) by nonlinear regression to the equation $[P] = [P]_{max}(1-e^{-kt})$, where [P] is product concentration, $[P]_{max}$ is the maximum product concentration, k is the reaction rate constant, and t is time.

To assess the status of global DNA methylation upon NtROS1 expression in human cells, wild-type and D1359N NtROS1 coding sequences were cloned into the pIRES-eGFP-puro plasmid (Clontech, USA) at the *Sac*I and *Cla*I restriction

Oligonucleotides used in this work

ID	Sequence (5′→3′)
	Reverse transcription PCR
RTPCRfwd	AGAAGGAGATATAACTATGTCATTCATTAGAAGACGGAAACCG
RTPCRrev	GTGGTGGTGATGGTGATGGCCGTTTTCATCTGGCTTTCCTTTAGTCC
	Site-directed mutagenesis
D1359Nfwd	CCTGTCAACACAAACGTTGGC
D1359Nrev	GAAAGCAAGGTGGTGAAGTGT
	Enzyme activity and specificity studies
C1	GCTTGTACTTTAGCGCATTGATTCTCACCACG
C2	CGTGGTGAGAATCAATGCGCTAAAGTACAAGC
M1	GCTTGTACTTTAGMGCATTGATTCTCACCACG (M = mC)
M2	CGTGGTGAGAATCAATGMGCTAAAGTACAAGC (M = mC)
H1	GCTTGTACTTTAGHGCATTGATTCTCACCACG (H = hmC)
H2	CGTGGTGAGAATCAATGHGCTAAAGTACAAGC (H = hmC)

sites. HEK293 Phoenix cells ($1.2 \cdot 10^6$) were transfected with 5 μg of the plasmid by the calcium phosphate method and grown in a monolayer in DMEM with 10 % fetal calf serum (HyClone, USA). After 24 and 48 h, the medium was changed with the addition of 3 μg/ml puromycin. Transfection efficiency was determined by flow cytometry (NovoCyte 3000, ACEA Biosciences, USA) by detection of the fluorescence of eGFP encoded by the same plasmid. NtROS1 expression in the transfected cells was confirmed by reverse transcription PCR using β-actin mRNA as a control. Genomic DNA was isolated from the cells ($5 \cdot 10^6$) using a QIAamp DNA Mini Kit (Qiagen, the Netherlands), and the relative content of mC was determined using anti-mC antibodies (MethylFlash Methylated DNA Quantification Kit, EpiGentek, USA). The results were compared using Student's *t*-test.

Results and discussion

Plant mC-specific DNA glycosylases are proteins of considerable size: for example, AtDME and AtROS1, as well as their DML2 and DML3 paralogs, are over 1000 amino acid residues long (Fig. 1). The extended N-terminal regions of these polypeptides are unstructured, although some of their parts are necessary for enzyme activity. The C-terminal regions contain a conserved HhH catalytic domain and an iron-sulfur cluster (FeS cluster), characteristic of many DNA glycosylases that recognize oxidative DNA damage, as well as an RNA-binding motif (RNA Recognition Motif, RRM) and a CXXC type permuted zinc finger unique to the DME/ROS1 family (see Fig. 1, *a*). Unlike in all other HhH superfamily DNA glycosylases, the catalytic domain in DME/ROS1 proteins is disrupted by a long non-conserved insert (Ponferrada-Marín et al., 2011). Based on the literature data on the AtROS1 protein, we previously cloned the NtROS1 cDNA fragment encoding amino acid residues 754–1796 (Gruber et al., 2018). This region contains all elements necessary for the catalytic activity in AtROS1 (Hong et al., 2014).

Since the structures of the DME/ROS1 family proteins are currently unknown, for a more detailed understanding of the organization of the NtROS1 catalytic fragment, we carried out homology modeling based on the AtDME and AtROS1 models from the AlphaFold template collection (Jumper et al., 2021). The two resulting models were almost identical except for the structure of the long non-homologous regions. The disrupted HhH domain folded into the α-helical structure characteristic of DNA glycosylases of this superfamily, in which a DNA binding groove with the catalytic residues Lys1341 and Asp1359 is evident (see Fig. 1, *b*). In addition, several more peripheral α-helices buttress the HhH domain and the FeS cluster and are obviously important for maintaining their structure. The FeS cluster, zinc finger, and RRM motif form separate structural elements that leave free access to the DNA binding groove (see Fig. 1, *b*). All disordered regions of the structure are also located on the side of the protein globule opposite to the DNA binding groove.

