ISSN 0006-2979, Biochemistry (Moscow), 2025, Vol. 90, No. 4, pp. 513-521 © Pleiades Publishing, Ltd., 2025.
Published in Russian in Biokhimiya, 2025, Vol. 90, No. 4, pp. 571-579.

Restriction–Modification Systems Specific toward GGATC, GATGC, and GATGG. Part 2. Functionality and Structure

Sergey Spirin[1,2,3,a]*, Alexander Grishin[4,5], Ivan Rusinov[1], Andrei Alexeevski[1,3], and Anna Karyagina[1,4,5]

[1]*Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 119234 Moscow, Russia*

[2]*Higher School of Economics National Research University, 109028 Moscow, Russia*

[3]*NRC "Kurchatov Institute" - SRISA, 117218 Moscow, Russia*

[4]*Gamaleya National Research Center for Epidemiology and Microbiology, Ministry of Healthcare of the Russian Federation, 123098 Moscow, Russia*

[5]*All-Russia Research Institute of Agricultural Biotechnology, 127550 Moscow, Russia*

[a]*e-mail: sas@belozersky.msu.ru*

Received January 21, 2025
Revised March 23, 2025
Accepted March 25, 2025

Abstract—The structural and functional basics of protein functionality of restriction–modification systems recognizing GGATC/GATCC, GATGC/GCATC, and GATGG/CCATC sites have been studied using bioinformatics methods. Such systems include a single restriction endonuclease and either two separate DNA methyltransferases or a single fusion DNA methyltransferase with two catalytic domains. It is known that some of these systems methylate both adenines in the recognition sites to 6-methyladenine, but the role of each of the two DNA methyltransferases remained unknown. In this work, we proved the functionality of most known systems. Based on the analysis of structures of related DNA methyltransferases, we hypothesized which of the adenines within the recognition site is modified by each of the DNA methyltransferases and suggested a possible molecular mechanism of changes in the DNA methyltransferase specificity from GATGG to GATGC during horizontal transfer of its gene.

DOI: 10.1134/S0006297925600152

Keywords: restriction–modification system, 3D structure, DNA methyltransferase, restriction endonuclease, horizontal gene transfer

INTRODUCTION

Restriction–modification (RM) systems protect prokaryotic cells from the invasion of foreign (e.g., viral) DNA [1]. RM systems have been traditionally divided into several types [2]. Type II RM systems encode two or more proteins; one is a restriction endonuclease (REase), which specifically recognizes and cleaves a cognate DNA sequence, while the other protein is a methyltransferase (MTase), which modifies the host DNA and prevents its hydrolysis by the REase. Unlike most Type II RM systems, subtype IIA RM systems are characterized by non-palindromic asymmetric recognition sequences, which necessitates the presence of either two different MTases or a single fused MTase with two catalytic centres to methylate both DNA strands [3, 4].

In Part 1 of this work [Evolution and Ecology, Biochemistry (Moscow), vol. 90, issue 4], we described the evolution of subtype IIA RM systems with the specificity towards GGATC/GATCC, GATGC/GCATC, or GATGG/CCATC. The REases of all such systems are homologous to each other, which is also true for their MTases. These MTases methylate adenines

* To whom correspondence should be addressed.

to N^6-methyladenines (6mA) in both DNA strands of the recognition sequence. Some of these systems possess two separate MTases, while others carry a single fused MTase with two catalytic MTase domains. We have shown that the fusion and separation of the two MTases have occurred multiple times during the evolution of such RM systems. While all the MTase domains of these systems belong to the same Methyltransf12 protein family (according to the Pfam database) [5], they can be classified based on the sequence similarity into two well defined groups, which we designated as A and B. In every RM system selected for this study, the two MTase domains always belonged to different groups. Evidently, these two different domains are responsible for the modification of different DNA strands. However, the experimental information on which MTase methylates which DNA strand is lacking. In this article, which is Part 2 of our work, we analyzed available biochemical data, amino acid sequences, and 3D structures of homologous MTases to determine specific nucleotides methylated by each of the MTases, predicted MTase regions responsible for the recognition of cognate DNA sequences, and proposed a hypothetical mechanism that explains changes in the specificity of one of the MTases during its evolution, presumably, in the course of its horizontal transfer from an RM systems with a different specificity.

MATERIALS AND METHODS

The list of RM systems was extracted from REBASE, v. 303 as of 28.02.2023 [6]. The amino acid sequences were aligned using Muscle [7], and the alignments were visualized in Jalview [8]. The boundaries between the N- and C-terminal domains of fused MTases were determined by comparing their sequences with the sequences of single-domain MTases. The evolutionary domains in the sequences of RM system proteins and domain families were identified with HMM profiles from the Pfam database [5]. Protein phylogenies were inferred using FastME [9] with midpoint rooting of the resulting phylogenetic tree. The trees were visualized with MEGA7 [10] and iTOL [11]. CD-HIT [12] was used for clustering proteins with desired sequence identity levels. The structures of MTases and MTase-DNA complexes were predicted with AlphaFold2 using ColabFold [13, 14] and visualized and analyzed in PyMOL [15]. Sequence LOGOs were generated by WebLogo [16].

