Universitetet i Stavanger

Faculty of Science and technology

MASTER'S THESIS

Study program/Specialization: **Biological Chemistry**	**Spring semester 2017** **Restricted access**
Author: **Prashanna Guragain**	---------- (signature of author)
Faculty Supervisor: **Svein Bjelland**	
Title of master's thesis: **DNA glycosylase activities for *N*4,5-dimethylcytosine**	
Credits (ECTS): **60**	
Key Words: **Methylation** **DNA Glycosylases** **Fpg** **Nei** **Enzyme Kinetics** **Dimethylcytosine**	Pages: **60** + Supplementary materials: **38** Date/year: **July 2017**

Acknowledgement

This master thesis concludes my two-year Master's Degree in Biological Chemistry at University of Stavanger (UiS). It was an outcome of an extensive scientific research conducted at the Center for Organelle Research (CORE), led by Prof. Svein Bjelland, located at Måltidets Hus, ipark, Stavanger.

First, I want to express my gratitude to my main supervisor Svein Bjelland for providing me this opportunity to work with him as a part of his research group, exploring the field of epigenetics and DNA damage. I would also like to thank him for motivating me, giving me advice and feedback during my research work and writing. I would like to express my gratitude to my research team, Marina Alexeeva and Almaz Tesfahun for their guidance during the experiments in the laboratory in CORE. Their guidance in practical work helped me a lot to improve the finesse working skills in glycosylase assays and gel electrophoresis. I would also like express thanks to our laboratory engineer Xiang Ming Xu for his guidance in health and safety regulations for working in the research laboratory at CORE and assistance in ordering the materials. There were some snags during the experiments and I would like to acknowledge Prof. Lutz Eichacker for generous ideas regarding gel electrophoresis and a special gratitude to Prof. Peter Ruoff for his time and plentiful suggestions for enzyme kinetics and instrumental optimization which aided me a lot in my work. I would like to thank MSc student Aysha Arshad for mutual support with similar problems. I would also like to thank Prof. Geir Slupphaug, Prof. Magnar Bjørås, Prof. Primo Schär, David Schürmann and Hanne Korvald for the gift of enzymes that were crucial in this project. I would also like to thank Tanya for supporting and motivating me in every aspect during my stay in Stavanger. I would like to thank everyone who has encouraged me, motivated me during my time at UiS and CORE.

Finally, I am grateful to my brother for always being there for me as a friend and I am forever indebted to my parents who selflessly encouraged me to explore new directions in life and pursue my own destiny. If not for them this journey would not have been possible so I dedicate this milestone to them.

Abstract

Although the epigenetic DNA base 5-methylcytosine (m^5C) has an important role in cellular functions, damaging chemical alterations to m^5C have been little studied. For example, while much knowledge exists on erroneous methylation of the four common bases in DNA, almost none studies have been conducted on methylation damage to m^5C resulting in double and triple methylated bases.

Certain methylases can convert m^5C into N^4,5-dimethylcytosine ($m^{N4,5}C$) in DNA *in vitro*, and there is a possibility of the presence of $m^{N4,5}C$ *in vivo*. We investigated the ability of various DNA glycosylases to process DNA containing $m^{N4,5}C$ at a specific site, and we report that *Escherichia coli* Fpg protein and endonuclease VIII (Nei) exhibit activity for $m^{N4,5}C$ in DNA *in vitro*. Fpg removes $m^{N4,5}C$ most efficiently opposite non-cognate C followed by T, while no activity was detected opposite A and cognate G. In contrast, Nei incises at $m^{N4,5}C$ in DNA most efficiently opposite cognate G followed by A and T, whereas almost no activity was detected opposite C. Nei and Fpg thus seem to complement each other in the repair of $m^{N4,5}C$ in DNA. Plasmids containing $m^{N4,5}C$ placed opposite G, C, A and T should separately be transformed into *E. coli* wild-type, *fpg*$^-$, *nei*$^-$ and *fpg*$^-$ *nei*$^-$ cells to study the *in vivo* consequences of these repair functions. Our findings describe for the first time the repair of a further methylated epigenetic base in DNA.

List of research papers

i. Bjelland, S., Tesfahun, A., Alexeeva, M., Tomkuviene, M., Arshad, A., Guragain, P., Klimasauskas, S., Jørgensen, K.B., Klungland, A. and Robertson, A.B. **Spontaneous and enzymatic modifications of the epigenetic DNA base 5-methylcytosine as targets for repair and mutagenesis.** *Manuscript.*

ii. Tesfahun, A., Guragain, P., Alexeeva, M., Arshad, A., Tomkuviene, M., Lærdahl, J.K., Klungland, A., Klimasauskas, S. and Bjelland, S. **Excision of the double methylated base *N*4,5-dimethylcytosine from DNA by *Escherichia coli* Fpg protein.** *Manuscript.*

iii. Alexeeva, M., Guragain P., Tesfahun, A., Arshad, A., Tomkuviene, M., Korvald, H., Bjørås, M., Lærdahl, J.K., Klungland, A., Klimasauskas, S., and Bjelland, S. ***Escherichia coli* Nei protein excises *N*4,5-dimethylcytosine from DNA.** *Manuscript.*

Table of Contents

List of Figures

List of Tables

Abbreviations

AP, apurinic/apyrimidinic (abasic)

AlkA, 3-methyladenine-DNA glycosylase II

BER, base excision repair

CpG, 5′-C-phosphate-G-3′

Cy3, Cyanine 3

Fpg, formamidopyrimidine-DNA glycosylase

hNEIL1, human Nei-like 1 endonuclease

hNEIL2, human Nei-like 2 endonuclease

hNEIL3, human Nei-like 3 endonuclease

hOGG1, human 8-oxoguanine-DNA glycosylase

hSMUG1, human single mono-functional uracil-DNA glycosylase 1

hTDG, human thymine-DNA glycosylase

hUNG, human uracil-DNA glycosylase

MBD4, methyl-CpG-binding domain protein 4

m^5C, 5-methylcytosine

$m^{N4,5}C$, *N*4,5-dimethylcytosine

MPG, methylpurine-DNA glycosylase

Nei, endonuclease VIII

Nth, endonuclease III

PMT, photomultiplier tube

SAM, *S*-adenosylmethionine

Tag, 3-methyladenine-DNA glycosylase I

UDG, uracil-DNA glycosylase

1 Introduction

1.1 DNA damages

Because the cellular DNA replication machinery is not perfect, genomic errors arise from DNA polymerase (mis)incorporation of an incorrect or mismatched normal base during replication (*e.g.*, an A opposite C) which is potentially mutagenic (Kim et al., 2012). Moreover, the exposure to different endogenous and exogenous agents causes the decomposition of DNA and is responsible for its limited chemical stability (Lindahl, 1993). Historically, exogenous agents have received much attention because they provide an extra burden to our survival and also measures can be taken to minimize exposure to them. One example is ionizing radiation, which in addition to oxidative damages forms single- and particularly double-strand breaks; the most devastating damages known. Another example is carcinogens found in cigarette smoke like polycyclic aromatic hydrocarbons (PAHs), the latter attaches to DNA bases forming large adducts. For detoxification, many carcinogens including PAHs are hydroxylated by Cytochrome P450 enzymes in an oxygen-requiring reaction to make them more water soluble; however, some intermediates formed possesses a highly reactive electrophilic center which easily can form DNA adducts. During replication, if DNA adducts are bypassed incorrectly by a polymerase, mutations can occur which may result in tumor development if growth controlling genes are involved (Pfeifer et al., 2002). A third example is the induction in DNA of cyclobutane pyrimidine dimers (CPDs) and 6–4 photoproducts by ultraviolet (UV) light heavily involved in skin carcinogenesis (Krwawicz et al., 2007; Pfeifer et al., 2005).

