Excision of the doubly methylated base \(N^4,5\)-dimethylcytosine from DNA by *Escherichia coli* Nei and Fpg proteins

Marina Alexeeva\(^1,\*\), Prashanna Guragain\(^1,\*\), Almaz N. Tesfahun\(^1,\*\), Miglé Tomkuvienė\(^2,\*\), Aysha Arshad\(^1\), Rūta Gerasimaitė\(^2,\*\), Audronė Rukšenaitė\(^2\), Giedrė Urbanavičiūtė\(^2\), Magnus Björås\(^3,4\), Jon K. Laerdahl\(^4\), Arne Klungland\(^4,5\), Saulius Klimašauskas\(^2\) and Svein Bjelland\(^1\)

\(^1\)Department of Chemistry, Bioscience and Environmental Technology-Centre for Organella Research, Faculty of Science and Technology, University of Stavanger, PO Box 8600 Forus, 4021 Stavanger, Norway

\(^2\)Department of Biological DNA Modification, Institute of Biotechnology, Vilnius University, Vilnius 10257, Lithuania

\(^3\)Institute for Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway

\(^4\)Department of Microbiology, Oslo University Hospital, Rikshospitalet, 0372 Oslo, Norway

\(^5\)Department of Molecular Medicine, Institute of Basic Medical Sciences, University of Oslo, 0317 Oslo, Norway

PG, 0000-0001-6502-9646; SK, 0000-0002-1395-2030

Cytosine (C) in DNA is often modified to 5-methylcytosine (m\(^5\)C) to execute important cellular functions. Despite the significance of m\(^5\)C for epigenetic regulation in mammals, damage to m\(^5\)C has received little attention. For instance, almost no studies exist on erroneous methylation of m\(^5\)C by alkylating agents to doubly or triply methylated bases. Owing to chemical evidence, and because many prokaryotes express methyltransferases able to convert m\(^5\)C into \(N^4,5\)-dimethylcytosine (m\(^4,5\)C) in DNA, m\(^4,5\)C is probably present *in vivo*. We screened a series of glycosylases from prokaryotic to human and found significant DNA incision activity of the *Escherichia coli* Nei and Fpg proteins at m\(^4,5\)C residues *in vitro*. The activity of Nei was highest opposite cognate guanine followed by adenine, thymine (T) and C. Fpg-complemented Nei by exhibiting the highest activity opposite C followed by lower activity opposite T. To our knowledge, this is the first description of a repair enzyme activity at a further methylated m\(^5\)C in DNA, as well as the first alkylated base allocated as a Nei or Fpg substrate. Based on our observed high sensitivity to nuclease S1 digestion, we suggest that m\(^4,5\)C occurs as a disturbing lesion in DNA and that Nei may serve as a major DNA glycosylase in *E. coli* to initiate its repair.

This article is part of a discussion meeting issue ‘Frontiers in epigenetic chemical biology’.

1. Introduction

The major epigenetic mark in mammalian DNA is 5-methylcytosine (m\(^5\)C) [1]. The recent decade has provided substantial new knowledge on m\(^5\)C function and turnover, especially the elucidation of its removal through oxidation of the 5-methyl group by the TET (ten–eleven translocation) family of dioxygenases followed by base excision repair (BER) initiated by the thymine-DNA glycosylase (TDG) [1–3]. In prokaryotic DNA, m\(^5\)C shares importance with \(N^4\)-methylcytosine (m\(^4\)C) and \(N^\text{Me}\)-methyladenine (m\(^\text{MeA}\)) as canonical methyl-base modifications [4,5] which participate in biological functions such as protection against DNA cleavage in restriction–modification systems and DNA repair. Recently, m\(^\text{MeA}\) was discovered as a widespread second epigenetic methylated base in eukaryotes [6,7]. All three base modifications are products of enzyme-catalysed transfer of a methyl group from S-adenosyl-l-methionine (SAM) to the cognate unmodified base by a DNA methyltransferase (MTase) [4,8].
Most human promoters contain CpG islands, which are genomic regions with significant overrepresentation of the CpG sequence. There is a strong correlation between promoter methylation and gene repression [1], and aberrant methylation is associated with many disease states including cancer [9,10]. Consequently, interference with m^5C integrity may result in deregulation of genes in addition to mutagenicity and cytotoxicity.