Following the terminology introduced in the Part 1 of this study [Evolution and Ecology, Biochemistry (Moscow), vol. 90, issue 4], we will further refer to the RM systems specific toward GGATC/GATCC as ‘red’, GATGG/CCATC – as ‘green’, and GATGC/GCATC – as ‘blue’. Hereinafter (except Fig. 4), the recognised sites for ‘red’, ‘green’, and ‘blue’ systems will be designated as GGATC, GATGG, and GATGC, respectively, i.e., we will use the sequence of the chain containing adenine that, according to our results, is methylated by a group A MTase.

RESULTS AND DISCUSSION

In Part 1 of our study [Evolution and Ecology, Biochemistry (Moscow), Issue 4, vol. 90], we determined that all RM systems specific toward GGATC, GATGC, or GATGG contained one REase domain of the RE_AlwI family and two MTase domains of the Methyltransf12 family. In total, REBASE v. 303 contained 493 such systems with two single-domain MTases and 227 systems with fused two-domain MTases. Based on their sequence similarity, all single-domain MTases, as well as the N- and C-terminal domains of fused two-domain MTases, were divided into two groups designated A and B. Two MTases of the same system or two domains of a fused MTase always belonged to the different groups.

Functionality of the RM systems. The enzymatic activity of several REases belonging to the studied RM systems has been demonstrated experimentally. Thus, the ‘blue’ REase SfaNI hydrolyzed DNA five nucleotides downstream of the recognition sequence GCATC in the top strand and nine nucleotides downstream in the bottom strand, which is conventionally designated as GCATC (5/9) [17]. According to REBASE, the ‘red’ REase AlwI hydrolyzes GGATC (4/5). The ‘green’ REase McaCI recognizes the sequence CCATC, although the exact site of hydrolysis remains unknown.

It was shown that mutations of the amino acid residues E418, D456, E469, and E482 in the nickase Nt.BstNBI, which is homologous to the studied REases (Fig. 1), resulted in the loss of its catalytic activity [18]. Amino acid residues corresponding to D456 and E482 of Nt.BstNBI were absolutely conserved in 411 reference sequences chosen from the clusters of studied REases with 98% sequence identity; the residues corresponding to E418 were conserved in 410 sequences and replaced by D in only one sequence. Amino acid residues corresponding to E469 were conserved in 407 sequences, replaced by K in two closely related REases, are deleted in two other sequences.

Based on the X-ray analysis of the structure of Nt.BspD6I nickase, which is 100% identical to Nt.BstNBI, it was proposed that H489 is another residue essential for the nickase enzymatic activity [19]. The corresponding amino acid residue in the sequences of studied REases were strictly conserved. Altogether, these data indicate that the majority of REases studied in this work are functional.

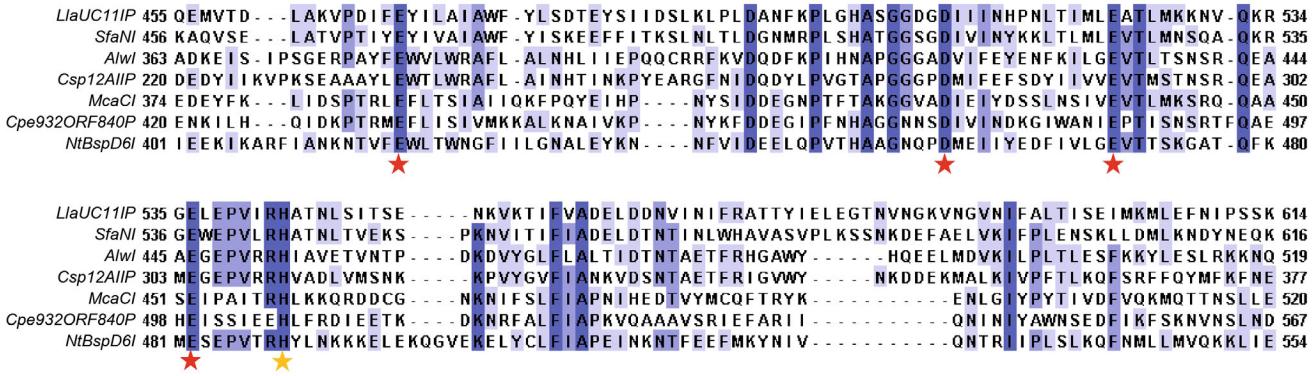

Fig. 1. Fragment of the multiple sequence alignment of the studied REases and Nt.BstNBI nickase. Red asterisks indicate amino acid residues E418, D456, E469, and E482 of nickase Nt.BstNBI that were experimentally confirmed as important for its catalytic activity; yellow asterisk indicates H489 presumably participating in the catalysis.

The MTases studied in our work were 6mA MTases that belong to the α group according to the classification based on the mutual position of the conserved motifs within their sequences [20]. The absolute majority of the studied MTases possessed both the S-adenosylmethionine-binding motif F-x-G-x-[G/A] and the catalytic motif D-[P/T]-P-Y, which indicates the functionality of these enzymes. The ability to methylate DNA has been experimentally demonstrated by PacBio for almost a hundred of these MTases.