To analyze the catalytic activity and substrate specificity of NtROS1, we performed a cleavage reaction of double-stranded oligonucleotides containing a CpG dinucleotide in which the cytosine was unmethylated, methylated, or hydroxymethylated in one or both chains (Fig. 2). The strand to be cleaved was ^{32}P-labeled at the 5′-end. NtROS1 showed virtually no activity on the substrate containing an unmethylated CpG site, which is consistent with the literature data claiming that the C5 position of the cytosine must carry a substitution to be cleaved by AtROS1 (Morales-Ruiz et al., 2006). Activity towards mC and hmC was observed; however, its level differed markedly for both types of substrates. Comparing the efficiency of cleavage of M1/C2 and M1/M2 substrates (see Fig. 2, lanes *5* and *8*) with H1/C2 and H1/H2 substrates (lanes *11* and *14*), one can see that the enzyme prefers to excise mC over hmC both from CpG sites modified at only one strand and from fully modified sites. Small differences in the cleavage of M1/C2, M1/M2 and M1/H2 substrates (lanes *5*, *8*, and *17*) indicate that modifications in the complementary strand have little effect on the removal of mC, and hmC in the complementary strand might even increase the cleavage. The enzyme showed no activity against DNA substrates containing uracil

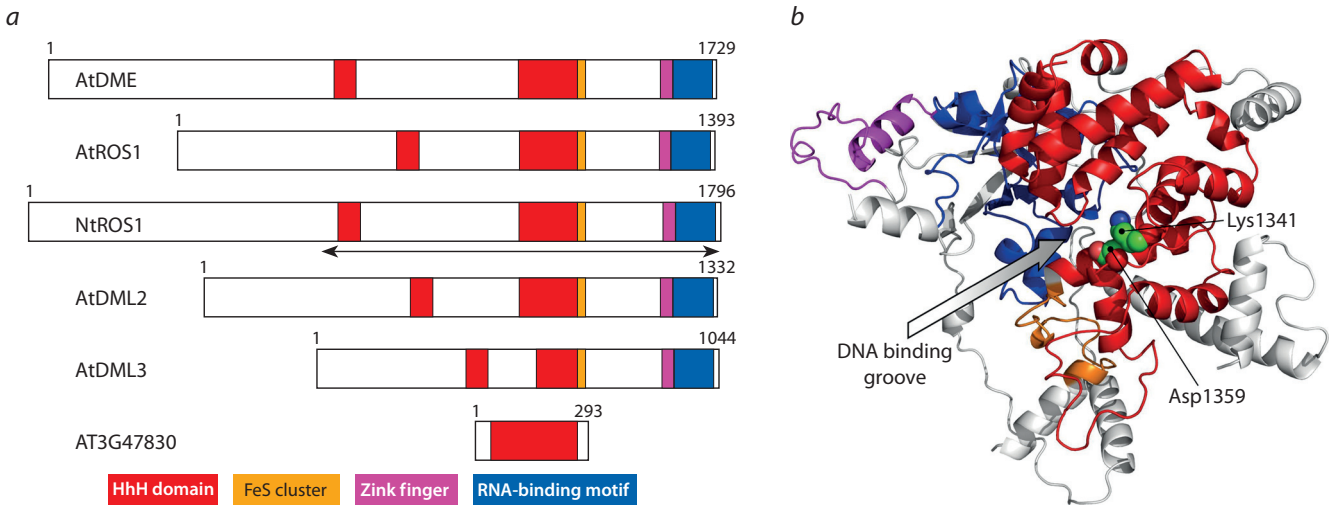

Fig. 1. Structure of NtROS1. *a*, Organization of DME/ROS1 family proteins. Arrows mark the catalytic fragment of NtROS1. *b*, Model of the NtROS1 catalytic fragment (long disordered regions are not shown).

The HhH domain is shown in red in both figures, the FeS cluster is in orange, the CXXC type zinc finger is in purple, and the RRM RNA-binding motif is in blue.

