

Relative Efficiency of Recognition of 5-Methylcytosine and 5-Hydroxymethylcytosine by Methyl-Dependent DNA Endonuclease *GlaI*

D. V. Petrova^a, M. B. Naumenko^b, D. V. Khantakova^b, I. R. Grin^{a, b}, and D. O. Zharkov^{a, b, 1}

^a*Institute of Chemical Biology and Fundamental Medicine, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia*

^b*Novosibirsk State University, Novosibirsk, 630090 Russia*

Received June 24, 2019; revised July 9, 2019; accepted July 25, 2019

Abstract—Only a limited number of tools are available to study cytosine methylation in DNA. One of the representatives of the recently discovered methyl-dependent DNA endonucleases is an enzyme *GlaI*. It is of great interest for determining the methylation status of eukaryotic genomic DNA due to its ability to cleave only methylated DNA. However, the ability of the *GlaI* endonuclease to recognize oxidized derivatives of 5-methylcytosine (^mC), in particular another epigenetic base, 5-hydroxymethylcytosine (^{hm}C), has not yet been characterized. It is not possible to fully use the potential of *GlaI* in the analysis of methylation due to the notable occurrence of the latter in the DNA of mammals. In this study, the efficiency of cleavage of DNA substrates with various combinations of ^mC and ^{hm}C by methyl-dependent DNA-endonuclease *GlaI* was compared; the kinetic parameters of cleavage reactions for fully methylated and fully hydroxymethylated recognition site were determined. It was shown that in most cases *GlaI* recognized substrates containing ^mC better than substrates containing ^{hm}C in the same positions. The most effective hydrolysis of substrates containing modifications in the sequence 5'-GCGC-3'/3'-CGCG-5' required the presence of ^{hm}C not only in the central but also in the edge positions in both DNA chains as in the case of ^mC.

Keywords: epigenetic methylation, 5-methylcytosine, 5-hydroxymethylcytosine, *GlaI* endonuclease

DOI: 10.1134/S1068162019060323

INTRODUCTION

Endonucleases of restriction are widely used currently in molecular biology. Restriction–modification systems are used by many prokaryotes to protect their own genome from foreign genetic material entering the cell [1, 2]. As a rule, the restriction–modification system includes two enzymes: DNA methyltransferase, which methylates certain recognition sites of the host DNA, thereby protecting it from the action of its own restrictases, and restriction endonuclease that cleaves unmethylated foreign DNA. In particular, type II restriction endonucleases differ from others in that they cleave DNA within the recognition site or in close proximity to it; they function as homodimers and do not require ATP hydrolysis for activity. They recognize palindromic sequences of 4–8 nucleotides and cleave them with the formation of a double-stranded gap.

The endonucleases capable to selectively recognize and cleave DNA containing different methylated bases

are of particular interest for epigenetic studies. Some of them, for example, endonuclease *MspJI*, that cleaves DNA containing 5-methylcytosine (^mC), are typical restriction endonucleases with a characteristic catalytic domain [3]. Others, such as endonuclease *DpnI*, specific for DNA containing *N*⁶-methyladenine, have a unique structure and, in all likelihood, perform the cell functions that are not associated with protection against foreign genetic material [4]. The recently discovered enzyme *GlaI*, isolated from the bacterial isolate GL29 of an unidentified species of the Microbacteriaceae family of the order Actinomycetales, belongs to such atypical endonucleases [5, 6]. *GlaI* endonuclease is capable of cleaving the 5'-G^mC↓GC-3' recognition site containing ^mC, but is not capable of hydrolyzing unmodified DNA; it is now increasingly being used to analyze the methylation status of eukaryotic DNA [6–13]. Although the sequence 5'-G^mCGC-3' is the minimal recognition site of the endonuclease *GlaI*, the most efficient cleavage of DNA occurs at sites containing at least three bases ^mC or ^{hm}C; and the palindromic sequence 5'-G^mCG^mC-3'/3'-^mCG^mCG-5' with four bases ^mC shows the maximal level of cleavage [6].

Abbreviations: ^{hm}C, 5-hydroxymethylcytosine; ^mC, 5-methylcytosine.

¹ Corresponding author: phone: +7 (383) 363-51-87; fax +7 (383) 363-51-53; e-mail: dzharkov@niboch.nsc.ru.