It has been known for a long time that the CpG islands are mutational hotspots in eukaryotic cells [11]. The classical explanation is that 5-methylation of cytosine enhances the rate of spontaneous deamination under physiological conditions, and its deamination product, thymine, is repaired less efficiently than the deamination product of cytosine, uracil. However, most of the thymines generated from m^5C deamination are replaced with cytosine prior to replication, where species-specific repair systems often are involved. In Escherichia coli, the very-short-patch repair system repairs the T-G mismatches, while in mammalian cells the BER pathway initiated by TDG is most important. These mechanisms are extensively investigated and are comprehensively described elsewhere [12–14].

In contrast to hydrolytic damage to m^5C, few reports exist on damage inflicted by reactive oxygen species [15–19] and on damage inflicted by reactive oxygen species [15–19] and methylating agents or by by methylating agents and by MTases has received little attention [20]. Many prokaryotes enzymatically modify cytosine at either the N^4- or 5-position [5]. As m^5C has been shown to be a substrate for enzymatic N^4-methylation [20,21], we constructed model DNA substrates containing N^4,5-dimethylcytosine (m^4,5C) and used them to screen a series of DNA glycosylases for their activity towards this dimethylated base in DNA. Here we report, for the first time to our knowledge, enzymatic excision of doubly methylated cytosine, by showing that the MTases have received little attention [20]. Thus, the possible further methylation of m^5C to di- or trimethylated bases either by methylating agents or by by methylating agents and by MTases has received little attention [20].

2. Material and methods

(a) Preparation of m^4,5C-containing oligonucleotide substrate

Equimolar amounts of single-stranded forward (Fw) 5'-[Cy3]-CCCTC-GAGTTA[U]CATGGATC[Tag]-CGATCGATCC-3' and reverse (Rev) 5'-GGTGC[Tag]-CGATCGATCC[Tag]-CGTTCTCACAAACCTAGACGAATTCCG-3' were annealed to form a 30 nucleotide (nt) DNA duplex. The duplex oligonucleotide (10 nmol) was incubated with 1.6 nmol of M.Mval and 110 nmol of SAM in M.Mval buffer (50 mM Tris-HCl (pH 9.0), 1 mM dithiothreitol (DTT)) at 37 °C overnight. An aliquot of the reaction was analysed by digestion with Mval restriction endonuclease to confirm complete methylation (electronic supplementary material, figure SIA). M.Mval was removed by thermal denaturation (20 min at 65 °C) followed by centrifugation. Denatured oligonucleotides (5 min at 95 °C) were separated, analysed and purified (electronic supplementary material, figure SIB) on an integrated high-performance liquid chromatography/electrospray ionization/mass spectrometry system (Agilent 1290 Infinity) equipped with a Zorbax Eclipse Plus C18 (3.5 μm; 2.1 mm) by elution with a linear gradient of solvents A (5 mM ammonium acetate; pH 7.0 in water) and B (5 mM ammonium acetate, pH 7.0 in methanol) at a flow rate of 0.17 ml min^{-1} at 55 °C as follows: 0–13 min, 5–20% B; 13–20 min, 20–21% B; 20–25.5 min, 21–30% B; 25.5–27 min, 30–100% B. High-resolution mass spectra of samples were acquired on an Agilent Q-TOF 6520 mass analyser (100–3200 m/z range, negative ionization mode). The results were analysed with Agilent MassHunter Qualitative Analysis software (electronic supplementary material, figure SIC).

(b) Control oligonucleotide substrates

Synthetic 5'-labelled oligodeoxyribonucleotides 5'-[Cy3]-CCCTC-GAGTTA[U]CATGGATC[Tag]-CGATCGATCC-3' (Fw 60 nt; 11 nt incision product), 5'-[Cy3]-TAGACATGGCTCCGTAGTA[U]CATGGATC[Tag]-CGATCGATCC-3' (Fw 60 nt; 20 nt incision product) and 5'-[Cy3]-CCCTC-GAGTTA[U]CATGGATC[Tag]-CGATCGATCC-3' (Fw 30 nt; 11 nt incision product), containing dU at a specific internal site (Sigma-Aldrich), were annealed to equimolar amounts of complementary (Rev) strands with G opposite U.