Hypothetical mechanism of DNA recognition by MTases and prediction of the methylation sites. The 3D structures are available for three MTases from the MethyltrasfD12 family: M.EcoT4Dam (PDB IDs:1YFJ, 1YFL, 1YF3, 1Q0S, 1Q0T) [21], M1.DpnII (2DPM) [22], and M.EcoKDam (2G1P, 2ORE, 4GOL, 4GOM, 4GON, 4GOO, 4GBE, and 4RTJ-4RTS) [23]. All three MTases are homologous to group B MTases from the studied RM systems (see Fig. S1 in the Online Resource 1). M.EcoT4Dam [21] contained several groups of conserved amino acid residues that formed three clusters on the protein surface (see Fig. 2b in the paper [21]). The first group included residues located near the catalytic site (green asterisks in Fig. S1 in the Online Resource 1), the second and the third groups contained conserved residues from the target-recognition domain (TRD; yellow asterisk in Fig. S1 in the Online Resource 1) and conserved β-hairpin (red asterisks in Fig. S1 in the Online Resource 1), respectively.

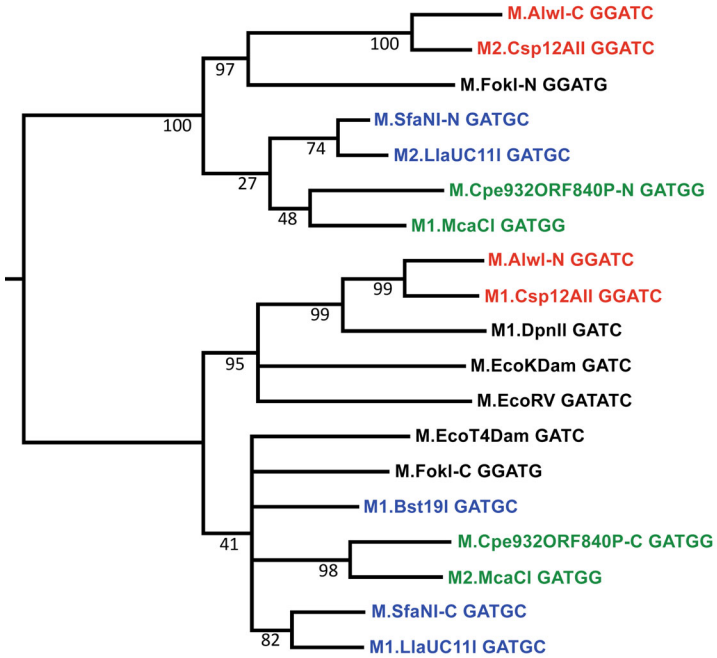

Fig. 2. Phylogenetic tree of 12 MTases specific toward GGATC (red), GATGC (green), and GATGG (blue), MTases M.EcoKDam, M.EcoT4Dam, and M1.DpnII with known 3D structures, and well-studied MTases M.EcoRV, M.FokI, and M1.Bst19I. The phylogeny for the fused two-domain MTases was inferred for the N- and C-terminal domains separately. Numbers indicate the bootstrap support values; branches with lower than 25% support were removed (collapsed).