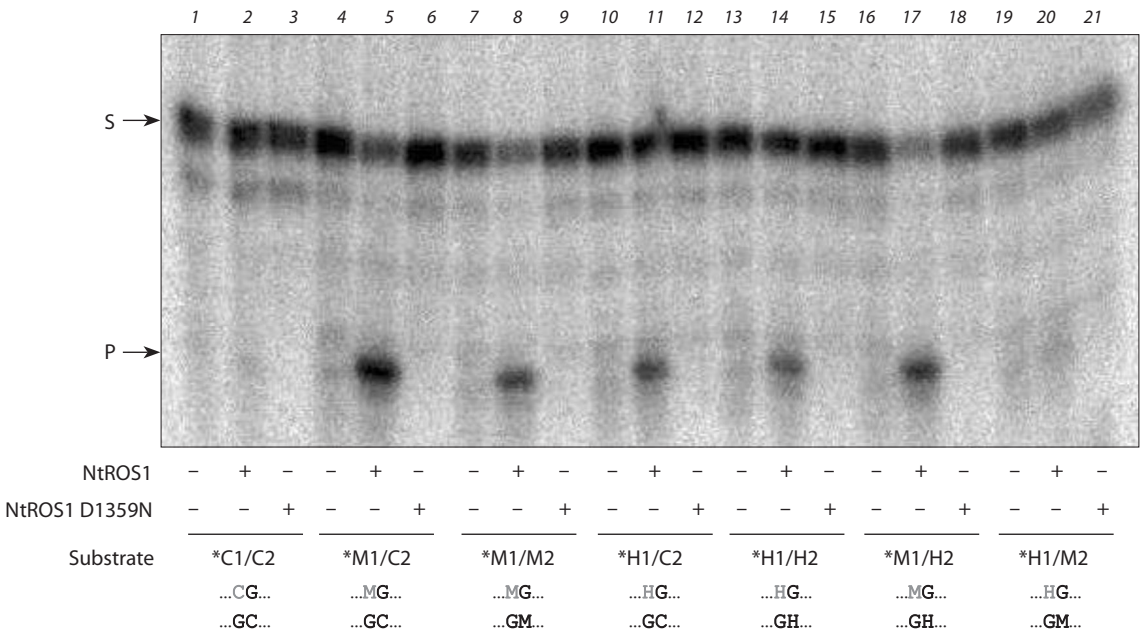

Fig. 2. Cleavage of substrates (50 nM) with wild-type NtROS1 and NtROS1 D1359N (100 nM) for 60 min.

The methylation context of the CpG site in each substrate is indicated below the gel image. Lanes *1*, *4*, *7*, *10*, *13*, *16*, and *19*, substrates without the enzyme; lanes *2*, *5*, *8*, *11*, *14*, *17*, and *20*, wild-type NtROS1; lanes *3*, *6*, *9*, *12*, *15*, *18*, and *21*, NtROS1 D1359N. Arrows mark the substrate (S) and the cleavage product (P). In the schematic representations of substrates below the gel, the excised base is shown in gray.

or 8-oxoguanine. The NtROS1 D1359N mutant, as expected, exhibited no glycosylase activity against any substrate with various combinations of mC and hmC, which confirms the importance of the Asp residue in the ROS1 catalytic center for the removal of modified bases from DNA.

AtROS1 was reported to be an enzyme with a very low turnover number (Ponferrada-Marín et al., 2009), so it was not feasible to use steady-state kinetics to characterize the activity of NtROS1. To determine the apparent reaction rate constant, we used the conditions close to the single-turnover kinetic regime ($[E]_0 > [S]_0$). Under these conditions, all DNA substrate rapidly binds to the enzyme, and the reaction rate is limited not by substrate binding or product release, but by the chemical step of the pseudo-first order reaction, the hydrolysis of the *N*-glycosidic bond of the modified nucleotide (Porello et al., 1998). Thus, following the accumulation of the product over time (Fig. 3, *a*), one can estimate the reaction rate constant. The time courses for all substrates are shown in Fig. 3, *b*, and the rate constants calculated from these data are summarized below:

Substrate	$k \cdot 10^4$, s^{-1}
*M1/C2	6.5±1.1
*M1/M2	2.4±0.5
*M1/H2	3.0±0.7
*H1/C2	1.9±0.3
*H1/M2	1.6±0.6
*H1/H2	2.0±0.3
*C1/C2	Not cleaved

* ^{32}P-labeled strand.