(c) Repair enzymes

Nei (2.1 μM), Fpg (13 μM), Nth (0.7 μM), Ung (1.95 μM), hSMUG1 (0.33 μM), Nfo (83 nM) and PseT (0.29 μM) were obtained from New England Biolabs. M.U (100 units ml^{-1}) was obtained from Trevigen. hUNG (hUNG584 with His-tag; 820 μM) was a gift from Bodil Kavli and Geir Slupphaug. hOOG1 (with His-tag; 5.4 μM) was produced as described [24]. hTDG (with His-tag; 15 μM) was a gift from David Schurmann and Primo Scharf [25]. hUNG1 (full length with His-tag; 4.5 μM), hUNG2 (full length with His-tag; 5.3 μM) and hUNG3 (residues 1–301 with His-tag; 8.6 μM) were purified as described [26,27]. More details are provided in 'electronic supplementary material, Repair enzyme details'.

(d) Production and purification of M.Mval

The mvaIM gene was cloned into the expression plasmid vector pET15b through the NdeI and BamHII sites, resulting in addition of an N-terminal His_6 affinity tag. E. coli ER2566 cells carrying the pET15b-mvaIM plasmid and a T7 RNA polymerase gene on a separate plasmid, pACAR1 (G Vilkaitis 2004, unpublished data) were grown at 37 °C in LB medium containing ampicillin (100 μg ml^{-1}) and chloramphenicol (25 μg ml^{-1}) to OD_{600} of 0.8. IPTG (isopropyl b-D-thiogalactopyranoside; 1 mM) was added and cells were grown at 16 °C for 16 h. Cells were then sonicated and the supernatant was applied onto a HiTrap Chelating HP column (GE Healthcare). The protein was eluted with a 10–500 mM imidazole gradient, and pooled fractions were dialysed against a storage buffer (50 mM Tris-HCl (pH 8.0), 1 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl, 2 mM 2-mercaptoethanol, 50% glycerol). All purification steps were carried out at 4 °C.

(e) Assays for excision of m^4,5C from DNA and incision of m^4,5C-containing DNA

Purified DNA glycosylases at varying concentrations were incubated with m^4,5C-single-stranded DNA (ssDNA; electronic supplementary material, figure S1D), or with m^4,5C-double-stranded DNA (dsDNA) where m^4,5C was placed opposite G, A, C or T (named m^4,5C-G-DNA, m^4,5C-A-DNA, m^4,5C-C-DNA and m^4,5C-T-DNA, respectively; figure 1a) at 37 °C in a suitable reaction buffer (Nei buffer: 10 mM Tris-HCl (pH 8.0), 75 mM NaCl, 1 mM EDTA; Fpg buffer: 10 mM Bis-Tris-propane-
HCl (pH 7.0), 10 mM MgCl₂, 1 mM DTT, 0.1 mg ml⁻¹ bovine serum albumin (BSA), for other glycosylases see figure legends of the respective electronic supplementary figures) in a final volume of 20 µl, if not otherwise stated. Reactions were terminated by the addition of 20 mM EDTA, 0.5% (w/v) sodium dodecyl sulfate and proteinase K (150 µg ml⁻¹) followed by incubation at 37 °C for 10 min, precipitation of DNA with ethanol and solubilization of the precipitate in water (if not otherwise stated, 10 µl) [28]. m⁴,⁵C-DNA glycosylase activity was determined by NaOH-mediated (0.1 M final concentration) incision of the resulting abasic site (90 °C for 10 min; electronic supplementary material, figure S2A), where incision of the DNA without alkaline treatment measured the m⁴,⁵C-DNA incision activity (electronic supplementary material, figure S2B). Following addition of loading solution containing 80% (v/v) formamide, 1 mM EDTA and 1% (w/v) blue dextran (10 µl), and incubation at 95 °C for 5 min to denature DNA, the vial was cooled on ice followed by quick centrifugation at 4 °C. The samples (5 µl) were subjected to polyacrylamide gel electrophoresis (PAGE) using a gel (20% (w/v) polyacrylamide : bis-acrylamide 37.5 : 1) containing 8 M urea. The PAGE was performed using a Tris–borate–EDTA buffer system (89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8.0). Visualization and quantification were performed by fluorescence imaging analysis using ImageQuant Software (Molecular Dynamics).