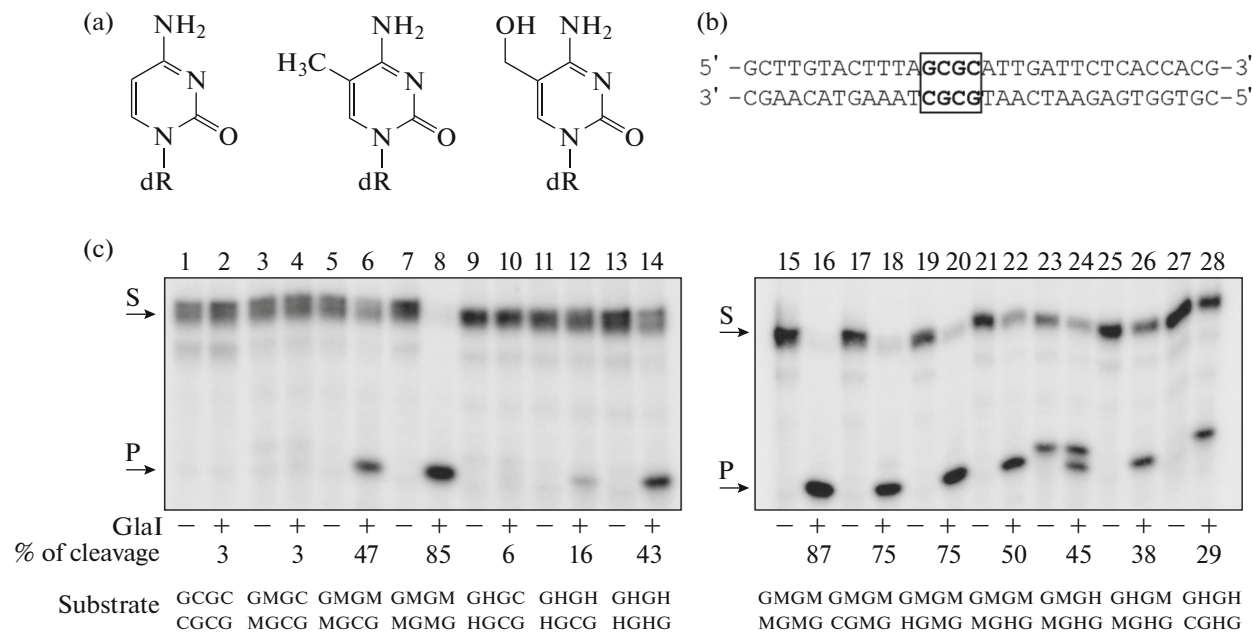


Fig 1. Cleavage of substrates containing various combinations of ^mC and ^{hm}C in the recognition site of endonuclease Glai. (a) The structures of cytosine, 5-methylcytosine and 5-hydroxymethylcytosine. dR is the residue of deoxyribose in DNA. (b) The general sequence of the substrates used in the study. A fragment of the sequence in which the bases C were replaced by ^mC and ^{hm}C in various combinations is highlighted. The radioactive label in all cases was in the upper specified chain. (c), the electropherogram of a representative experiment. The enzyme was present in the tracks with even numbers. The arrows indicate the non-cleaved substrate (S) and the cleavage product (P). In the shown sequences, M is for ^mC and H is for ^{hm}C.

DNA methylation in CpG dinucleotides to form ^mC (Fig. 1a) is a special way of controlling the activity of a wide range of genes, which has maximally developed in higher eukaryotes [14, 15]. The effect of ^mC on gene activity is mainly mediated by proteins containing a methyl binding domain, a portion of the polypeptide chain capable of recognizing a fully methylated CpG dinucleotide in DNA [16]. The proteins containing it (MeCP2, MBD1, MBD2, MBD3, etc.) form complexes with histone deacetylases or possess the activity of histone-specific methyltransferase, chromatin remodeling factors, etc. [17]. These functional interactions lead to condensation of chromatin in the region of methylated CpG-rich segments and suppression of transcription. The level of methylation of promoter regions shows a strong negative correlation with transcriptional activity [18, 19]. Recent studies have shown that the oxidized 5-methylcytosine derivative 5-hydroxymethylcytosine (^{hm}C; Fig. 1a), formed due to the directed oxidation of ^mC in CpG sites by TET1, TET2 and TET3 dioxygenases, also plays an epigenetic role in the mammalian genome [20]. The promoter regions of actively expressed genes are enriched not with ^mC but with ^{hm}C, which is considered as an activating epigenetic marker [21, 22]. Proteins that specifically bind ^{hm}C in DNA were found

[23, 24], but there is no clear understanding of how this base affects transcriptional activity yet.