Despite all challenges inflicted by exposures from the environment, it is nevertheless chemical insults originating as a consequence of the cell chemistry itself that poses the major threat to DNA integrity. Such reactions *e.g.* alter DNA directly or damage deoxynucleoside monophosphates (dNMPs) before their incorporation into DNA during DNA synthesis. Since the major component of a cell is water, apurinic/apyrimidinic (AP; abasic) sites formed by hydrolytic depurination rather than depyrimidination (base loss) are the most abundant DNA damage, and the hydrolytic deamination of cytosine to uracil generate the potentially mutagenic U:G mismatch (Lindahl, 1993). Aerobic organisms have an extra challenge in protecting themselves from the oxidation of their DNA by reactive oxygen species (ROS) produced as a

byproduct of respiration, which in eukaryotic species make mitochondria an important source of ROS, in addition to ROS generated by the immune system and during microsomal metabolism in mammals. The most studied oxidatively damaged base is 7,8-dihydro-8-oxoguanine (oxo^{8}G), which induces the G to T transversion; a common somatic mutation in cancer (Grollman et al., 1993). In addition, a multitude of purine and pyrimidine base damages has been identified in DNA exposed to ROS, and has been extensively reviewed (Bjelland et al., 2003).

Bases in DNA altered chemically by hydrolysis or oxidation can be mutagenic and potentially carcinogenic or inhibit DNA replication or transcription causing cellular toxicity. This also applies to a third major and spontaneous insult, alkylation or methylation of the DNA bases at different positions by endogenous or exogenous alkylating agents as *e.g.* the methyl donor *S*-adenosylmethionine (SAM) in cells, which is the theme of the next chapters.

1.1.1 Common DNA bases damaged by methylation

Methylating agents are very cytotoxic and normally mutagenic, the former justifying their use in the treatment of some cancers. Depending on the nature of the methylating agent, its reaction mechanism and secondary structure of target DNA, methylating agents alkylate DNA at many sites producing a variety of base lesions and phosphotriesters. Methylating agents can react at 12 different sites on the DNA base ring nitrogens, exocyclic oxygens and oxygens in the sugar-phosphate backbone (Figure 1), whereas the proportion of the alkylation occurring on different sites depends on the mode of action (S_N1 or S_N2 nucleophilic substitution; Figure 2) of the methylating agent (Sedgwick, 2004; Sedgwick et al., 2002).

Endogenous SAM is an efficient methyl group donor in most cellular transmethylation reactions, which weakly methylates the DNA in a non-enzymatic manner by producing similar pro-carcinogenic and pro-mutagenic lesions as formed by some carcinogenic chemical methylating agents. Similar to methylsulfonium compounds and various nucleophiles, the transmethylation reaction by SAM also occurs by the S_N2 mechanism as observed from the spectrum of the product detected (Barrows et al., 1982; Rydberg et al., 1982). 7-Methylguanine (m^{7}G) and 3-methyladenine (m^{3}A) are the major DNA lesions where the former alteration is

principally harmless but the latter is cytotoxic blocking DNA replication and thus a major threat to the cell (Lindahl, 1993). However, m^7G depurinates rapidly creating cytotoxic abasic sites in DNA (Philip et al., 1996).

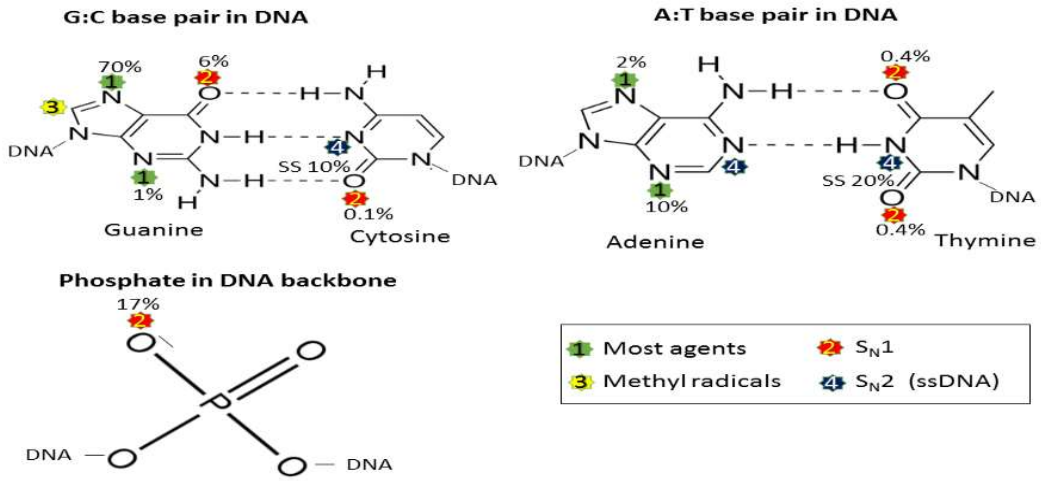


Figure 1: Methylation sites on the bases and sugar-phosphate backbone of DNA.
Oxygen atoms in DNA that are most frequently methylated by S_N1 agents like N-methyl-N′-nitro-N-nitroguanidine (MNNG) are indicated by red circles. S_N2 agents like methylmethane sulphonate that methylates ssDNA are indicated by blue circles. The yellow circle shows site methylated by methyl radicals and green circle indicate sites methylated by most agents. The percentage indicate the relative abundance of each modification. Adapted from (Sedgwick, 2004).

Table 1: DNA lesions in cells caused by methylation.

Lesion	Mode of formation	Agents responsible	Effects
3-Methyladenine (m^3A)	Methylation of adenine by S_N1 methylating agents and SAM	SAM, MMS	Cytotoxic, blocks DNA replication
7-Methylguanine (m^7G)	Methylation of guanine by S_N1 methylating agents and SAM	SAM, MMS	Harmless but abasic sites are cytotoxic
1-Methyladenine (m^1A)	Alkylation by S_N2 chemical agents	MMS and methyl halides	Cytotoxic, blocks DNA replication
3-Methylcytosine (m^3C)	Alkylation by S_N2 chemical agents	MMS and methyl halides	Cytotoxic, blocks DNA replication
O^6-Methylguanine ($m^{O6}G$)	Genomic alkylation by endogenous nitrosamines and S_N1 agents	MNNG, MNU	Cytotoxic, G:C→A:T transition mutation
O^4-Methylthymine ($m^{O4}G$)	Genomic alkylation by endogenous nitrosamines and S_N1 agents	MNNG, MNU	A:T→G:C mutation

Adapted from (Kim et al., 2012; Sedgwick, 2004).

S_N1 agents like *N*-methyl-*N*-nitrosourea (MNU) and *N*-methyl-*N′*-nitro-*N*-nitrosoguanidine (MNNG) are mutagenic and react more readily with oxygens in DNA creating the major adduct O^6-methylguanine ($m^{O6}G$) and the minor adduct O^4-methylthymine ($m^{O4}T$). $m^{O6}G$ is a highly mutagenic lesion which mispairs with thymine during replication resulting in G:C → A:T transition mutations (Sedgwick, 2004). S_N2 agents like methylmethane sulfonate (MMS) and methyl halides react with DNA generating 1-methyladenine (m^1A) and 3-methylcytosine (m^3C), which predominantly arise in single-stranded DNA (ssDNA) as opposed to double-stranded DNA (dsDNA) where the reactive sites are protected by base pairing (Figure 1). These major base lesions block DNA replication (Bodell et al., 1979; Sedgwick, 2004).

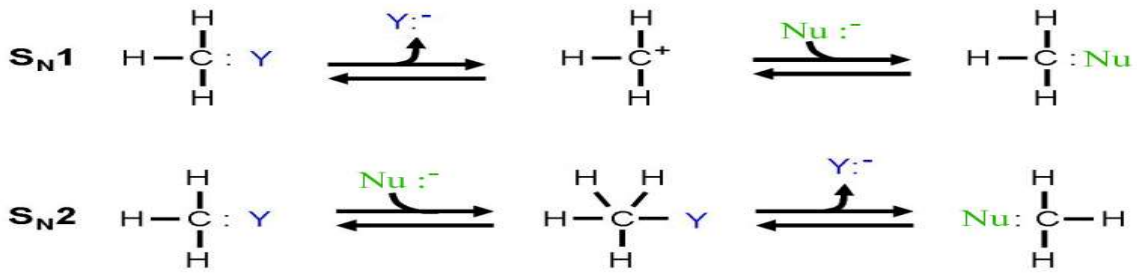


Figure 2: Nucleophilic substitution reactions.
When a departing ion (Y) is replaced by an electron-rich nucleophile (Nu), a nucleophilic substitution reaction occurs. In the monomolecular S_N1 reaction, the leaving group departs before the nucleophile arrives. In the bimolecular S_N2 reaction, both groups are involved in a transient intermediate phase. In DNA methylation, methylating agents generate the leaving group and bases are the source of the nucleophile. Adapted from (Sedgwick, 2004).