(f) Preparation of m⁴,⁵C-containing plasmid DNA
The pBR322 plasmid DNA was isolated from E. coli strains ER2267 (ddcm⁻), methylated at C(T/A)G sites or ER2566 (ddcm⁺), unmodified DNA) using standard protocols. Plasmid DNA (15 µg) was incubated with 0.4 µM M.MvaI in 0.2 ml of reaction buffer containing 10 mM Tris-HCl (pH 8.5), 10 mM MgCl₂, 100 mM KCl, 0.2 mg ml⁻¹ BSA, 2 mM 2-mercaptoethanol and

---

Figure 1. E. coli Nei and Fpg proteins incise at m⁴,⁵C in DNA. (a) Structure of 5'-labelled dsDNA substrates for the determination of DNA glycosylase activity at modified cytosines (X) placed against different opposing bases (Y). Stars (*) indicate nuclease-resistant phosphorothioate bonds; [Cy3], cyanine3 fluorophore. (b) Nei exhibits activity towards m⁴,⁵C (but not C, m⁶C or m⁷C) when placed opposite G in DNA. Nei (8.3 pmol) was incubated with the DNA substrate (1 pmol) at 37 °C for 1 h (final volume, 20 µl). (c) Nei exhibits the highest activity towards m⁴,⁵C when opposite G in DNA. For experimental details, see electronic supplementary material, figure S3. The average (± s.d.) of four independent experiments is presented. (d) Fpg exhibits the highest activity towards m⁴,⁵C when placed opposite C, with lower activity opposite T, in DNA. DNA substrate (1 pmol) at 37 °C 1 h. U·G-DNA (60 nt; 1 pmol) incubated without (lane 10) and with hUNG, the latter followed by NaOH/heat treatment (lane 11), was used as negative control for active hUNG, respectively, which was used to convert U·G-DNA into AP-DNA to demonstrate active Fpg (i.e. lyase activity; lane 12). Reaction products were analysed by denaturing PAGE at 200 V for 2 h. The average (± s.d.) of three independent experiments is presented in (d)(ii). (Online version in colour.)
0.5 mM SAM at 37°C for 4 h. DNA was then linearized with R.HindIII, extracted with phenol : chloroform (1:1) and chloroform, and precipitated with ethanol. The completeness of the 
MvmaI-methylation in reaction aliquots was confirmed by R.MvmaI restriction endonuclease digestion and agarose gel electrophoresis (see electronic supplementary material, figure S9).

(g) Nuclease S1 assay

Doubly methylated DNA (or mono-methylated / unmethylated controls) (0.5 μg) was incubated with 20 units S1 nuclease (Thermo Fisher Scientific) in supplier buffer for 1 h at 37°C. Reactions were stopped by proteinase K (Thermo Fisher Scientific) treatment and analysed by 1.5% agarose gel electrophoresis.

3. Results

To investigate possible enzymatic excision of m⁴,5C from DNA, a 5’ fluorescently labelled oligodeoxyribonucleotide ssDNA oligomer, m⁴,5C-ssDNA, was alone, or annealed to a complementary strand with m⁴,5C placed opposite either cognate G or non-cognate A, T or C (figure 1g, lane 1), exposed to a series of available DNA glycosylases from E. coli or human (see below). Enzymatic base excision by so-called mono-functional DNA glycosylases results in an alkali-labile apurinic/apyrimidinic (AP) site, which can be monitored by the extent that NaOH cleaves this site by a β/δ-elimination reaction [29] (electronic supplementary material, figure S2A). Bi-functional DNA glycosylases like Nei and Fpg will in addition cleave the resulting AP-DNA, and these consecutive activities can be monitored by PAGE under denaturing conditions without NaOH/heat treatment (electronic supplementary material, figure S2B).