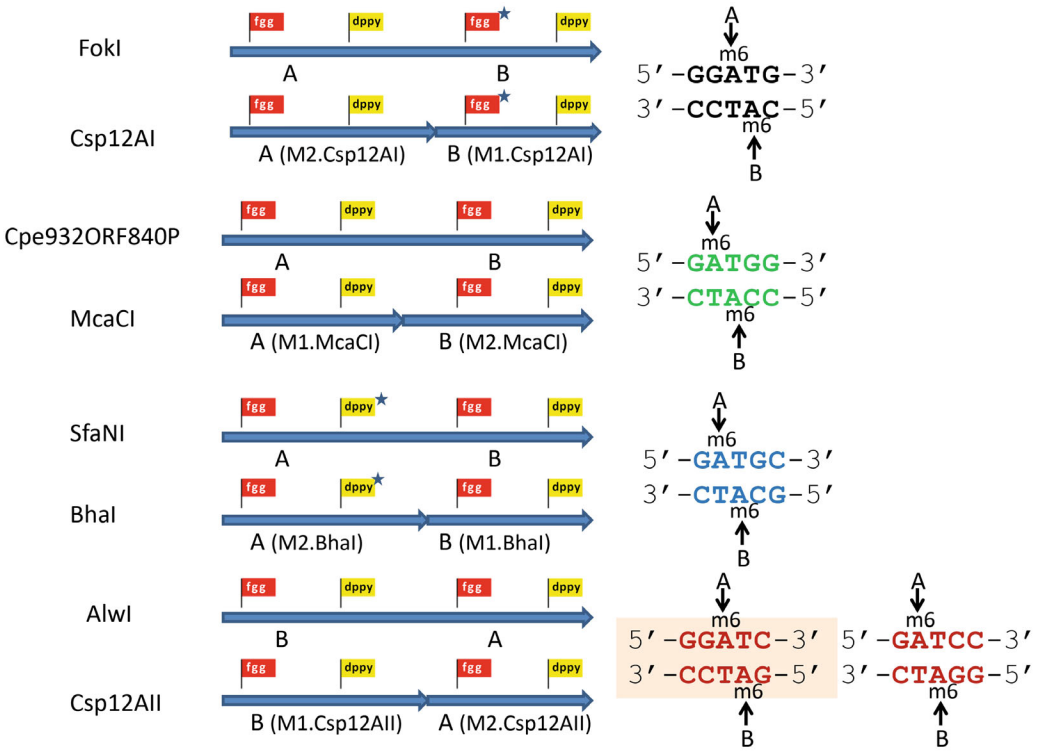

Fig. 3. Gene organization and target sequences of representative MTases recognizing GGATG, GATGG, GATGC, and GGATC sites. Based on the information from REBASE website, red and yellow flags indicate positions of conserved motifs F-x-G-x-G and D-P-P-Y, respectively; blue asterisks indicate non-perfect matches between the canonical motifs and the actual sequences (F-x-G-x-A and D-T-P-Y, respectively). Two possible methylation variants are shown for GGATC, with pink background indicating a more probable variant.

Both DNA strands in the recognition sequences of the studied RM systems contained one adenine residue each. Several 3D structures of MethyltrasfD12 family MTases, namely M.EcoT4Dam (PDB IDs: 1YFJ, 1YFL, 1YF3, and 1Q0T), and M.EcoKDam (PDB ID 2G1P, and 4RTJ-4RTS) were obtained in complex with DNA [21, 23]. Since these MTases were homologous to the group B MTases studied in our work (Fig. 2 and S1 in the Online Resource 1), it was possible to predict which adenine is methylated by which MTase in the RM systems recognizing GATGG (‘green’) and GATGC (‘blue’) sequences. Specifically, multiple sequence alignment of group B MTases with M.EcoKDam and M.EcoT4Dam identified a conserved arginine residue (R116 in M.EcoT4Dam and R124 in M.EcoKDam; the first red asterisk in Fig. S1 in the Online Resource 1). In M.EcoKDam, this residue is responsible for the recognition of the guanine residue in the DNA strand complementary to the strand containing methylated adenine of the GATC recognition sequence [23]. Presumably, this means that group B MTases methylate adenine in the ATC subsequence. This hypothesis was supported by the experimental data on the nucleotides methylated by the ‘green’ MTases M1.Hpy300VI (group A) and M2.Hpy300VI (group B) [24], M.FokI [25] specific towards GGATG (the N- and C-terminal domains of this enzyme are homologous to group A and group B MTases, respectively), and MTase M1.Bst19I which is similar to group B MTases [26]. Although M1.Bst19I recognizes GATGC site, its sequence differs significantly from other MTases with the same specificity. M1.Bst19I did not cluster together with the rest of ‘blue’ MTases in the phylogenetic tree (Fig. 2). The RM system containing M1.Bst19I was not included in the list of RM systems studied in this work, because the amino acid sequence of the corresponding REase Bst19I was not available.

Therefore, we predict that group B MTases belonging to the systems with the GATGG and GATGC recognition sites (‘green’ and ‘blue’, respectively) methylate adenines that are complementary to the third (T) nucleotide. Group A MTases likely methylate the second (A) nucleotide in these sequences. However, no such prediction can be made for the systems recognizing GGATC (red), since in this case, both DNA strands contain the ATC subsequence (Fig. 3).

To determine the most probable methylation pattern for the MTases with GGATC recognition sequence, we analyzed the 3D structures of the MTase complexes with DNA (Fig. 4). The models of all proteins, including M.EcoKDam and M.EcoT4Dam, were generated with ColabFold, since the crystal structures for