1.1.2 The major epigenetic DNA base can also be damaged including by methylation

The inheritable changes in gene expression without any alteration in DNA sequences are epigenetics and fundamental mechanisms of cancer are related to the change in the epigenome. DNA methylation is one of the important mechanisms regulating gene expression where the major epigenetic mark is cytosine methylated in the 5-position occurring enzymatically in CpG dinucleotides with SAM as the methyl donor. Mostly, gene transcription is inhibited either by blockade of transcription factors and their binding sites or by recruitment of methyl-binding domain protein that mediates inhibition of gene expression (Chen et al., 2014). The methylation quantities and patterns differ between cells and disposition of m^5C throughout the genome is

highly regulated which has profound effects on cell identity and transcriptional profiles and in the case of aberrations of m^5C profiles contributes to tumorigenesis and improper development (Klungland et al., 2016). Among all epigenetic modifications, cytosine methylation is the only known symmetric modification with an established maintenance mechanism. Modification of m^5C on CpG dinucleotides permits the inheritance of methylation patterns through DNA replication. The methylation patterns can be authentically propagated through DNA replication from parental strand to unmethylated daughter strand as carried out by DNMT1 DNA methyltransferase complemented by DNMT3 methyltransferase (Breiling et al., 2015). It has been established that the oxidation of m^5C is catalyzed by the TET enzyme family (Figure 4) and can result in 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine (Ito et al., 2011; Klungland et al., 2016).

In bacteria, N^6-methyladenine ($m^{N6}A$) outshines the m^5C base modification and has been shown to be essential for viability of several bacteria being associated with genome protection via restriction-modification systems, a bacterial defense mechanism against phages and plasmids that are able to distinguish between host and invader (Breiling et al., 2015; Heyn et al., 2015). Additional roles are that $m^{N6}A$ guides the discrimination between original and newly synthesized DNA strand in DNA mismatch repair, as well as reduces the stability of base-pairing and therefore is able to support transcriptional initiation by lowering the necessary energy required to open duplexes (Heyn et al., 2015).

The CpG sequences are mutational hotspots and about 35–40% of the mutations in human genetic diseases are base substitutions at these spots. The traditional explanation is that the hydrolytic deamination of m^5C to thymine, which is less efficiently repaired than uracil formed in the same way from cytosine, causes G:C to A:T transition mutations (Cooper et al., 1988). In addition, m^5C is damaged by ROS where some of the damages are specific to m^5C like 5-formylcytosine (f^5C), 5,6-dihydroxy-5,6-dihydro-5-methylcytosine (m^5C-gly), 5-hydroxy-5-methyl-4-aminohydantoin (m^5C-hyd) etc. (*Manuscript i.*) and some are generated enzymatically as an intermediate in m^5C demethylation as described below. Furthermore, m^5C can be damaged by methylation to double and triple methylated bases which also is presented in more detail in further chapters.

1.2 DNA base damage repair and epigenetic demethylation

Since the DNA in all cells is exposed to endogenous and exogenous agents causing many different types of harmful damages some occurring in quite large amounts, they must be corrected efficiently. The genome integrity and a relatively low mutation frequency are maintained by several different types of DNA repair mechanisms including the error-correcting exonuclease function of the replicative DNA polymerases, or else the outcomes would lead to *e.g.* disease initiation and progression in humans (Kim et al., 2012). There are various types of DNA repair mechanisms but here we will emphasize much on Base excision repair mechanism.

There are two direct reversal mechanisms where the damaging methyl group in the bases is removed by methyltransferases or by oxidative demethylation DNA dioxygenases. The main mutagenic DNA lesion $m^{O6}G$ and the minor lesion $m^{O4}T$ are repaired by $m^{O6}G$-DNA methyltransferase transferring the methyl group to a nucleophilic cysteine residue in the active site of the protein. This reaction causes the irreversible inactivation of the protein making reaction quick, stoichiometric but also costly by consuming the transferase (Sedgwick, 2004). The constitutive Ogt protein and the inducible Ada protein are the *E. coli* $m^{O6}G$-DNA methyltransferases, the former protecting against the endogenous DNA methylation damages whereas Ada is induced by and protects against damages caused by external agents. The higher alkyl adducts and $m^{O4}T$ are repaired by Ogt while Ada efficiently repairs $m^{O6}G$ (Sedgwick, 2004). Oxidative demethylation catalyzed by AlkB is a direct reversal mechanism being a part of, as Ada, the adaptive response in *E. coli*. AlkB is a member of the α-ketoglutarate-dioxygenase superfamily that oxidizes the aberrant methyl group in m^1A and m^3C in both ssDNA and dsDNA, and is dependent on oxygen, α-ketoglutarate and Fe(II) to react (Begley et al., 2003; Li et al., 2013). The damaging methyl group is hydroxylated to decompose resulting in the release of CO_2, succinate, and formaldehyde, where m^1A and m^3C are reversed to the unmodified adenine and cytosine, respectively (Trewick et al., 2002). AlkB also provides protection against various epoxides that generate hydroxyalkyl adducts and ethylating and propylating agents. 1-Ethyladenine is regenerated by AlkB where oxidation of the ethyl group is followed by the release of acetaldehyde (Sedgwick, 2004). Under alkylation stress, AlkB is regarded as the versatile gatekeeper of the genomic integrity that can repair all simple N-alkyl

adducts occurring at the Watson-Crick base pairing interface of the four DNA bases (Li et al., 2013).

In another important DNA repair pathway called nucleotide excision repair (NER), which removes a wide range of helix-distorting DNA adducts and bulky lesions including UV damages, the damage is removed together with up- and downstream deoxynucleotides comprising 12–13 nucleotides in *E. coli* and 24–32 nucleotides in humans. In addition, non-bulky alkylation adducts like $m^{O6}G$ and $m^{N6}A$ are also repaired by NER. In *E. coli*, the recognition and incision at both sides of the damage are carried out by an excinuclease complex of the three proteins UvrA, UvrB, and UvrC and requires ATP. The resulting gap in DNA is filled by DNA polymerase I and the nick is sealed by DNA ligase. In human cells, components of the exinuclease complex remain bound to the post-incision gap to prevent non-specific degradation of the single-stranded region, and DNA polymerase ε or δ or both attached to proliferating cell nuclear antigen (PCNA) carry out the repair synthesis which is followed by ligation by one of three DNA ligases (Petit et al., 1999).

1.2.1 Base excision repair

Correction of most forms of alkylated, oxidized and deaminated bases in DNA is carried out by the base excision repair (BER) pathway (Kim et al., 2012), which in *E. coli* is initiated by one of seven and in mammalian cells one of nine DNA glycosylases which cleaves the N–C1′ glycosylic bond between the damaged base and the deoxyribose moiety (Krokan et al., 2000; Seeberg et al., 1995). In the presence of a plethora of normal bases, DNA glycosylases search for, recognize and then accommodate the damaged base in their active site pocket after kinking the DNA and flipping the substrate base out from its partnership with the opposite base (Krokan et al., 2013). These enzymes are classified according to the catalytic mechanism as either mono-functional or bifunctional, where the former creates an AP site following excision of the damaged base whereas the latter is able to incise the DNA strand after base excision. Mono-functional glycosylases use a water molecule as a nucleophile to attack the deoxyribose C atom while bifunctional glycosylases use an active site amino moiety as a nucleophile to create a covalent Schiff base between the deoxyribose C and the protein, eventually incising the DNA strand (Kim et al., 2012). Methylated and deaminated bases like uracil are recognized and