Interestingly, we observed that E. coli Nei exhibited significant activity for m⁴,5C in dsDNA (figure 1b, lane 8), while no m⁴,5C-DNA incision occurred without the enzyme (figure 1b, lane 4). Control experiments were carried out with identical DNA substrates where m⁴,5C was replaced by m¹C, m⁴C or C, showing that Nei exhibited no activity towards the normal cytosines (figure 1b, lanes 5–7). The activity was highest when m⁴,5C was placed opposite cognate G (figure 1c and electronic supplementary material, figure S3, lane 6), and substantially lower opposite non-cognate A and T (figure 1c and electronic supplementary material, figure S3, lanes 8 and 9, respectively), while close to zero opposite C (figure 1c and electronic supplementary material, figure S3, lane 7). Uracil-DNA glycosylase of E. coli (Ung) was used to form AP-DNA to confirm active Nei, due to the ability of the latter to recognize and incise AP sites (electronic supplementary material, figure S3, lane 12), while NaOH/heat treatment of AP-DNA was used to confirm active Ung (electronic supplementary material, figure S3, lanes 10 and 11). As the activity of Nei for m⁴,5C-G was the highest, as well as this being the normal arrangement in vivo, further studies on Nei were only carried out using m⁴,5C-G-DNA.

We also observed that E. coli Fpg exhibited significant activity for m⁴,5C in DNA, although only when opposite pyrimidines on the complementary strand and with the highest activity opposite C (figure 1d, lanes 7 and 9). No m⁴,5C-DNA incision occurred without enzyme (figure 1d, lanes 1–5) or when m⁴,5C was placed opposite G or A (figure 1d, lanes 6 and 8, respectively). Surprisingly, Fpg also exhibited similar activity for m¹C and m⁴C, and even, although lower, for C itself when placed opposite C (AN Tesfahun, P Guragain, M Alexeeva, A Arshad, M Tomkuvienė, JK Laerdel, A Klungland, S Klimauskas, S Bjelland 2017, unpublished results). This prompted us to conclude that the activity of Fpg on m⁴,5C (figure 1d) is the manifestation of a hitherto unknown (methyl)C-C mismatch activity, rather than a specific m⁴,5C repair function of Fpg. This activity is presently under investigation. All other enzymes tested (electronic supplementary material, figures S4–S6), that is, E. coli Ung, mismatch-specific uracil-DNA glycosylase (Mug) and endonuclease III (Nth) as well as human uracil-DNA glycosylase (hUNG), single-strand-selective mono-functional uracil-DNA glycosylase (hSMUG1), hTDG and 8-oxoguanine-DNA glycosylase (hOGG1) showed no detectable DNA glycosylase activity for m⁴,5C at the concentrations and under the incubation conditions employed. Because of previously detected activity for base lesions placed opposite all normal bases and also in ssDNA for enzymes in the uracil-DNA glycosylase superfamily (UDG), Ung, hUNG, Mug and hSMUG1 were tested for m⁴,5C removal in all these contexts without showing any activity (electronic supplementary material, figures S4A, S5A, S4B and S5B, respectively).

Based on the known substrate preference, the family 2 UDG hTDG, and the bi-functional glycosylases Nth and hOGG1 [19], were only examined for DNA incision at the damage site in the dsDNA context, both showing no detectable activity (electronic supplementary material, figures S4C and S5C). As opposed to E. coli Nei, we also found no m⁴,5C-specific activity by the human orthologues hNEIL1 (full length; electronic supplementary material, figure S6A), hNEIL2 (full length; electronic supplementary material, figure S6B) or hNEIL3 (truncated; amino acids 1–301; electronic supplementary material, figure S6C), neither opposite G, C, A or T nor placed in ssDNA, the latter only having been analysed for NEIL3. Active NEIL1 and NEIL2 were confirmed by their active AP lyase functions for dsDNA, while active NEIL3 was verified by active AP lyase function for ssDNA.