The discovery of the existence of ^{hm}C and its special role in the epigenome of higher eukaryotes posed a question about the specificity of molecular biological tools used for the analysis of epigenetic methylation. In particular, it is not clear with what efficiency ^mC-specific restriction endonucleases are capable to cleave the DNA containing ^{hm}C recognition sites or various combinations of ^mC and ^{hm}C. In this study, we compared the efficiency of the cleavage of methylated and hydroxymethylated DNA with methyl-dependent DNA endonuclease Glai.

RESULTS AND DISCUSSION

Methyl-dependent DNA endonuclease Glai has recently become popular as an epigenome analysis tool, suitable for the study of cytosine methylation in the DNA of both mammals and plants [6–13]. Although this enzyme is one of the few endonucleases available, which requires ^mC in the 5'-GCGC-3' recognition site, its activity varies quite strongly depending on the total amount of ^mC in this site. According to the literature, the difference between activities of the worst and best substrates, 5'-G^mCGC-3'/3'-CG^mCG-5'

Table 1. Kinetic constants of the Glal endonuclease cleavage reaction at the fully methylated and fully hydroxymethylated 5'-GCGC-3' site

Substrate	K_M , nM	V'_{max} , nM min ⁻¹ *	V'_{max}/K_M , min ⁻¹	Relative specificity
5'-G ^m CG ^m C-3'/3'- ^m CG ^m CG-5'	9.3 ± 5.0	13 ± 1	1.3 ± 0.7	1
5'-G ^{hm} CG ^{hm} C-3'/3'- ^{hm} CG ^{hm} CG-5'	41 ± 16	6.6 ± 0.6	0.16 ± 0.06	0.12

* V'_{max} , the measured V_{max} , normalized per unit of enzyme activity.

and 5'-G^mCG^mC-3'/3'-^mCG^mCG-5', is about 20 times [6]. Glal recognizes with less efficiency the sites other than 5'-GCGC-3' of the general form 5'-RCGY-3' (R is a purine base and Y is a pyrimidine base) [6].

The ability of the Glal enzyme to cleave substrates containing some combinations of C, ^mC and ^{hm}C in an optimal recognition site was investigated (Fig. 1b). The observed efficiency to cleave substrate 5'-G^mC-G^mC-3'/3'-^mCG^mCG-5' (Fig. 1, lanes 7–8 and 15–16) compared to other substrates confirmed that the recognition site containing four ^mC is optimal for hydrolysis by endonuclease Glal. Replacement of all ^mC by ^{hm}C reduced the amount of the cleavage product by two times (Fig. 1, lanes 13–14). Replacement of one ^mC of the complementary chain to C or ^{hm}C in the immediate vicinity of the cleavage site significantly reduced the cleavage efficiency (Fig. 1, lanes 5–6 and 21–22). Further deterioration of cleavage was observed after the replacement of two ^mC to ^{hm}C directly around the hydrolyzed bond (Fig. 1, lanes 25–26). On the other hand, the replacement of one edge ^mC in the recognition site of a complementary chain with an unmethylated nucleotide or its oxidized derivative also affected the cleavage efficiency, but to a lesser extent (Fig. 1, lanes 17–20). An unmodified recognition site, or a site containing no more than 2 modified nucleotides, was hardly cleaved by the enzyme (Fig. 1, lanes 1–4 and 9–10).

The study was conducted to determine the stationary kinetics of cleavage of DNA substrates containing the sites 5'-G^mCG^mC-3'/3'-^mCG^mCG-5' and 5'-G^{hm}C-G^{hm}C-3'/3'-^{hm}CG^{hm}CG-5' and the kinetic parameters K_M and V_{max} of the reaction for a quantitative comparison of efficiency of recognition of ^mC and ^{hm}C (Table 1). It can be seen that the maximal rate for a fully methylated substrate after normalization of V_{max} by the amount of the enzyme was approximately twice as high as for a fully hydroxymethylated substrate. The Michaelis constant for a fully hydroxymethylated sub-

strate was approximately four times higher than for a fully methylated substrate, indicating less stability of the enzyme-substrate complex in the presence of ^{hm}C bases in the site of recognition. The enzyme specificity proportional to the V_{max}/K_M ratio was approximately 8.4 times higher for a fully methylated recognition site compared to a fully hydroxymethylated site.