In all cases, hmC was a worse substrate for NtROS1 compared to mC. Based on these results, we conclude that NtROS1 cleaves hemimethylated CpG sites most efficiently, while hmC in the context of HG/GM, on the contrary, is the worst substrate for this enzyme. The kinetic constants are consistent with the qualitative data on the relative cleavage efficiency for different substrates (see Fig. 2).

Many slow-turnover DNA glycosylases have lower AP lyase activity in comparison with their DNA glycosylase activity. In this case, after the removal of the modified base, a part of the reaction product exists as an AP site for a long time, and the true amount of the product can be revealed only upon treatment with alkali or nucleophilic amines (Porello et al., 1998). However, in the case of NtROS1, additional treatment with NaOH did not lead to a noticeable increase in the accumulation of the reaction product (see Fig. 3, *c*). Apparently, the rate of the reaction catalyzed by NtROS1 is limited by the hydrolysis of the *N*-glycosidic bond, which was also suggested for the enzyme from *A. thaliana* (Hong et al., 2014).

To assess the ability of the NtROS1 protein to act as a demethylase when expressed in mammalian cells, we constructed plasmids based on the pIRES-eGFP-puro vector encoding wild-type NtROS1 and its catalytically inactive mutant NtROS1 D1359N. Using anti-mC antibodies, we have estimated the level of this epigenetic base in HEK293 cells after transfection with these plasmids. When cells were observed post-transfection, the proliferation of cells with the pIRES-eGFP-puro-NtROS1 plasmid was reduced by about 30 % compared to the control cells transfected with the plasmid with no insert and the cells transfected with the plasmid encoding the catalytic mutant. The fraction of live cells was the same in all cases, which indicates that the cell cycle may be slower in the presence of active NtROS1 due to the need to repair a large number of breaks introduced into DNA at mC residues. An analysis of the mC level revealed a ~2-fold decrease relative to control samples when wild-type NtROS1

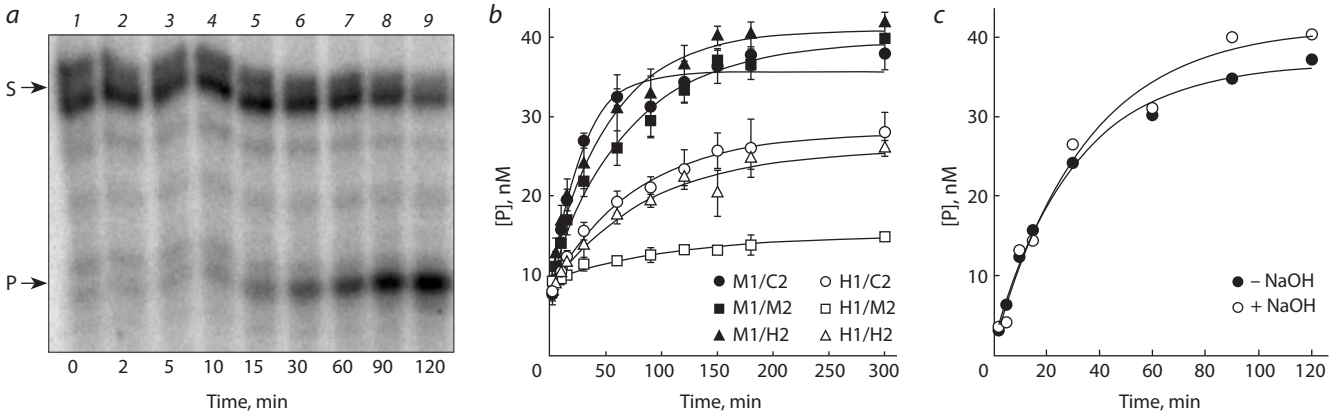

Fig. 3. Time course of the hydrolysis of various substrates by NtROS1.

a, Representative gel showing the accumulation of the reaction products after cleavage of the M1/M2 substrate for 0–120 min; *b*, cleavage of the M1/C2, M1/M2, M1/H2, H1/C2, H1/M2 and H1/H2 substrates (mean ± S.D. for 3–4 experiments are shown); *c*, cleavage of the M1/M2 substrate without and with NaOH treatment for the complete elimination of AP sites.