removed by mono-functional DNA glycosylases while oxidized and ring saturated bases are mostly processed by bifunctional DNA glycosylases. After base excision and DNA incision, BER proceeds by two sub-pathways, *i.e.* short patch-BER (SP-BER) and long patch-BER (LP-BER) (Krwawicz et al., 2007). The second step in the BER pathway is that typically an AP endonuclease, but also sometimes the AP lyase function of a bifunctional DNA glycosylase, incises the abasic site created by the glycosylase wherein the phosphodiester bond is broken generating a single-stranded break in the DNA strand which contained the damage. Incision by an AP endonuclease results in a free 3′-OH end to prepare for the third repair replication step, while a 5′-deoxyribose phosphate (dRP) lyase function is needed to prepare for the final ligation step. If the abasic site is incised by β- or β/δ-elimination reaction as carried out by an AP lyase, the abasic site remnant is left behind as a 3′-α,β-unsaturated/saturated aldehyde (3′-PUA) or a 3′-phosphate, which need to be removed by a 3′-phosphodiesterase or a 3′-polynucleotide phosphatase function, respectively, to form the 3′-OH end required for replication. Shortly: if the incision in step two is performed by an AP endonuclease, a 5′-dRP lyase is required for the final ligation step; if performed by an AP lyase, a 3′-phosphodiesterase or 3′-polynucleotide phosphatase function is required for the following replication event. Principally, the AP endonuclease and 3′-phosphodiesterase activity are functions of the same protein, while the 5′-dRP lyase function is usually provided by the (repair) polymerase. Mostly, in case or repair by bifunctional glycosylases SP-BER takes place, whereas in case of mono-functional glycosylases, if the 5′-dRP moiety is removed by 5′-dRP lyase, SP-BER takes place otherwise the repair function proceeds by LP-BER. The final DNA ligase step seals the nick in DNA by utilizing either ATP-dependent or NAD^+-dependent phosphoanhydride hydrolysis to create a covalent phosphodiester bond between 3′-OH and 5′-PO_4 (Figure 6) (Kim et al., 2012; Seeberg et al., 1995; Subramanya et al., 1996).

1.2.1.1 BER in *E. coli*

Uracil-DNA glycosylase (UDG) which has been classified into six families initiates the repair of uracil in DNA in all cells (Lee et al., 2011). *E. coli* Ung is a family 1 UDG coded by the *ung* gene which is active against uracil on ssDNA and dsDNA regardless of the partner base on the opposite strand. In very diverse organisms such as humans, *E. coli* and herpes simplex virus the active site pocket of the family 1 UDGs is highly conserved and provides a unique and specific

binding site for uracil disfavoring binding of virtually any other base (Pearl, 2000). Thus, Ung is specific for uracil in addition to exhibiting some activity for some uracil derivatives like dihydroxyuracil (dhU) and 5-hydroxyuracil (h^5U) (Table 2). Family 2 UDGs like Mug are mismatch-specific glycosylases requiring guanine on the complementary strand; they may have a broader specificity removing both uracil and thymine from mismatches with guanine (Barrett et al., 1998; Pearl, 2000). In addition, Mug has additional specificity for the removal of lesions like h^5U, 5-formyluracil (f^5U), 5-hydroxymethyluracil (hm^5U) and 5-hydroxycytosine (h^5C) from DNA (Table 2).

In *E. coli*, the repair of alkylated bases in DNA is mediated by the two DNA glycosylases 3-methyladenine glycosylase I (Tag) and 3-methyladenine glycosylase II (AlkA). Tag is a constitutive enzyme and has high activity for m^3A but can also remove 3-methylguanine (m^3G) from DNA, a major cytotoxic lesion in DNA (Bjelland et al., 1993). The AlkA enzyme exhibits a broad substrate specificity but comprises only 10% of the methylbase-removing glycosylase activity in cells (Seeberg et al., 1995), being a part of the adaptive response to alkylation. Thus, the *in vivo* concentration of AlkA increases ten-fold after the cells have been exposed to a sublethal dose of a methylating agent (Seeberg et al., 1995). AlkA removes m^3G, m^3A (Bjelland et al., 1993), m^7G, O^2-methylcytosine ($m^{O2}C$), O^2-methylthymine ($m^{O2}T$), hm^5U (Seeberg et al., 1995), 8-methylguanine (m^8G) which is induced by methyl radicals, the deamination product hypoxanthine, ethenobases and f^5U from DNA. Indeed, a weak activity for the normal bases in DNA illustrates the considerable promiscuity of AlkA in base selection (Sedgwick, 2004).

The oxidized purine and pyrimidine lesions in DNA are removed by a set of enzymes, and first to be identified in *E. coli* was Nth (endonuclease III) which recognized DNA bases damaged by ionizing radiation. Fpg (formamidopyrimidine-DNA glycosylase) was the next enzyme to be identified in *E. coli*, which in addition to formamidopyrimidines remove oxo^8G, dhU, spiroiminodihydantoin (Sp) and guanidinohydantoin (Gh) from DNA. Another enzyme identified much later is Nei (endonuclease VIII), which has sequence homology to Fpg but significant substrate overlaps with Nth (Table 2) (Lee et al., 2017). Nth, Fpg and Nei are bifunctional DNA glycosylases (Figure 5).

Following DNA incision by bifunctional glycosylases in the SP-BER pathway, the 3′-PUA after β-elimination and 3′-P after β/δ-elimination are removed by the exonuclease activity of Xth (exonuclease III). Although 95% of the total removal of the 3′-blocking ends from DNA is performed by Xth (Demple et al., 1986), Nfo (endonuclease IV) is also involved (Figure 5). Then, DNA polymerase I (Pol I) incorporates the complementary deoxynucleotides and the nick is sealed by DNA ligase (Krwawicz et al., 2007).

In the case of mono-functional glycosylases, the abasic sites created are incised by the AP endonucleases Xth and Nfo generating the 3′-OH and 5′-dRP ends. The repair process further continues by LP-BER or SP-BER depending on the types of the enzymes involved. If the 5′-dRP is removed by the 5′-dRP-lyase activity of Pol I, Fpg or Nei, SP-BER takes place. If the 5′-dRP is not removed prior to or simultaneously with repair synthesis, the repair proceeds by LP-BER where the synthesis by Pol I displaces the dRP-containing strand and Pol I also cleaves the displaced strand by its 5′–3′ exonuclease activity; 2–8 nucleotides are removed and replaced followed by ligation by DNA ligase (Krwawicz et al., 2007).

1.2.1.1.1 *E. coli* Fpg and Nei

Fpg protein was discovered in Lindahl′s laboratory as a DNA glycosylase that removes the 2,6-diamino-4-hydroxy-5-*N*-methylformamidopyrimidine (mfapyG) lesion from alkylated DNA exposed to alkaline conditions, where mfapyG is formed by ring-opening of the m^7G imidazole ring. Nei was discovered in Wallace′s laboratory as a glycosylase that recognizes oxidized pyrimidines (Prakash et al., 2012). Both enzymes are bifunctional DNA glycosylases that incises the AP site by β/δ-elimination. In addition, Fpg and Nei are capable of cleaving a 5′-preincised AP site at the 3′-site with its dRPase activity (Jiang et al., 1997; Serre et al., 2002).

Besides mfapyG, Fpg removes fapyA, fapyG, and oxo^8G and its derivatives Gh and Sp when positioned opposite C. It has also activity for oxidized pyrimidines like thymine glycol (Tg) and 5,6-dihydrouracil (dHU) opposite G, and h^5U opposite C (Guo et al., 2010). Fpg shows opposite base specificity with strong interaction of an active site arginine residue for C compared to G and T while showing no interaction with A (Fromme et al., 2002). Nei recognizes a broad

spectrum of oxidized bases such as Tg, dhU, dihydrothymine (dHT), fapyA, fapyG, Sp and Gh (Table 2). Although Nei has a structural and sequence similarity with Fpg (Figure 3) and these enzymes recognize some common lesions, Nei also exhibits a significant substrate overlap with Nth (Lee et al., 2017).

As opposed to hUNG and 8-oxoguanine-DNA glycosylase (hOGG1) where the extruded substrate bases are recognized by specific interactions between the damaged base and their active site pocket amino acids, Fpg and Nei show much more flexibility regarding substrate base accommodation. However, the structure of Fpg indicates an efficient interaction of a structural loop with oxo^{8}G which is involved in the stabilization of the lesion in active site pocket. This contrasts with Nei which lacks this structural loop and is much less efficient in removing oxo^{8}G from DNA.