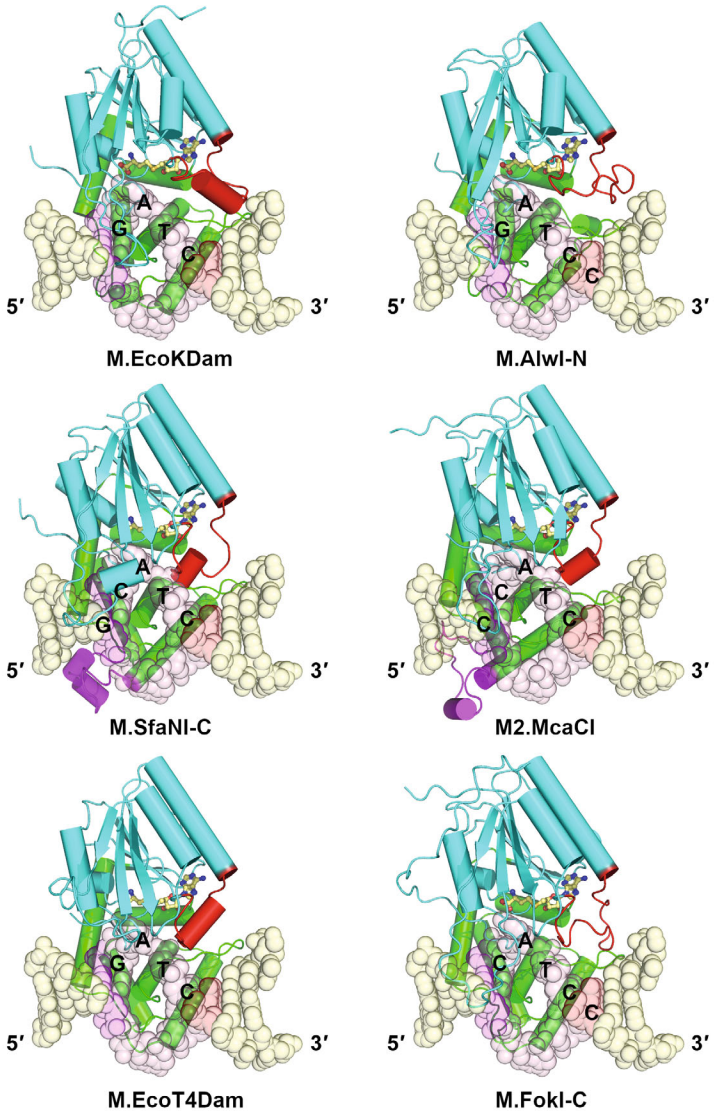

Fig. 4. 3D models of group B MTases and their homologs in complex with DNA. Cyan, MTase catalytic domain; green, TRD; red, loop potentially involved in the recognition of the nucleotide pairs shown to the right in the figure; magenta, region presumably involved in the recognition of nucleotides shown to the left in the figure. In the DNA molecule, pale pink indicates four nucleotides that correspond to the GATC recognition sequence of M.EcoKDam; light lilac and pink correspond to flanking nucleotide pairs. Nucleotides designated with letters and 5′ and 3′ denotations are given for the strands containing adenine residues methylated by MTases B

M.EcoKDam and M.EcoT4Dam lacked coordinates for several loops. For M.EcoKDam and M.EcoT4Dam, the regions present in the crystal structures did not differ significantly from the obtained models (RMSD values for comparison of M.EcoKDam model with its X-ray structure (2G1P) and of M.EcoT4DAM model with its structure (1YFL) were 0.225 and 0.633 Å, respectively). The positions of DNA and S-adenosylhomocysteine were modelled from 2G1P. MTase models were aligned with the 2G1P structure over the catalytic domain (coloured cyan in Fig. 4) to unify their spatial orientation.

We hypothesized that the conserved arginine residue (R124 in M.EcoKDam) of group B MTases always interacts with guanine in the DNA strand complementary to the strand containing the methylated adenine in the ATC motif. If this adenine is a part of the GGATC sequence, then group B MTase (e.g., N-terminal domain of M.AlwI) would have amino acid residues specifically recognizing the two 5′ nucleotides; on the other hand, if this adenine is a part of the complementary GATCC sequence, then the group B MTase would contain residues recognizing both 5′ and 3′ nucleotides. As shown in Fig. 4, the structure of the N-terminal domain of M.AlwI possesses a loop located close to the nucleotides adjacent to the methylated adenine from the 3′ direction (the loop is shown in red in Fig. 4 and underlined with red in the sequence alignment in Fig. S1 in the Online Resource 1). At the same time,

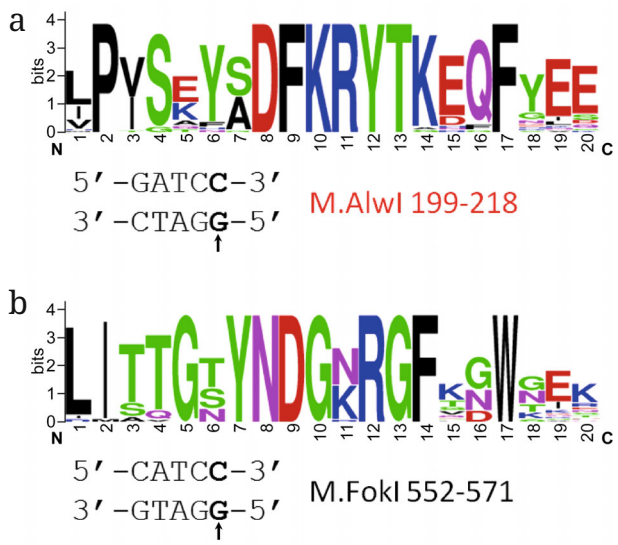

Fig. 5. LOGO diagrams of the sequence alignments of group B MTases loops (shown in red in Fig. 4). a) Diagram constructed for the loops of GGATC-specific MTases, homologous to residues 199-218 of M.AlwI N-terminal domain (VPISEYSDFKRYTKEQFYLE). b) Diagram constructed for the loops of GGATG-specific MTases, homologous to residues 552-571 of M.FokI C-terminal domain (LITTGSYNDGNRGFKDWNRL). The recognition sequences of the respective MTases are shown under each LOGO diagram; the nucleotide pair presumably recognized by the protein loop is highlighted in bold; G residue indicated by the arrow likely binds to the conserved arginine.