Determination of the m⁴,5C-G-DNA incision activity as a function of protein concentration showed that the extent of incision follows the same equation that describes the dependence of V₀ as a function of [S] in the Michaelis–Menten kinetic analysis (electronic supplementary material, figure S7A and B). However, it should be noted that the enzyme/substrate concentration ratio used was above 1:1, i.e. approaching single-turnover conditions. A similar profile was observed when the incision activity was determined as a function of time (electronic supplementary material, figure S7C and D), using an eightfold excess of enzyme over substrate. Kinetic analysis of the m⁴,5C-G-DNA incision activity of Nei indicated Michaelis–Menten behaviour (electronic supplementary material, figure S8), resulting in an apparent Kₘ of 311 ± 80 mM, a Vₘₐₓ of 3.61 ± 0.46 mM min⁻¹ and a kₐp of 0.0087 ± 0.0011 min⁻¹. These parameters were also determined with 1 mM of DTT in the incubation buffer, as often used in glycosylase assays, resulting in virtually identical values (data not shown). The obtained values are quite close to those determined for other Nei substrates [22,30,31].

To confirm the presence of a 3’-phosphorylated product following the incision of m⁴,5C-G-DNA by Nei, we treated Nei-exposed DNA with T4 polynucleotide kinase (PseT), which specifically removes phosphate from the 3’-end [32,33], generating a slower-migrating product in PAGE (figure 2). As opposed to control incubation in the absence of enzyme or
incubation with Nei alone which only formed 3’-phosphate, incubation with Nei followed by endonuclease IV (Nfo) [34] showed additional formation of a 3’-OH product (figure 2).

A similarly efficient processing of the Nei product was achieved by addition of PseT, which converted the 3’-phosphate into the 3’-OH product. This confirms that the Nei-mediated incision of m\textsuperscript{4,5}C·G-DNA forms a 3’-phosphate, and accords with the presently accepted knowledge that Nfo may displace Nei from the resultant AP site as well as remove the 3’-phosphate [35], both resulting in the 3’-OH product necessary for repair synthesis.

To assess a possible impact of m\textsuperscript{4,5}C on the helical DNA structure (figure 3a), we also performed probing of modified plasmid DNA with nuclease S1, which is specific for ssDNA regions including DNA mismatches and lesions that disrupt dsDNA structure [38]. This DNA contained \textit{Cm}\textsuperscript{m4,5}C\textsubscript{CWGG} sites as a result of \textit{in vivo} methylation by \textit{E. coli} Dcm MTase followed by \textit{in vitro} methylation with R.Mval (electronic supplementary material, figure S9). The experiment showed extensive cleavage of the \textit{m}\textsuperscript{4,5}C·G-DNA at CCWGG sites by S1, yielding a fragmentation pattern similar to that obtained by R.Mval digestion of unmodified control DNA. These results clearly demonstrate that the presence of \textit{m}\textsuperscript{4,5}C·G base pairs, as opposed to those involving \textit{m}\textsuperscript{4}C or \textit{m}\textsuperscript{5}C, leads to local distortion of the helical DNA structure that is readily detectable by a structure-sensitive enzyme \textit{in vitro} (figure 3b).

4. Discussion

All three types of biological DNA methylation (i.e. \textit{m}\textsuperscript{4}C, \textit{m}\textsuperscript{5}C and \textit{m}\textsuperscript{6}A) occur in the major groove of the DNA helix where they can be readily sensed by cellular proteins and interpreted as biological signals. Importantly, these modifications do not alter or preclude the coding capacity of the target nucleobases, preserving the original content of the genome [8]. Here we present the first report to our knowledge describing a repair activity for a further methylated \textit{m}\textsuperscript{6}C residue in dsDNA, \textit{m}\textsuperscript{4,5}C, in which two types of biological methylation occur in a single nucleobase. We show that the \textit{E. coli} Nei DNA glycosylase/lyase initiates BER of \textit{m}\textsuperscript{4,5}C in DNA by strand incision at the lesion site \textit{in vitro} (figure 1b). The activity observed was highest when the lesion was placed opposite cognate G, and 3–4 times lower opposite A and T, while almost no activity was detected opposite C (figure 1c). We also show that \textit{E. coli} Fpg exhibits a similar activity for \textit{m}\textsuperscript{4,5}C in DNA, removing \textit{m}\textsuperscript{4,5}C most efficiently opposite

![Figure 2](http://rstb.royalsocietypublishing.org/)  
**Figure 2.** Definition and processing of the 3’-end following Nei-mediated incision of \textit{m}\textsuperscript{4,5}C·G-DNA. DNA substrate (figure 1a; 1 pmol) was incubated without (lane 1) or with Nei (2.1 pmol; lanes 2–4) at 37°C for 30 min, followed by no addition (lanes 1 and 2), addition of 0.083 pmol endonuclease IV (Nfo; lane 3) or addition of 0.29 pmol T4 polynucleotide kinase (PseT; lane 4) and incubation for an additional 30 min (final volume, 10 μl). Reaction products were analysed by denaturing PAGE at 500 V for 4 h. P, phosphate. (Online version in colour.)