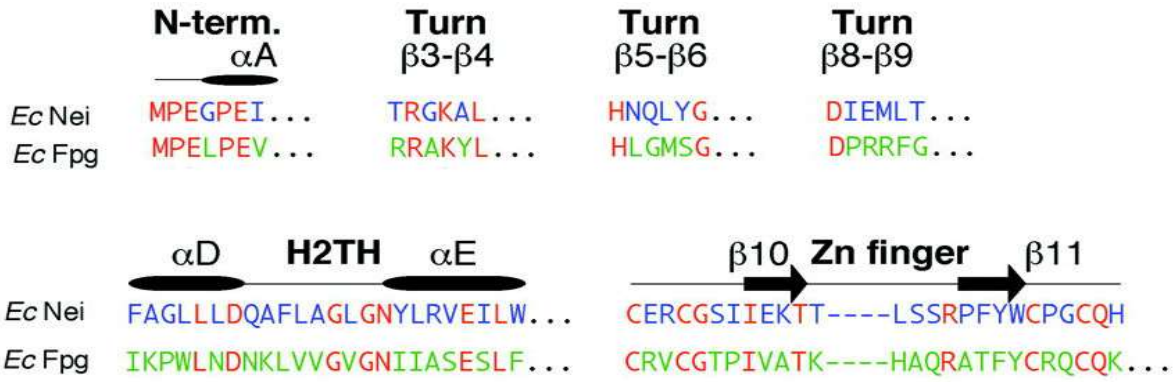


Figure 3: Sequence similarity of Fpg and Nei protein.
The alignment of *E. coli* Fpg and Nei in putative DNA-contacting regions. The residues which are absolutely conserved in both proteins are shown in red. Adapted from (Takao et al., 2002).

1.2.1.2 BER in mammalian cells

In contrast to bacterial counterparts, mammalian DNA glycosylases have N-terminal extensions which have a crucial role in targeting enzymes to nuclei or mitochondria and in interaction with other proteins that may have a role in BER (Krokan et al., 2000). Uracil in nuclear DNA is known to be removed by four distinct DNA glycosylases: hUNG2, single strand specific mono-functional uracil-DNA glycosylase (hSMUG1), thymine-DNA glycosylase (hTDG) and methyl CpG domain-binding protein 4 (MBD4). In addition, hSMUG1 recognizes other substrates like

hm^{5}U and f^{5}U in ss and dsDNA. hTDG is specific for thymine and derivatives opposite G like hm^{5}U and f^{5}U, and some years ago it was reported to be involved in m^{5}C demethylation by removing TET-oxidation products (Hashimoto et al., 2012; Ito et al., 2011; Krokan et al., 2013).

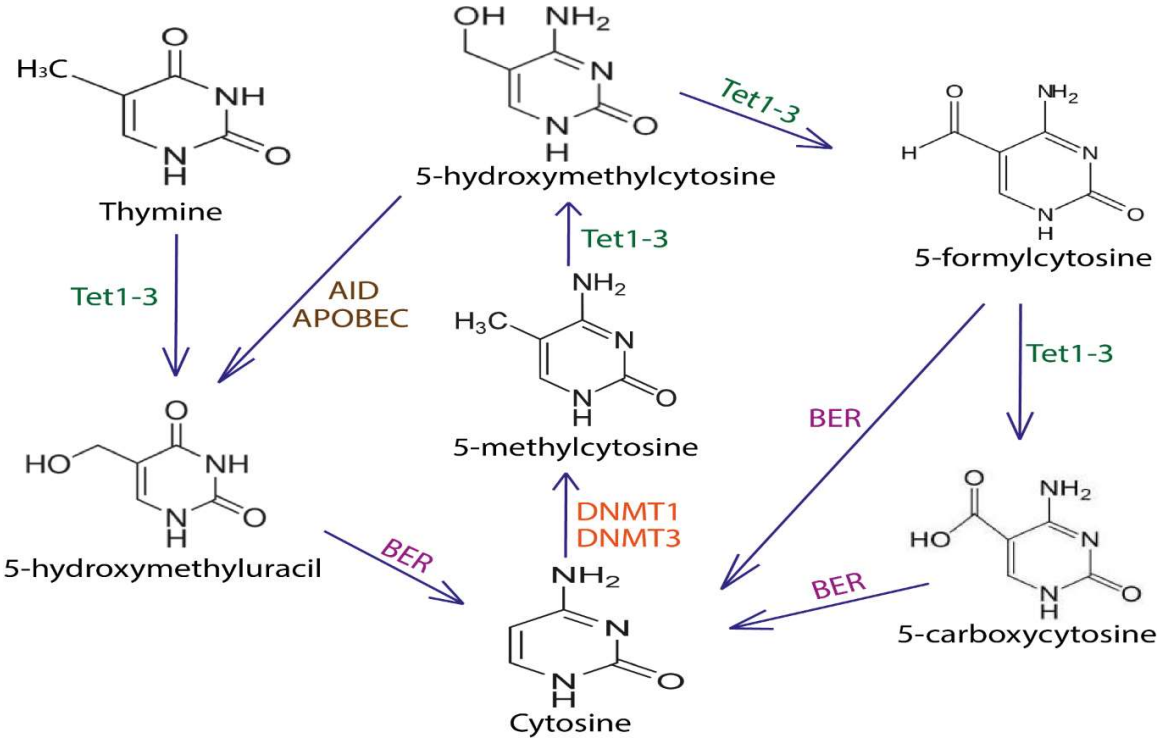


Figure 4: Cytosine modifications and possible demethylation pathways.
Cytosine is converted to m^{5}C by two DNA methyltransferases, DNMT1 and DNMT3. m^{5}C is oxidized by Tet1, Tet2, or Tet3 (Tet1–3) to hm^{5}C which can be further oxidized by Tet1–3 yielding f^{5}C and 5-carboxycytosine (ca^{5}C). Both f^{5}C and ca^{5}C are removed by hTDG and further processed by BER pathway. Alternatively, hm^{5}C is enzymatically deaminated by AID or APOBEC to hm^{5}U; which is removed by hSMUG1 or hTDG and processed by BER pathway. Thymine can also be oxidized by Tet1–3 resulting in hm^{5}U. Altogether these pathways could result in unmodified cytosine. Adapted from (Klungland et al., 2016).

Lesions in mammalian DNA caused by oxidation are mainly removed by endonuclease III homolog 1 (hNTH1), hOGG1 and Nei-like 1/2/3 [hNEIL(1/2/3)] glycosylases. The removal of oxo^{8}G and corresponding ring-fragmented purine fapyG is catalyzed by hOGG1. hNTH1 is involved in the removal of fapyA/fapyG along with oxidized pyrimidines. The oxidized purines and pyrimidines in DNA is removed by three hNEIL glycosylases. hNEIL1 has a broad substrate specificity and recognizes oxidized pyrimidines like h^{5}U, h^{5}C as well as ring saturated

dHT and dHU. It has also been shown to recognize oxo^{8}G in dsDNA. In mismatched dsDNA, hNEIL2 is specific for h^{5}U and oxidized derivatives of cytosines. hNEIL1 and hNEIL2 also removes Sp and Gh from both ss and dsDNA and 8-oxoadenine (oxo^{8}A) when paired with C. hNEIL3 has preference of lesions in ssDNA and in bubble structures and recognizes Sp, Gh, fapyA, fapyG but not oxo^{8}G (Krokan et al., 2013). The single human methylpurine-DNA glycosylase (MPG) has a wide substrate range including several alkylated purines, including m^{7}G and m^{3}G, 1,*N*6-ethenoadenine (εA) and hypoxanthine and shares this property with the bacterial enzyme AlkA even though they are not related in amino acid sequence (Krokan et al., 2013; Krokan et al., 2000).

In mammalian cells, SP-BER is carried out by the core proteins presented in Table 2. The 3′-PUA generated after β-elimination by bifunctional glycosylases like hOGG1 and hNTH1 is removed by AP endonuclease 1 (APE1), a homolog to *E. coli* Xth. After β/δ-elimination by hNEIL1 and hNEIL2, polynucleotide kinase 3′-phosphatase (PNKP), a bifunctional enzyme involved in repair, hydrolyzes the 3′-phosphate to 3′-OH by its 3′-phosphatase activity. PNKP also exhibits 5′-kinase activity to phosphorylate 5′-OH to 5′-phosphate ends. Both SP-BER and LP-BER are initiated by DNA polymerase β (Pol β) in mammalian cells, which also removes 5′-dRP by its 5′-dRPase activity, to prepare for strand closure by DNA ligase I or III (LIG1/3). It is believed that LIG3 is more important than LIG1 in SP-BER. It has also been argued that LIG1 is involved in nuclear repair whereas LIG3 is important for mitochondrial repair. Poly(ADPribose) polymerase 1 (PARP1) and X-ray repair cross complementing 1 (XRCC1) along with various other proteins participate but are thought not to be required for all SP-BER (Krokan et al., 2013; Krwawicz et al., 2007).