Thus, it can be concluded from the comparison of the cleavage efficiency of various combinations of C, ^mC and ^{hm}C that the ^{hm}C bases are recognized by the Glal enzyme somewhat worse than the ^mC bases. This may be due to the presence of an additional hydroxyl group of ^{hm}C in the large DNA groove. According to X-ray diffraction analysis, it does not interact with other atoms in the groove [25] and therefore can easily come into contact with donors and acceptors of hydrogen bonds in the protein; in addition, its volume can sterically affect the binding of Glal to DNA. The central position of the sequence 5'-GCGC-3' is of the utmost importance for recognition by the enzyme Glal, while modification of the edge cytosines contributes to the recognition less. It is, in all probability, due to the degeneracy of the recognition site of the enzyme Glal, which cleaves with some effectiveness all the 5'-RCGY-3' sequences containing ^mC [6]. The ability of Glal to recognize ^{hm}C in sequences other than optimal 5'-GCGC-3' may be the subject of further research. In general, it can be noted that the less effective cleavage of ^{hm}C-rich DNA sites by Glal should be taken into account in any analysis of epigenetic status with this enzyme.

EXPERIMENTAL

Materials. Oligodeoxyribonucleotides (Fig. 1b) were synthesized by phosphoramidite solid-phase method on an ASM-800 automatic synthesizer (OOO Biosset, Russia) using commercially available phosphoramidites of ^mC and ^{hm}C (Glen Research, United States). Polynucleotide kinase of bacteriophage T4

was from Biosan (Russia) and DNA endonuclease *GlaI* was from Sibenzyme (Russia).

DNA substrates of endonuclease *GlaI* were obtained by introduction of the ^{32}P -label at the 5'-end of oligonucleotides by standard methods using $\gamma[^{32}\text{P}]\text{-ATP}$ and polynucleotide kinase of bacteriophage T4 [26]. Labeled oligonucleotides were separated from the remaining $\gamma[^{32}\text{P}]\text{-ATP}$ by reversed phase chromatography on the C_{18} NenSorb sorbent (DuPont, United States) and annealed with a double molar excess of the necessary unlabeled complementary chain.

To study the activity of *GlaI* endonuclease, the reaction mixture contained 33 mM Tris acetate (pH 7.9), 10 mM $\text{Mg}(\text{CH}_3\text{COO})_2$, 66 mM CH_3COOK , and 1 mM dithiothreitol (Sibenzyme buffer Y), the substrate at a concentration of 50 nM or 100 nM and 0.35 U/ μL of the enzyme. The mixture was incubated at 30°C for 30 min; the reaction was stopped by adding an equal volume of 80% formamide containing 20 mM sodium ethylenediaminetetraacetate, 0.1% xylene cyanol (w/v), and 0.1% bromophenol blue (w/v), and heated at 95°C for 5 min. The reaction products were analyzed by electrophoresis in 20% polyacrylamide gel containing 8 M urea, followed by radioluminescent scanning with Typhoon FLA 9500 (GE Healthcare, United States). The results were calculated with the Quantity One v 4.6.3 program (BioRad, United States).

To determine the parameters of stationary kinetics of DNA-endonuclease *GlaI* with substrates 5'- $\text{G}^{\text{m}}\text{C}-\text{G}^{\text{m}}\text{C}-3'/3'-\text{mCG}^{\text{m}}\text{CG}-5'$ and 5'- $\text{G}^{\text{hm}}\text{CG}^{\text{hm}}\text{C}-3'/3'-\text{hmCG}^{\text{hm}}\text{CG}-5'$, the reaction mixture contained 3 mM Tris-acetate (pH 7.9), 10 mM $\text{Mg}(\text{CH}_3\text{COO})_2$, 66 mM CH_3COOK , 1mM dithiothreitol, the substrate at a concentration of 10–700 nM and the enzyme at a concentration of 0.02 or 0.05 U/ μL for methylated and hydroxymethylated substrate, respectively. The mixture was incubated at 30°C for 5 min. The reaction was stopped; the products were separated and analyzed as described above.

FUNDING

The work was supported by the Program of the Presidium of the RAS "Fundamental Sciences to Medicine," project no. AAA-A17-117112340087-5.

COMPLIANCE WITH ETHICAL STANDARDS

The work has no studies involving humans or animals as subjects of the study.

Conflict of Interest

The authors state that there is no conflict of interest.