![Figure 3](http://rstb.royalsocietypublishing.org/)  
**Figure 3.** The \textit{m}\textsuperscript{4,5}C residue disrupts normal DNA structure. (a) Structural considerations of the \textit{m}\textsuperscript{4,5}C residue in dsDNA [20,36,37] show how rotation around the N4–C4 bond of \textit{m}\textsuperscript{4,5}C relieves the steric clash between the methyl groups (upper panel) but at the same time disrupts the normal G·C hydrogen-bonding pattern (lower panel), precluding a normal Watson–Crick DNA structure. (b) Impact of \textit{m}\textsuperscript{4,5}C on DNA structure probed by susceptibility to nuclease S1 digestion. pBR322 plasmid DNA containing \textit{m}\textsuperscript{4,5}C, mono-methylated (either \textit{m}\textsuperscript{4}C or \textit{m}\textsuperscript{5}C) or unmethylated cytosine (c), respectively, at CC(A/T)GG sites (electronic supplementary material, figure S9) was incubated with nuclease S1 and analysed by agarose gel electrophoresis. R.Mval lane, commercial pBR322 digested with R.Mval; L, GeneRuler DNA Ladder Mix.
performed by the bi-functional DNA glycosylase Nei as reported here. The damaged base excision step 1 and the AP site incision step 2 are presented here show that the m
4,5C residue specifically sensitizes DNA for S1 nuclease digestion (figure 3b), supporting a proposed irregular (open) structure of the m
4,5C -G base pair. In its cytidine form, m
4,5C is generally more stable to deamination and hydrolysis/depyrimidination than cytosine itself, m
5Ca and m
5Cc [43], suggesting considerable in vivo stability and the need for active repair.

Our results indicate that the active site of at least two bacterial DNA repair enzymes is able to accommodate the m
4,5C residue during catalysis. This contrasts with several tested helix–hairpin–helix (HhH) or UDG DNA glycosylases of E. coli and human cells (electronic supplementary material, figures S4, S5) like Ung, Mug, Nth, hUNG, hSMUG1, hTDG and hOGG1, which showed no detectable activity under the incubation conditions employed. Curiously, this also applies to the human NHEJ orthologues hNEIL1, hNEIL2 and hNEIL3 (electronic supplementary material, figure S6). To our knowledge, m
4,5C is the first methylated base found to be a substrate for Nei. Nei, as well as Fpg, has a flexible and spacious active-site pocket that can bind a variety of nucleobase lesions, providing numerous direct or water-mediated atomic contacts [44]. Indeed, m
4,5C is quite similar to some other oxidized pyrimidine or purine N base substrates [19]. Unlike Nei and Fpg, DNA glycosylases in the HhH superfamily, for example E. coli Nth and hOGG1, and the UDG superfamily, including E. coli Ung and Mug, as well as hUNG, hSMUG1 and hTDG, bind their target base lesions in deeply buried and relatively rigid recognition pockets [44], and these DNA glycosylases, therefore, appear to be mainly specialized for damaged bases of a single type and with little structural diversity.

Although relatively rare, the presence of cytosine-N4 MTases potentially acting at Dcm-methylated cytosines has been described in E. coli. For example, the M.EcoNI MTase produces m
4,5C at CCGNNNAGG sites (BP Antón, T Clark, M Boitano, J Korlach, RJ Roberts 2012, unpublished observations; http://rebase.neb.com/cgi-bin/erefget?5672), which may overlap with highly abundant Dcm Cnm
5C(T/A)GG sites in certain genomic contexts. Similar double methylation can also occur in other bacterial species or strains containing both types (m
5Ca and m
5Cc) of MTases operating on overlapping targets in genomic DNA. Foreign MTase genes may be acquired by horizontal transfer of mobile elements of plasmid-borne restriction–modification systems or by viral infections. m
4,5C may also arise as a consequence of chemical insults to the cellular DNA.