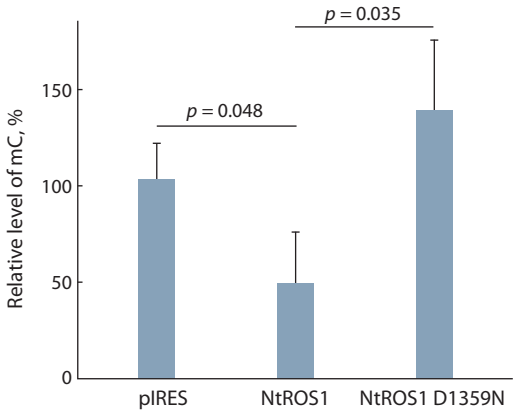

Fig. 4. Relative amounts of mC in the DNA of HEK293 Phoenix cells transfected with the pIRES-eGFP-puro plasmid carrying no insert or carrying an insert encoding the NtROS1 catalytic fragment or its inactive variant NtROS1 D1359N.

Mean ± S.D. of 3 experiments are shown.

was expressed and the absence of statistically significant changes with the expression of NtROS1 D1359N (Fig. 4). In general, it can be considered that the transient expression of the catalytic domain of 5-methylcytosine–DNA glycosylase NtROS1 in human cells indeed leads to the global erasure of mC epigenetic marks. The amount of hmC in cells could not be measured using anti-hmC antibodies, probably because this modified base is about an order of magnitude less abundant than mC (Yu et al., 2012; Zahid et al., 2016).

Thus, it can be concluded that NtROS1 is a DNA glycosylase specific for 5-methylcytosine and, to a lesser extent, for 5-hydroxymethylcytosine. Its biological functions, as in the case of AtROS1, most likely consist in the regulation of methylation status and gene expression during embryonic development (Yamamuro et al., 2014) or in the response to infections and abiotic stress (Gong et al., 2002; Le et al., 2014), and the regulation of silencing of transgenes and transposable elements (Gong et al., 2002; Kapoor et al., 2005; Zhu et al., 2007). The mechanism of RNA-dependent addressing of DME/ROS1 family proteins to specific demethylation sites, which has not yet been elucidated, is of great interest. The RRM motif in these polypeptides is homologous to the motifs responsible for nonspecific interactions with RNA in many other proteins (Cléry et al., 2008) and, by its position in the structure (see Fig. 1, *b*), could bind small RNAs complementary to the DNA stretch 5′ of the targeted mC. Zinc fingers of the CXXC type are used by many mC-recognizing proteins; however, they predominantly bind unmethylated DNA and are presumably required for accurate positioning of the enzyme in the presence of several methylation sites located at a short distance (Iyer et al., 2011).

The prospects for using NtROS1 and other proteins of the DME/ROS1 family as tools for genetic technologies largely depend on the possibility of reducing the size of the catalytic fragment. A deletion of the long insert between the two parts of the HhH domain in AtROS1 fully preserves its activity, but a deletion of the C-terminal tail after the FeS cluster results in an inactive enzyme (Hong et al., 2014). Judging from the structural models of AtROS1 and NtROS1, the HhH and RRM domains interact with each other, and shortening the protein here may only be done by trimming the insert between them. In any case, DME/ROS1 proteins, including NtROS1, represent a promising scaffold for enzyme engineering to analyze epigenetic methylation status and control gene activity.

Conclusion

The study of the demethylating DNA glycosylase ROS1 from *Nicotiana tabacum* reported in this work presents the only biochemical investigation of ROS1 beyond its homologue from *Arabidopsis thaliana*. In addition to 5-methylcytosine, NtROS1 showed the ability to remove 5-hydroxymethylcytosine from DNA, but the efficiency of this reaction was lower than for 5-methylcytosine, apparently because plants rarely if at all use 5-hydroxymethylcytosine as an epigenetic marker. The observed decrease in global methylation upon expression of NtROS1 in human cells suggests that this protein or its optimized variants can be used as a tool for epigenetic regulation, either on its own or as an active module in constructs targeted to certain genome regions.

References

Agius F., Kapoor A., Zhu J.-K. Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl. Acad. Sci. USA.* 2006;103(31):11796-11801. DOI 10.1073/pnas.0603563103.