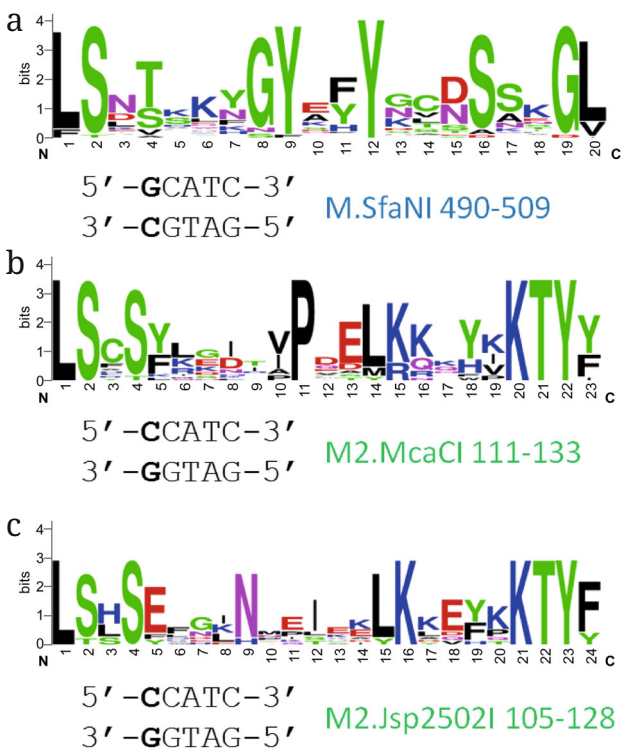

Fig. 6. LOGO diagrams of the sequence alignments of group B MTase loops that are shown in magenta in Fig. 4. a) Diagram constructed for loops of GATGC-specific MTases (‘blue’), homologous to residues 490-509 of M.SfaNI C-terminal domain (LSNSKMYGYNYYKTSSAKGL). b and c) Diagrams constructed for the 23-residue (b) and 24-residue (c) loops of GATGG-specific MTases (‘green’), homologous to residues 111-133 of M2.McaCI (LSCSYLSITVPDELKKKYVKTYY). The recognition sequences of the respective MTases are shown under each LOGO diagram; the nucleotide pair presumably recognized by the protein loop is highlighted in bold.

M.AlwI lacks protein segments coloured in magenta in Fig. 4; these segments are present in the structures of MTases (‘green’ M2.McaCI and C-terminal domain of ‘blue’ M.SfaNI) whose recognition sequences extend toward the 5′ end relative to the methylated adenine (underlined with magenta in the sequence alignment in Fig. S1 in the Online Resource 1). Therefore, it seems more plausible that the N-terminal domain of M.AlwI methylates adenine in GATCC and not in GGATC.

Group B MTases that recognize GGATC (‘red’) or GGATG (M.FokI-like) contain the conserved motif Y-x-D-x-x-R (Fig. 5) potentially involved in the decoding of guanine marked with an arrow in Fig. 5 that could possibly interact with the conserved arginine residue of this motif.

Group B MTases recognizing GATGC and GATGG (‘blue’ and ‘green’, respectively) contain the L-S-x-[S/T] motif (Fig. 6) at the beginning of the loop (coloured in magenta in Fig. 4). The rest of the loop sequences differ between the ‘blue’ and ‘green’ group B MTases, probably because they recognise GC nucleotide pairs oriented differently with respect to the methylated adenines. These GC pairs are probably recognized by some residues from the G-Y-x–x-Y-x-x-x-S motif in ‘blue’ MTases (Fig. 6a) and K-x-x-x-x-K-T-Y-[F/Y] motif in ‘green’ MTases (Fig. 6, b and c). ‘Green’ MTases contain two variants of the loop (23 and 24 amino acid residues in length). The sequences of these variants are similar at the edges but differ in the central part.

Changes in the MTase specificity after horizontal gene transfer. In several RM systems, including Cup11541IV, Hfe11613I, and Hmu12714II whose specificity towards GATGC was confirmed by PacBio, group A MTases was more similar to the MTases recognizing GATGG (‘green’) than to the majority of MTases with the GATGC specificity (‘blue’). Group B MTases and REases from these systems, however, did not demonstrate such anomalous clustering (note the positions of Hmu12714II and Hfe11614I proteins on the trees in Fig. 7, a-c). Presumably, an ancestral group A MTase has changed its specificity from GATGG to GATGC upon the horizontal transfer of its gene. To suggest the molecular mechanisms responsible for this change, we compared the sequences of group A MTases recognizing GATGG and GATGC (Fig. 7). The specificity of these MTases was confirmed by PacBio. All ‘blue’ group A MTases, including those more similar to the ‘green’ group A enzymes in majority of their