REFERENCES

1. Wilson, G.G. and Murray, N.E., *Ann. Rev. Genet.*, 1991, vol. 25, pp. 585–627.
2. Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S.Kh., Dryden, D.T.F., Dybvig, K., Firman, K., Gromova, E.S., Gumpert, R.I., Halford, S.E., Hattman, S., et al., *Nucleic Acids Res.*, 2003, vol. 31, pp. 1805–1812.
3. Cohen-Karni, D., Xu, D., Apone, L., Fomenkov, A., Sun, Z., Davis, P.J., Morey, Kinney S.R., Yamada-Mabuchi, M., Xu, S.-Y., Davis, T., Pradhan, S., Roberts, R.J., and Zheng, Y., *Proc. Natl. Acad. Sci. U. S. A.*, 2011, vol. 108, pp. 11040–11045.
4. Cantalupo, G., Bucci, C., Salvatore, P., Pagliarulo, C., Roberti, V., Lavitola, A., Bruni, C.B., and Alifano, P., *FEBS Lett.*, 2001, vol. 495, pp. 178–183.
5. Chernukhin, V.A., Nayakshina, T.N., Abdurashitov, M.A., Tomilova, Yu.E., Mezentseva, N.V., Dedkov, V.S., Mikhnenkova, N.A., Gonchar, D.A., and Degtyarev, S.Kh., *Biotekhnologiya*, 2006, no. 4, pp. 31–35.
6. Tarasova, G.V., Nayakshina, T.N., and Degtyarev, S.Kh., *BMC Mol. Biol.*, 2008, vol. 9, p. 7.
7. Abdurashitov, M.A., Chernukhin, V.A., Gonchar, D.A., and Degtyarev, S.Kh., *BMC Genomics*, 2009, vol. 10, p. 322.
8. Kravets, A.P., Myusse, T.A., Litvinchuk, A.V., Ostermiller, Sh., Vengzhen, G.S., and Grodzinskii, D.M., *Tsitol. Genet.*, 2010, vol. 44, pp. 18–22.
9. Wood, R.J., McKelvie, J.C., Maynard-Smith, M.D., and Roach, P.L., *Nucleic Acids Res.*, 2010, vol. 38. e107.
10. Syeda, F., Fagan, R.L., Wean, M., Avvakumov, G.V., Walker, J.R., Xue, S., Dhe-Paganon, S., and Brenner, C., *J. Biol. Chem.*, 2011, vol. 286, pp. 15344–15351.
11. Fagan, R.L., Wu, M., Chedin, F., and Brenner, C., *PLoS One*, 2013, vol. 8. e78752.
12. Rand, K.N., Young, G.P., Ho, T., and Molloy, P.L., *Nucleic Acids Res.*, 2013, vol. 41. e15.
13. Sun, Y., Sun, Y., Tian, W., Liu, C., Gao, K., and Li, Z., *Chem. Sci.*, 2018, vol. 9, pp. 1344–1351.
14. Zemach, A. and Zilberman, D., *Curr. Biol.*, 2010, vol. 20, pp. R780–R785.
15. Lee, T.-F., Zhai, J., and Meyers, B.C., *Proc. Natl. Acad. Sci. U. S. A.*, 2010, vol. 107, pp. 9027–9028.
16. Ballestar, E. and Wolffe, A.P., *Eur. J. Biochem.*, 2001, vol. 268, pp. 1–6.
17. Baubec, T., Ivánek, R., Lienert, F., and Schübeler, D., *Cell*, 2013, vol. 153, pp. 480–492.
18. Weber, M. and Schübeler, D., *Curr. Opin. Cell Biol.*, 2007, vol. 19, pp. 273–280.
19. Singer, M., Kosti, I., Pachter, L., and Mandel-Gutfreund, Y., *Nucleic Acids Res.*, 2015, vol. 43, pp. 3498–3508.
20. Branco, M.R., Ficiz, G., and Reik, W., *Nat. Rev. Genet.*, 2011, vol. 13, pp. 7–13.
21. 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., et al., *Nature*, 2011, vol. 473, pp. 394–397.
22. 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., and He, C., *Cell*, 2012, vol. 149, pp. 1368–1380.
23. Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H.C., Mensinga, A., Brinkman, A.B., Lephikov, K., Muller, U., et al., *Cell*, 2013, vol. 152, pp. 1146–1159.
24. Takai, H., Masuda, K., Sato, T., Sakaguchi, Y., Suzuki, T., Suzuki, T., Koyama-Nasu, R., Nasu-Nishimura, Y., Katou, Y., Ogawa, H., Morishita, Y., Kozuka-Hata, H., Oyama, M., Todo, T., Ino, Y., et al., *Cell Rep.*, 2014, vol. 9, pp. 48–60.
25. Renciuk, D., Blacque, O., Vorlickova, M., and Spingler, B., *Nucleic Acids Res.*, 2013, vol. 41, pp. 9891–9900.
26. Sambrook, J. and Russell, D.W., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 2001.

Translated by E. Puchkov