Ballestar E., Wolffe A.P. Methyl-CpG-binding proteins. Targeting specific gene repression. *Eur. J. Biochem.* 2001;268(1):1-6. DOI 10.1046/j.1432-1327.2001.01869.x.

Baubec T., Ivánek R., Lienert F., Schübeler D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell.* 2013;153(2):480-492. DOI 10.1016/j.cell.2013.03.011.

Bochtler M., Kolano A., Xu G.-L. DNA demethylation pathways: additional players and regulators. *Bioessays.* 2017;39(1):1-13. DOI 10.1002/bies.201600178.

Branco M.R., Ficz G., Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* 2011;13(1):7-13. DOI 10.1038/nrg3080.

Choi C.-S., Sano H. Identification of tobacco genes encoding proteins possessing removal activity of 5-methylcytosines from intact tobacco DNA. *Plant Biotechnol.* 2007;24(3):339-344. DOI 10.5511/plantbiotechnology.24.339.

Choi W.L., Mok Y.G., Huh J.H. Application of 5-methylcytosine DNA glycosylase to the quantitative analysis of DNA methylation. *Int. J. Mol. Sci.* 2021;22(3):1072. DOI 10.3390/ijms22031072.

Choi Y., Gehring M., Johnson L., Hannon M., Harada J.J., Goldberg R.B., Jacobsen S.E., Fischer R.L. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell.* 2002;110(1):33-42. DOI 10.1016/s0092-8674(02)00807-3.

Cléry A., Blatter M., Allain F.H.-T. RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* 2008;18(3):290-298. DOI 10.1016/j.sbi.2008.04.002.

Devesa-Guerra I., Morales-Ruiz T., Pérez-Roldán J., Parrilla-Doblas J.T., Dorado-León M., García-Ortiz M.V., Ariza R.R., Roldán-Arjona T. DNA methylation editing by CRISPR-guided excision of 5-methylcytosine. *J. Mol. Biol.* 2020;432(7):2204-2216. DOI 10.1016/j.jmb.2020.02.007.

Fedorova O.S., Kuznetsov N.A., Koval V.V., Knorre D.G. Conformational dynamics and pre-steady-state kinetics of DNA glycosylases. *Biochemistry* (*Moscow*). 2010;75(10):1225-1239. DOI 10.1134/S0006297910100044.

Gong Z., Morales-Ruiz T., Ariza R.R., Roldán-Arjona T., David L., Zhu J.-K. *ROS1*, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell.* 2002;111(6):803-814. DOI 10.1016/s0092-8674(02)01133-9.

Gruber D.R., Toner J.J., Miears H.L., Shernyukov A.V., Kiryutin A.S., Lomzov A.A., Endutkin A.V., Grin I.R., Petrova D.V., Kupryushkin M.S., Yurkovskaya A.V., Johnson E., Okon M., Bagryanskaya E.G., Zharkov D.O., Smirnov S.L. Oxidative damage to epigenetically methylated sites affects DNA stability, dynamics, and enzymatic demethylation. *Nucleic Acids Res.* 2018;46(20):10827-10839. DOI 10.1093/nar/gky893.

Hong S., Hashimoto H., Kow Y.W., Zhang X., Cheng X. The carboxy-terminal domain of ROS1 is essential for 5-methylcytosine DNA glycosylase activity. *J. Mol. Biol.* 2014;426(22):3703-3712. DOI 10.1016/j.jmb.2014.09.010.

Iyer L.M., Abhiman S., Aravind L. Natural history of eukaryotic DNA methylation systems. *Prog. Mol. Biol. Transl. Sci.* 2011;101:25-104. DOI 10.1016/B978-0-12-387685-0.00002-0.

Jang H., Shin H., Eichman B.F., Huh J.H. Excision of 5-hydroxymethylcytosine by DEMETER family DNA glycosylases. *Biochem. Biophys. Res. Commun.* 2014;446(4):1067-1072. DOI 10.1016/j.bbrc.2014.03.060.

Jumper J., Evans R., Pritzel A., Green T., Figurnov M., Ronneberger O., Tunyasuvunakool K., Bates R., Žídek A., Potapenko A., Bridgland A., Meyer C., Kohl S.A.A., Ballard A.J., Cowie A., Romera-Paredes B., Nikolov S., Jain R., Adler J., Back T., Petersen S., Reiman D., Clancy E., Zielinski M., Steinegger M., Pacholska M., Berghammer T., Bodenstein S., Silver D., Vinyals O., Senior A.W., Kavukcuoglu K., Kohli P., Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583-589. DOI 10.1038/s41586-021-03819-2.