LP-BER takes place when a modification of the 5′-dRP moiety by oxidation or reduction prevents its excision by Pol β. First, Pol β falls off and proliferating cell nuclear antigen (PCNA) is recruited together with DNA polymerase ε (Pol ε) or DNA polymerase δ (Pol δ) which adds few nucleotides to the 3′-OH end and generates a flap containing the 5′-dRP end which is removed by flap endonuclease 1 (FEN-1) (Figure 5). Finally, the nick is sealed by a complex of LIG1 and PCNA (Krokan et al., 2013; Krwawicz et al., 2007).

Table 2: Human and *E. coli* DNA glycosylases and primary substrates.

Damaged base type	*Name*	*Gene*	*Mono/ bifunctional*	*Known substrates*
Human				
Deaminated	hUNG	*UNG*	M	ssU, U:G, U:A, h^5U, isodialuric acid, alloxan
	hSMUG1	*SMUG1*	M	ssU, U:A, U:G, f^5U, hm^5U, h^5U
	hTDG	*TDG*	M	U:G, T:G, f^5U:G, f^5U:A, hm^5U:G, hm^5U:A, εC in ss and dsDNA, f^5C, ca^5C
	MBD4	*MBD4*	M	U:G, T:G, hm^5U in CpG, εC in ss and dsDNA
Oxidized	hOGG 1	*OGG1*	M/B, β elimination	fapyG:C, oxo^8G:C
	hNEIL1	*NEIL*	B, via β/δ elimination	Tg, fapyG, fapyA, oxo^8G, h^5U, dHU, Sp and Gh in ss and dsDNA.
	hNEIL2	*NEIL*	B, via β/δ elimination	Similar to hNEIL1
	hNEIL3	*NEIL*	M/B, via β/δ elimination	ssAP, fapyG, fapyA, Sp and Gh in ssDNA
Alkylated	MPG/AAG	*MPG*	No	m^3A, m^3G, m^7G, hypoxanthine, εA, $m^{N6}A$, m^7A
Escherichia coli				
Oxidized	Fpg	*fpg*	B, via β/δ elimination, dRP lyase	fapyG:C, fapyG:A, Gh, Sh, fapyA, AP sites, me-fapy, oxo^8G, oxo^8A
	Nei	*nei*	B, via β/δ elimination, dRP lyase	fapyG, Gh, Sp, fapyA, Ap sites, Tg, dHT, Th^5, hmh, urea, h^5C, dhC, Ug, h^5U, dHU, dhU, β-ureidoisobutaric acid, Uh^5, oxanine, xanthine
	Nth	*nth*	B, via β elimination	fapyG, Gh, Sp, fapyA, Ap sites, Tg, dHT, Th^5, hmh, urea, h^5C, dhC, Ug, h^5U, dHU, dhU, mfapy, Cg, f^5U, εA degradation products
Methylated	Tag	*tag*	M	m^3A, m^3G, m^7G
	AlkA	*alkA*	M	m^3A, m^3G, m^7G, hm^5U, f^5U, hypoxanthine, $m^{O2}C$, $m^{O2}G$,
Deaminated	Mug	*mug*	M	ssU, U:G, T:G, ssT, hm^5U:A, hm^5U:G, f^5U:A, f^5U:G, h^5C:G, h^5U:G,
	Ung	*ung*	M	ssU, U:G, U:A, h^5U, dhU

Abbreviations: AP, apurinic/apyrimidinic; Cg, cytosine glycol; ca^5C, 5-carboxymethylcytosine; dhC, 5,6-dihydroxycytosine; dhU, 5,6-dihydroxyuracil; dHU, 5,6-dihydrouracil; dHT, 5,6-dihydrothymine; εA, 1,*N*6-ethenoadenine; εC, 3,*N*4-ethenocytosine; f^5C, 5-formylcytosine; f^5U, 5-formyluracil; fapyA, 4,6-diamino-5-formamidopyrimidine; fapyG, 2,6-diamino-4-hydroxy-5-formamidopyrimidine; Gh, guanidinohydantoin; h^5C, 5-hydroxycytosine; h^5U, 5-hydroxyuracil; hm^5U, 5-hydroxymethyluracil; hmh, 5-hydroxy-5-methyl hydantoin; m^3A, 3-methyladenine; $m^{N6}A$, 6-methyladenine; m^7A; 7-methyladenine; $m^{O2}C$, *O*2-methylcytosine; $m^{O2}G$, *O*2-methylguanine m^3G, 3-methylguanine; m^7G, 7-methylguanine; $m^{O4}T$, *O*4-methylthymine; mfapy, 2,6-diamino-4-oxo-5-(N-methyl)formamidopyrimidine; oxo^8A, 7,8-dihydro-8-oxoadenine; oxo^8G, 7, 8-dihydro-8-oxoguanine; Sp, spiroiminodihydantoin; ss, single-stranded; Tg, thymine glycol; Th^5, 5-hydroxy-6-hydrothymine; Ug, uracil glycol; Uh^5, 5-hydroxy-6-hydrouracil. Adapted from: (Krokan et al., 2013; Kim et al., 2012; Krwawicz et al., 2007; Seeberg et al., 1995; Lee et al., 2017; Klungland et al., 2016).