Additional indirect evidence for its harmful effects and efficient removal of genomic m
4,5C in vivo comes from previous experiments that found negligible levels of m
4,5C in DNA of E. coli cells carrying active mnaLM and dcm genes, although persisting production of m
4,5C in DNA led to the induction of the SOS response [20]. Indeed, because of its helix-desaturating features (figure 3), m
4,5C is probably also an efficient substrate for the nucleotide excision repair (NER) protein complex UvrABC in E. coli [12,45]. In particular, the SOS response induces several trans-lesion bypass polymerases probably able to insert e.g. C or T opposite m
4,5C in DNA [19]. We primarily expect Nei to initiate error-free BER of m
4,5C in DNA opposite cognate G (figure 4), which, as suggested above, may be bypassed by a replicative or trans-lesion polymerase to insert a non-cognate base opposite it. Then, Nei or Fpg-mediated error-prone repair can be a second line of defence if m
4,5C escapes the first Nei or maybe NER defence line and continues to be a substrate for Nei. Nei, as well as Fpg, has a flexible and spacious active-site pocket that can bind a variety of nucleobase lesions, providing numerous direct or water-mediated atomic contacts [44]. Indeed, m
4,5C is quite similar to some other oxidized pyrimidine or purine N base substrates [19]. Unlike Nei and Fpg, DNA glycosylases in the HhH superfamily, for example E. coli Nth and hOGG1, and the UDG superfamily, including E. coli Ung and Mug, as well as hUNG, hSMUG1 and hTDG, bind their target base lesions in deeply buried and relatively rigid recognition pockets [44], and these DNA glycosylases, therefore, appear to be mainly specialized for damaged bases of a single type and with little structural diversity.

Although relatively rare, the presence of cytosine-N4 MTases potentially acting at Dcm-methylated cytosines has been described in E. coli. For example, the M.EcoNI MTase produces m
4,5C at CCGNNNAGG sites (BP Antón, T Clark, M Boitano, J Korlach, RJ Roberts 2012, unpublished observations; http://rebase.neb.com/cgi-bin/erefget?5672), which may overlap with highly abundant Dcm Cnm
5C(T/A)GG sites in certain genomic contexts. Similar double methylation can also occur in other bacterial species or strains containing both types (m
5Ca and m
5Cc) of MTases operating on overlapping targets in genomic DNA. Foreign MTase genes may be acquired by horizontal transfer of mobile elements of plasmid-borne restriction–modification systems or by viral infections. m
4,5C may also arise as a consequence of chemical insults to the cellular DNA.

Additional indirect evidence for its harmful effects and efficient removal of genomic m
4,5C in vivo comes from previous experiments that found negligible levels of m
4,5C in DNA of E. coli cells carrying active mnaLM and dcm genes, although persisting production of m
4,5C in DNA led to the induction of the SOS response [20]. Indeed, because of its helix-desaturating features (figure 3), m
4,5C is probably also an efficient substrate for the nucleotide excision repair (NER) protein complex UvrABC in E. coli [12,45]. In particular, the SOS response induces several trans-lesion bypass polymerases probably able to insert e.g. C or T opposite m
4,5C in DNA [19]. We primarily expect Nei to initiate error-free BER of m
4,5C in DNA opposite cognate G (figure 4), which, as suggested above, may be bypassed by a replicative or trans-lesion polymerase to insert a non-cognate base opposite it. Then, Nei or Fpg-mediated error-prone repair can be a second line of defence if m
4,5C escapes the first Nei or maybe NER defence line and continues as a target for DNA synthesis. Because of high cellular toxicity, an ‘accepted’ price for m
4,5C removal is mutagenesis.

Data accessibility. The datasets supporting this article have been uploaded as part of the electronic supplementary material.
References


27. Leiro J et al. 2007 Structural basis for enzymatic excision of N4-methyladenine and N4,N6-methylcytosine from DNA. EMBO J. 26, 2206 – 2217. (doi:10.1038/sj.emboj.7601662)