Kapazoglou A., Drosou V., Argiriou A., Tsaftaris A.S. The study of a barley epigenetic regulator, *HvDME*, in seed development and under drought. *BMC Plant Biol.* 2013;13:172. DOI 10.1186/1471-2229-13-172.

Kapoor A., Agarwal M., Andreucci A., Zheng X., Gong Z., Hasegawa P.M., Bressan R.A., Zhu J.-K. Mutations in a conserved replication protein suppress transcriptional gene silencing in a DNA-methylation-independent manner in *Arabidopsis*. *Curr. Biol.* 2005;15(21):1912-1918. DOI 10.1016/j.cub.2005.09.013.

Le T.-N., Schumann U., Smith N.A., Tiwari S., Au P.C.K., Zhu Q.-H., Taylor J.M., Kazan K., Llewellyn D.J., Zhang R., Dennis E.S., Wang M.-B. DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in *Arabidopsis*. *Genome Biol.* 2014;15(9):458. DOI 10.1186/s13059-014-0458-3.

Lee T.-F., Zhai J., Meyers B.C. Conservation and divergence in eukaryotic DNA methylation. *Proc. Natl. Acad. Sci. USA.* 2010;107(20):9027-9028. DOI 10.1073/pnas.1005440107.

Li X., Qian W., Zhao Y., Wang C., Shen J., Zhu J.-K., Gong Z. Antisilencing role of the RNA-directed DNA methylation pathway and a histone acetyltransferase in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA.* 2012;109(28):11425-11430. DOI 10.1073/pnas.1208557109.

Li Y., Kumar S., Qian W. Active DNA demethylation: mechanism and role in plant development. *Plant Cell Rep.* 2018;37(1):77-85. DOI 10.1007/s00299-017-2215-z.

Liu R., How-Kit A., Stammitti L., Teyssier E., Rolin D., Mortain-Bertrand A., Halle S., Liu M., Kong J., Wu C., Degraeve-Guibault C., Chapman N.H., Maucourt M., Hodgman T.C., Tost J., Bouzayen M., Hong Y., Seymour G.B., Giovannoni J.J., Gallusci P. A DEMETER-like DNA demethylase governs tomato fruit ripening. *Proc. Natl. Acad. Sci. USA.* 2015;112(34):10804-10809. DOI 10.1073/pnas.1503362112.

Morales-Ruiz T., Ortega-Galisteo A.P., Ponferrada-Marín M.I., Martínez-Macías M.I., Ariza R.R., Roldán-Arjona T. DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proc. Natl. Acad. Sci. USA.* 2006;103(18):6853-6858. DOI 10.1073/pnas.0601109103.

Ono A., Yamaguchi K., Fukada-Tanaka S., Terada R., Mitsui T., Iida S. A null mutation of *ROS1a* for DNA demethylation in rice is not transmittable to progeny. *Plant J.* 2012;71(4):564-574. DOI 10.1111/j.1365-313X.2012.05009.x.

Ortega-Galisteo A.P., Morales-Ruiz T., Ariza R.R., Roldán-Arjona T. Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Mol. Biol.* 2008;67(6):671-681. DOI 10.1007/s11103-008-9346-0.

Parrilla-Doblas J.T., Roldán-Arjona T., Ariza R.R., Córdoba-Cañero D. Active DNA demethylation in plants. *Int. J. Mol. Sci.* 2019;20(19):4683. DOI 10.3390/ijms20194683.

Pastor W.A., Aravind L., Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* 2013;14(6):341-356. DOI 10.1038/nrm3589.

Pastor W.A., Pape U.J., Huang Y., Henderson H.R., Lister R., Ko M., McLoughlin E.M., Brudno Y., Mahapatra S., Kapranov P., Tahiliani M., Daley G.Q., Liu X.S., Ecker J.R., Milos P.M., Agarwal S., Rao A. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature.* 2011;473(7347):394-397. DOI 10.1038/nature10102.

Penterman J., Uzawa R., Fischer R.L. Genetic interactions between DNA demethylation and methylation in Arabidopsis. *Plant Physiol.* 2007;145(4):1549-1557. DOI 10.1104/pp.107.107730.