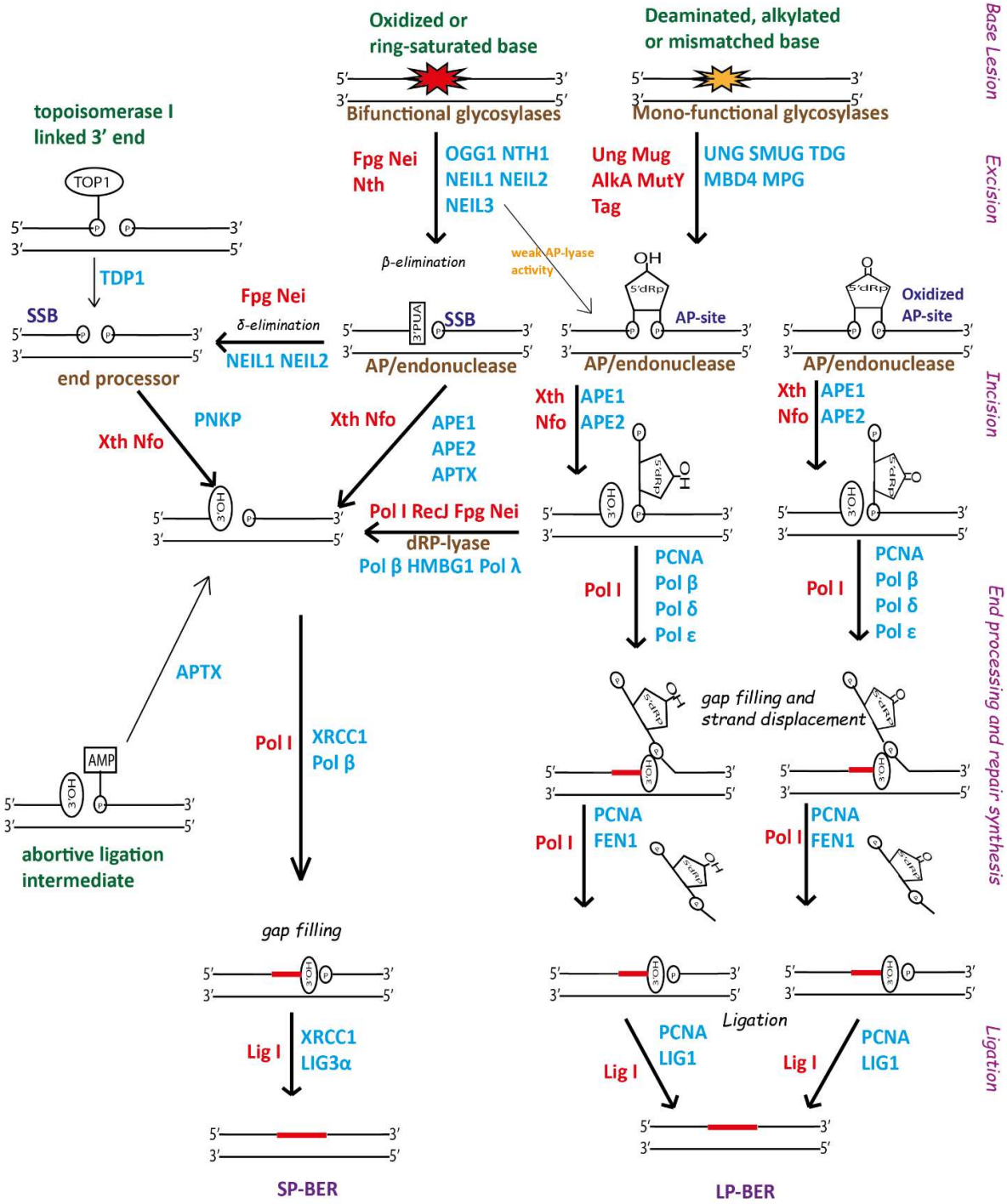


Figure 5: Model for base excision repair in mammals and bacteria.
P, phosphate, OH, hydroxyl group; 3′PUA, 3′-unsaturated aldehydes; 5′dRP, 5′-deoxyribose phosphate; AMP, adenylate group; APTX, Aprataxin; TOP1, topoisomerase I-linked 3′-end; SSB, single strand break; The types of DNA lesions repaired by common sub pathways of single strands break repair and base excision repair are marked in dark blue. *E. coli* enzymes are in left, in red; human enzymes are in right, in light blue. Adapted from (Krokan et al., 2013; Krwawicz et al., 2007).

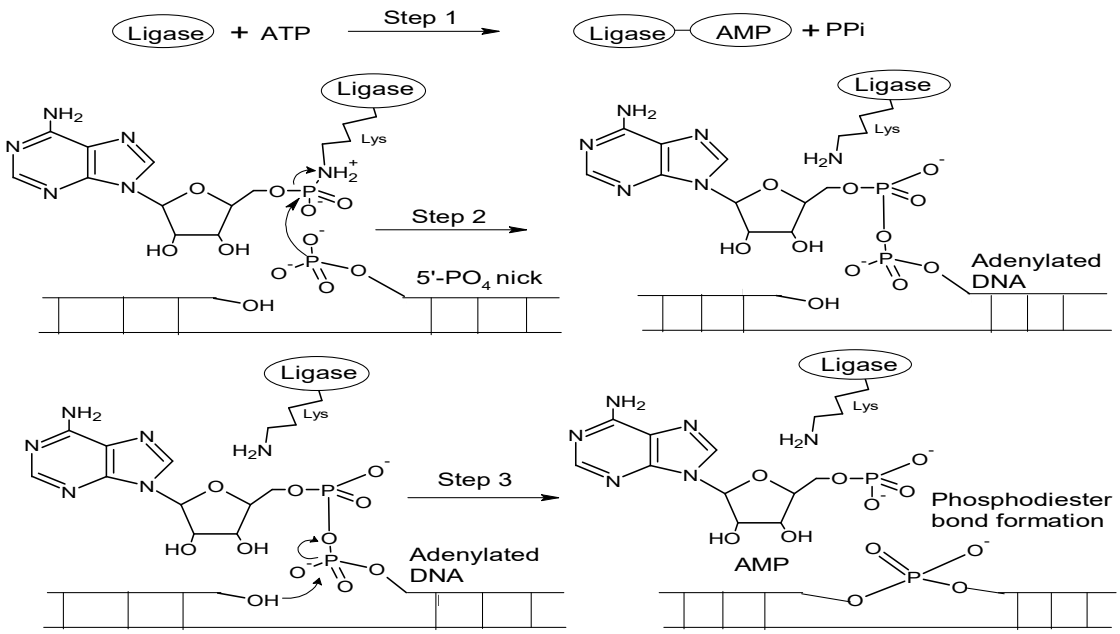


Figure 6: Nick sealing mechanism by DNA ligase
The catalysis by DNA ligase proceeds via a number of steps starting with ATP or NAD^+ hydrolysis to covalently link AMP and active site lysine releasing inorganic pyrophosphate (PP_i) or nicotinamide mononucleotide. Then, adenylated enzyme transfers the AMP moiety from lysine to the 5′-phosphate end of DNA forming a pyrophosphate linkage (5′P–5′P). The activated 5′-phosphate of DNA is attacked by 3′-hydroxyl group of adjacent DNA forming the phosphodiester bond and releasing AMP from the adenylated DNA intermediate (Subramanya et al., 1996).

1.3 *N*4,5-dimethylcytosine: generation and possible repair

The transfer of a methyl group from SAM to cytosine in the presence of a DNA methyltransferase generates *N*4-methylcytosine ($m^{N4}C$) or m^5C dependent on whether the methyltransferase targets either the *N*4- or 5-position of cytosine. *N*4,5-dimethylcytosine ($m^{N4,5}C$) is generated by the successive enzymatic modification of DNA by these two types of methyltransferases that recognizes identical or overlapping sequences (Klimasauskas et al., 2002). Out of the two reciprocal pathways proposed (Figure 7), the generation of dimethylated cytosine was only possible when the C5 position was methylated first followed by the N4 position in cytosine. The enzymatic incorporation of the methyl group in $m^{N4,5}C$ is not possible since C5-methyltransferase (C5-MT) has very highly conserved catalytic domain. Moreover, the interaction of the enzyme with both hydrogen atoms at *N*4 position and formation of

transient-covalent bond at the C6 position of the aromatic ring is needed to achieve the methylation at the C5 position in cytosine (Klimasauskas et al., 2002).

In vivo studies of *E. coli* harboring Dcm naturally and a plasmid producing MvaI methyltransferase showed presence of both mono-methylated cytosines, however, no $m^{N4,5}C$ was detected in DNA. It was suggested that $m^{N4,5}C$ was efficiently repaired in DNA, which might be indirectly supported because the *E. coli* SOS system was induced by the M.MvaI-producing plasmid (Klimasauskas et al., 2002).

The structural considerations of the planar conformations of this base suggest that the *cis* conformer appears less likely to occur than the *trans*-conformer because the former is sterically more demanding due to the clash between the two methyl groups at *N*4 and C5. However, *trans* conformer is free from the steric hindrance but it is not compatible with the Watson-Crick base pairing pattern resulting in a distortion in the β-helical structure which may disturb *e.g.* protein-DNA interactions (Klimasauskas et al., 2002). If m^5C is converted to $m^{N4,5}C$ in mammalian DNA it may cause epigenetic dysregulation in addition to toxicity and mutagenicity. Hence, we want to establish the possible repair of this double-methylated cytosine *in vitro.*

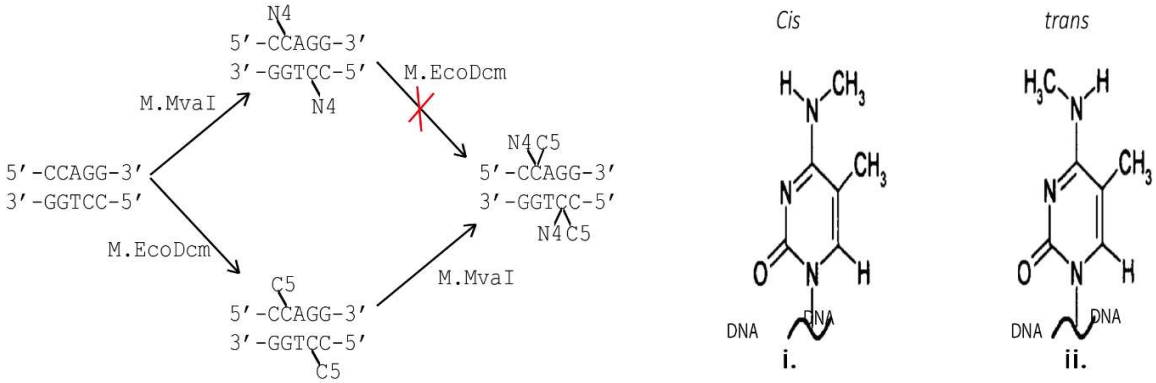


Figure 7: Formation and stereochemistry of *N*4,5-dimethylcytosine.
Enzymatic formation of $m^{N4,5}C$: The theoretical pathway that leads to the double methylation of cytosine is methylation at C5 first by enzyme M.Eco Dcm and then at N4 by M.MvaI (left panel). However, the upper pathway does not lead to the formation of dimethylcytosine. *Possible planar conformers of* $m^{N4,5}C$: *i.* The steric clash between the two groups highly disfavors *cis* conformer; *ii.* There is no steric hindrance in *trans* conformer but proper H-bonding in dsDNA is disfavored (right panel). Adapted from (Klimasauskas et al., 2002).