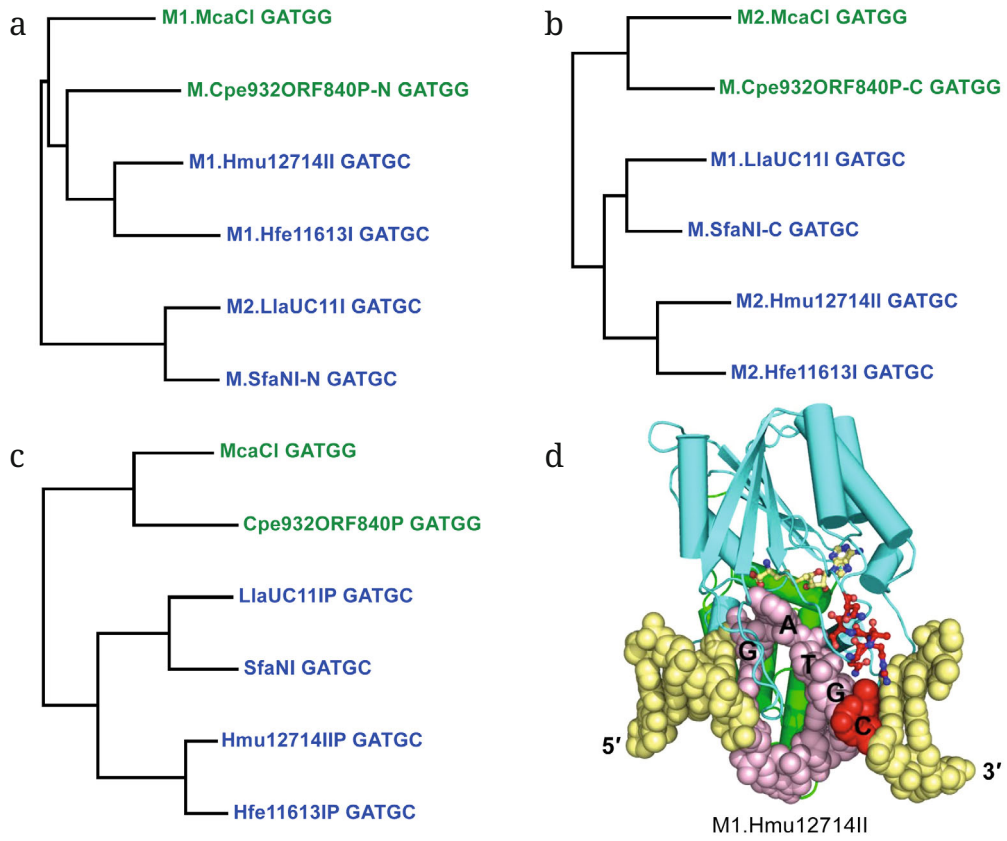

Fig. 7. Phylogenetic trees for the representative group A MTases (a), group B MTases (b), and REases (c) recognizing GATGC ('blue') or GATGG ('green'), and the 3D model of group A MTase M1.Hmu12714II (d). The GC pair corresponding to the 3′-end cytosine in GATGC is shown in red. S-adenosylhomocysteine and side chains of amino acid residues from the N-x-R-S-N motif are shown as ball-and-stick models; the catalytic domain and the TRD indicated with cyan and green, respectively.

amino acid sequences, contained in the C-terminal region the conserved motif N-x-R-S-N, which 'green' MTases lacked. As can be seen from the model of M1.Hmu12714II complex with DNA (Fig. 7d), this motif is positioned in the DNA major groove, with arginine side chain located close to the guanine complementary to the 3′-end C in GATGC. It can be speculated that the arginine residue of this motif is responsible for the enzyme specificity.

The N-x-R-S-N motif could have emerged independently as a result of convergent evolution. More likely, however, that this motif has been obtained by an MTase with the GATGG specificity through recombination with an MTase recognizing GATGC.

CONCLUSION

In this work, we analyzed the sequences and 3D structures of proteins from the RM systems recognizing GGATC, GATGC, and GATGG sequences, as well as several related proteins, to establish the functionality of the majority of these systems and to propose which adenine base is methylated by each of the two MTases in a RM system and how the MTase recognition sequence can change from GATGG to GATGC upon the horizontal transfer of an MTase gene.

Abbreviations. MTase, DNA methyltransferase; REase, restriction endonuclease; RM system, restriction–modification system; TRD, target recognition domain.

Supplementary information. The online version contains supplementary material available at https://doi.org/10.1134/S0006297925600152.

Contributions. S.S. and A.K. developed the concept and supervised the study; S.S., A.G., I.R., and A.K. curated the data, developed the software, and analyzed the data; S.S., A.G., and A.K. wrote the manuscript.

Funding. This work was supported by the Russian Foundation for Basic Research (project no. 21-14-00135).

Ethics approval and consent to participate. This work does not contain any studies involving human or animal subjects.