Ponferrada-Marín M.I., Parrilla-Doblas J.T., Roldán-Arjona T., Ariza R.R. A discontinuous DNA glycosylase domain in a family of enzymes that excise 5-methylcytosine. *Nucleic Acids Res.* 2011;39(4):1473-1484. DOI 10.1093/nar/gkq982.

Ponferrada-Marín M.I., Roldán-Arjona T., Ariza R.R. ROS1 5-methylcytosine DNA glycosylase is a slow-turnover catalyst that initiates DNA demethylation in a distributive fashion. *Nucleic Acids Res.* 2009;37(13):4264-4274. DOI 10.1093/nar/gkp390.

Porello S.L., Leyes A.E., David S.S. Single-turnover and pre-steady-state kinetics of the reaction of the adenine glycosylase MutY with mismatch-containing DNA substrates. *Biochemistry.* 1998;37(42): 14756-14764. DOI 10.1021/bi981594+.

Roldán-Arjona T., Ariza R.R., Córdoba-Cañero D. DNA base excision repair in plants: an unfolding story with familiar and novel characters. *Front. Plant Sci.* 2019;10:1055. DOI 10.3389/fpls.2019.01055.

Waterhouse A., Bertoni M., Bienert S., Studer G., Tauriello G., Gumienny R., Heer F.T., de Beer T.A.P., Rempfer C., Bordoli L., Lepore R., Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46(W1):W296-W303. DOI 10.1093/nar/gky427.

Wen S., Wen N., Pang J., Langen G., Brew-Appiah R.A.T., Mejias J.H., Osorio C., Yang M., Gemini R., Moehs C.P., Zemetra R.S., Kogel K.-H., Liu B., Wang X., von Wettstein D., Rustgi S. Structural genes of wheat and barley 5-methylcytosine DNA glycosylases and their potential applications for human health. *Proc. Natl. Acad. Sci. USA.* 2012;109(50):20543-20548. DOI 10.1073/pnas.1217927109.

Wu X., Zhang Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* 2017;18(9):517-534. DOI 10.1038/nrg.2017.33.

Yamamuro C., Miki D., Zheng Z., Ma J., Wang J., Yang Z., Dong J., Zhu J.-K. Overproduction of stomatal lineage cells in *Arabidopsis* mutants defective in active DNA demethylation. *Nat. Commun.* 2014;5:4062. DOI 10.1038/ncomms5062.

Yu M., Hon G.C., Szulwach K.E., Song C.-X., Zhang L., Kim A., Li X., Dai Q., Shen Y., Park B., Min J.-H., Jin P., Ren B., He C. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell.* 2012;149(6):1368-1380. DOI 10.1016/j.cell.2012.04.027.

Zahid O.K., Zhao B.S., He C., Hall A.R. Quantifying mammalian genomic DNA hydroxymethylcytosine content using solid-state nanopores. *Sci. Rep.* 2016;6:29565. DOI 10.1038/srep29565.

Zemach A., Zilberman D. Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr. Biol.* 2010;20(17):R780-R785. DOI 10.1016/j.cub.2010.07.007.

Zharkov D.O. Base excision DNA repair. *Cell. Mol. Life Sci.* 2008; 65(10):1544-1565. DOI 10.1007/s00018-008-7543-2.

Zhu J., Kapoor A., Sridhar V.V., Agius F., Zhu J.-K. The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in *Arabidopsis*. *Curr. Biol.* 2007;17(1):54-59. DOI 10.1016/j.cub.2006.10.059.

ORCID ID

N.V. Permyakova orcid.org/0000-0001-9291-1343
I.R. Grin orcid.org/0000-0002-5685-1248
D.O. Zharkov orcid.org/0000-0001-5013-0194

Acknowledgements. The study was supported by the Russian Foundation for Basic Research and the Government of the Novosibirsk Region (project No. 20-44-540007). The modeling of the structure of NtROS1 was supported by the Russian Ministry of Science and Education (State budget project No. 121031300056-8). DNA sequencing was performed at the SB RAS Genomics Core Facility.

Conflict of interest. The authors declare no conflict of interest.

Received November 23, 2021. Revised February 24, 2022. Accepted April 6, 2022.