Conflict of interest. The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. Williams, R. J. (2003) Restriction endonucleases: classification, properties, and applications, *Mol. Biotechnol.*, **23**, 225-244, https://doi.org/10.1385/mb:23:3:225.
2. Roberts, R. J. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes, *Nucleic Acids Res.*, **31**, 1805-1812, https://doi.org/10.1093/nar/gkg274.
3. Madhusoodanan, U. K., and Rao, D. N. (2010) Diversity of DNA methyltransferases that recognize asymmetric target sequences, *Crit. Rev. Biochem. Mol. Biol.*, **45**, 125-145, https://doi.org/10.3109/10409231003628007.
4. Vasu, K., and Nagaraja, V. (2013) Diverse functions of restriction-modification systems in addition to cellular defense, *Microbiol. Mol. Biol. Rev.*, **77**, 53-72, https://doi.org/10.1128/mmbr.00044-12.
5. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., and Bateman, A. (2020) Pfam: the protein families database in 2021, *Nucleic Acids Res.*, **49**, D412-D419, https://doi.org/10.1093/nar/gkaa913.
6. Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2014) REBASE – a database for DNA restriction and modification: enzymes, genes and genomes, *Nucleic Acids Res.*, **43**, D298-D299, https://doi.org/10.1093/nar/gku1046.
7. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.*, **32**, 1792-1797, https://doi.org/10.1093/nar/gkh340.
8. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview Version 2 – a multiple sequence alignment editor and analysis workbench, *Bioinformatics*, **25**, 1189-1191, https://doi.org/10.1093/bioinformatics/btp033.
9. Lefort, V., Desper, R., and Gascuel, O. (2015) FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program, *Mol. Biol. Evol.*, **32**, 2798-2800, https://doi.org/10.1093/molbev/msv150.
10. Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets, *Mol. Biol. Evol.*, **33**, 1870-1874, https://doi.org/10.1093/molbev/msw054.
11. Letunic, I., and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, *Nucleic Acids Res.*, **49**, W293-W296, https://doi.org/10.1093/nar/gkab301.
12. Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, *Bioinformatics*, **22**, 1658-1659, https://doi.org/10.1093/bioinformatics/btl158.
13. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) ColabFold: making protein folding accessible to all, *Nat. Methods*, **19**, 679-682, https://doi.org/10.1038/s41592-022-01488-1.
14. 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., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold, *Nature*, **596**, 583-589, https://doi.org/10.1038/s41586-021-03819-2.
15. DeLano, W. L. (2002) Pymol: An open-source molecular graphics tool, *CCP4 Newsl. Protein Crystallogr*, **40**, 82-92.
16. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: A sequence logo generator, *Genome Res.*, **14**, 1188-1190, https://doi.org/10.1101/gr.849004.
17. Gingeras, T. R., MIlazzo, J. P., and Roberts, R. J. (1978) A computer assisted method for the determination of restriction enzyme recognition sites, *Nucleic Acids Res.*, **5**, 4105-4127, https://doi.org/10.1093/nar/5.11.4105.
18. Higgins, L. S., Besnier, C., and Kong, H. (2001) The nicking endonuclease N.BstNBI is closely related to type IIS restriction endonucleases MlyI and PleI, *Nucleic Acids Res.*, **29**, 2492-2501, https://doi.org/10.1093/nar/29.12.2492.
19. Kachalova, G. S., Rogulin, E. A., Yunusova, A. K., Artyukh, R. I., Perevyazova, T. A., Matvienko, N. I., Zheleznaya, L. A., and Bartunik, H. D. (2008) Structural analysis of the heterodimeric type IIS restriction endonuclease R.BspD6I acting as a complex between a monomeric site-specific nickase and a catalytic subunit, *J. Mol. Biol.*, **384**, 489-502, https://doi.org/10.1016/j.jmb.2008.09.033.
20. Malone, T., Blumenthal, R. M., and Cheng, X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes, *J. Mol. Biol.*, **253**, 618-632, https://doi.org/10.1006/jmbi.1995.0577.
21. Yang, Z., Horton, J. R., Zhou, L., Zhang, X. J., Dong, A., Zhang, X., Schlagman, S. L., Kossykh, V., Hattman, S., and Cheng, X. (2003) Structure of the bacteriophage T4 DNA adenine methyltransferase, *Nat. Struct. Biol.*, **10**, 849-855, https://doi.org/10.1038/nsb973.
22. Horton, J. R., Liebert, K., Hattman, S., Jeltsch, A., and Cheng, X. (2005) Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase, *Cell*, **121**, 349-361, https://doi.org/10.1016/j.cell.2005.02.021.
23. Horton, J. R., Liebert, K., Bekes, M., Jeltsch, A., and Cheng, X. (2006) Structure and substrate recognition

of the *Escherichia coli* DNA adenine methyltransferase, *J. Mol. Biol.*, **358**, 559-570, https://doi.org/10.1016/j.jmb.2006.02.028.

24. Nell, S., Estibariz, I., Krebes, J., Bunk, B., Graham, D. Y., Overmann, J., Song, Y., Spröer, C., Yang, I., Wex, T., Korlach, J., Malfertheiner, P., and Suerbaum, S. (2018) Genome and methylome variation in *Helicobacter pylori* with a *cag* pathogenicity island during early stages of human infection, *Gastroenterology*, **154**, 612-623, https://doi.org/10.1053/j.gastro.2017.10.014.
25. Friedrich, T., Fatemi, M., Gowhar, H., Leismann, O., and Jeltsch, A. (2000) Specificity of DNA binding and methylation by the M.FokI DNA methyltransferase, *Biochim. Biophys. Acta*, **1480**, 145-159, https://doi.org/10.1016/s0167-4838(00)00065-0.
26. Tomilova, J. E., Kuznetsov, V. V., Abdurashitov, M. A., Netesova, N. A., and Degtyarev, S. K. (2010) Recombinant DNA-methyltransferase M1.Bst19I from *Bacillus stearothermophilus* 19: purification, properties, and amino acid sequence analysis, *Mol. Biol.*, **44**, 606-615, https://doi.org/10.1134/S0026893310040163.