1.4 Aim of the study

The major epigenetic mark in mammals is m^5C. Presently no knowledge exists on the presence and possible biological consequences of further methylated m^5C in their DNA. One such dimethylated cytosine is $m^{N4,5}C$, which can be introduced into a specific sequence and site in DNA by MvaI methyltransferase if it contains a correctly placed m^5C. At the start, the aim of this study was to use such DNA to

i. identify human and *E. coli* DNA glycosylases able to initiate BER of the $m^{N4,5}C$ residue *in vitro*,

which resulted in the identification of the *E. coli* Fpg and Nei proteins able to initiate such repair. Then the aim of the study was shifted to

ii. measure kinetic parameters of Fpg and Nei for $m^{N4,5}C$ in DNA.

2 Material and Methods

2.1 DNA substrates

2.1.1 $m^{N4,5}C$-containing substrates

The Cyanine 3 (cy3) labeled 30-mer polydeoxynucleotides containing m^5C lesion at 12th position and the unlabeled m^5C containing reverse polydeoxynucleotides strand were synthesized by Sigma-Aldrich. The polydeoxynucleotides were annealed and further methylated at the *N*4 position by the treatment with SAM and M.MvaI and purified in Laboratory for Biological DNA Modification, Institute of Biotechnology, Graiciuno 8, LT-2028 Vilnius, Lithuania (Klimasauskas et al., 2002). The substrates were end protected with phosphothiorate to protect from the degradation by nucleases. The forward labeled polydeoxynucleotide with $m^{N4,5}C$ lesion was annealed to its complementary strand having G, C, A, or T opposite the lesion in equimolar ratio by heating at 95°C for 5 min and gradual cooling of 1°C per min to 23°C for 2 h (Appendix A).

Table 3: DNA substrate oligonucleotides containing $m^{N4,5}C$.

Substrate	***Oligo***	***Sequence, 5′–3′***
S1 (30 nt)	Fw-$m^{N4,5}C$	Fw: [Cy3]C*G*G*TGAAGTAC[$m^{N4,5}C$]AGGAAGCGATTTCG A*C*C*C
	Rev-$m^{N4,5}C$	Rev: G*G*G*TCGAAATCGCTTC[$m^{N4,5}C$]TGGTACTTCA*C*C*G
S2 (30 nt)	Fw-$m^{N4,5}C$	Fw: [Cy3]C*G*G*TGAAGTAC[$m^{N4,5}C$]AGGAAGCGATTTCG A*C*C*C
	Rev-Gcomp	Rev: G*G*G*TCGAAATCGCTTCCTGGTACTTCA*C*C*G
S5 (30 nt)	Fw-$m^{N4,5}C$	Fw: [Cy3]C*G*G*TGAAGTAC[$m^{N4,5}C$]AGGAAGCGATTTCG A*C*C*C
	Rev-Acomp	Rev: G*G*G*TCGAAATCGCTTCCTAGTACTTCA*C*C*G
S6 (30 nt)	Fw-$m^{N4,5}C$	Fw: [Cy3]C*G*G*TGAAGTAC[$m^{N4,5}C$]AGGAAGCGATTTCG A*C*C*C
	Rev-Ccomp	Rev: G*G*G*TCGAAATCGCTTCCTCGTACTTCA*C*C*G
S7 (30 nt)	Fw-$m^{N4,5}C$	Fw: [Cy3]C*G*G*TGAAGTAC[$m^{N4,5}C$]AGGAAGCGATTTCG A*C*C*C
	Rev-Tcomp	Rev: G*G*G*TCGAAATCGCTTCCTTGTACTTCA*C*C*G

Note: Lesion highlighted in yellow and pink, complementary base in green and cyan. * end protection

2.1.2 Control oligonucleotide substrates

Polydeoxynucleotides containing dUMP at a specific site were supplied fluorescently labeled at the 5′-end with Cy3 by Sigma-Aldrich (Table 4). The labeled strands were annealed to equimolar amounts of a complementary strand in 10 mM Tris-HCl, pH 8.0, 50 mM NaCl, and 1 mM EDTA by heating at 95°C for 5 min and gradual cooling of 1°C per min to 23°C for 2 h.

Table 4: The control DNA substrate oligonucleotides containing uracil.

Control substrate	***Oligo***	***Sequence, 5′–3′***
S8 (60 nt) U at 21st position	Fw-Cy3U60	Fw: [Cy3]TAGACATTGCCCTCGAGGTAUCATGGATCCGATTTCGACCTCAAACCTAGACGAATTCCG
	Rev-Gcomp60	Rev: CGGAATTCGTCTAGGTTTGAGGTCGAAATCGGATCCATGGTACCTCGAGGGCAATGTCTA
S9 (60 nt) U at 12th position	Fw-Cy3U60	Fw: [Cy3]CCCTCGAGGTAUCATGGATCCGATCGATCCGATTTCGACCTCAAACCTAGACGAATTCCG
	Rev-Gcomp60	Rev: CGGAATTCGTCTAGGTTTGAGGTCGAAATCGGATCGATCGGATCCATGGTACCTCGAGGG
S10 (30 nt) U at 12th position	Fw-Cy3U30	Fw: [Cy3]CCCTCGATGTAUCATGGATCCGATCGATCC
	Rev-G30comp	Rev: GGATCGATCGGATCCATGGTACATCGAGGG
S11 (60 nt) T at 24th position	Fw-G60comp	Fw: TAGACATTGCCCTCGAGGTACCATGGATCCGATGTCGACCTCAAACCTAGACGAATTCCG
	Rev-T60FAM	Rev: [FAM]CGGAATTCGTCTAGGTTTGAGGTTGACATCGGATCCATGGTACCTCGAGGGCAATGTCTA

Note: Lesion highlighted in yellow, complementary base in green; FAM, Fluorescein.

2.2 Enzymes

Table 5: Various enzymes used in the study.

Enzyme	*Catalog/Lot no/Supplier*	*Dissolved in*
Fpg	M0240S; 8000 units/ml (16.67 pmol/µl); lot No. 0031005, 0061402, 0061405, 0081610; New England Biolabs	20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 0.5 mM ethylenediaminetetraacetic acid (EDTA), 200 µg/ml bovine serum albumin (BSA), 50% glycerol
Nei (endonuclease VIII)	M0299S, M0299L; 10000 units/ml (2.08 pmol/µl); lot No. 0091109, 0091605, 0091611; New England Biolabs	10 mM Tris-HCl, pH 8.0, 250 mM NaCl, 0.1 mM EDTA, 50% Glycerol
Nth (endonuclease III)	M0268S; 10000 units/ml (0.695 pmol/µl); lot No. 0031311; New England Biolabs	10 mM Tris-HCl, pH 7.4, 250 mM NaCl, 1 mM dithiothreitol (DTT), 0.1 mM EDTA, 200 µg/ml BSA, and 50% Glycerol
hSMUG1	M0336S; 5000 units/ml (0.33 pmol/µl); lot No. 0011405, 0011408, 0011512 New England Biolabs	10 mM Tris-HCl, pH 7.4, 250 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 200 µg/ml BSA, 50% glycerol, 0.15% Triton® X-100
hTDG	with His-tag; 15 pmol/µl Gift from Prof. Primo Schär and David Schürmann (University of Basel, Switzerland)	50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol
hOGG1	with His-Tag; aminoacid 12–327; 5.4 pmol/µl; Gift from Prof. Magnar Bjørås (Rikshospitalet, Norway)	10 mM 2-ethanesulfonic acid (MES), pH 6.0, 50 mM NaCl and 10 mM β-mercaptoethanol
hUNG	hUNGΔ84 with His-tag; 823 pmol/µl; Gift from Prof. Geir Slupphaug (NTNU, Norway)	20 mM Tris-HCl, pH 7.5, 60 mM NaCl, 1 mM EDTA, 1 mM DTT
hNEIL1	C-term His tag; full length, 4.49 pmol/µl; Gift from Prof. Magnar Bjørås (Rikshospitalet, Norway)	10 mM Tris-HCl, pH 7.0, 50 mM NaCl and 10 mM β-mercaptoethanol and 20% glycerol
hNEIL2	C-term His tag; full length, 0.2 µg/µl (5.3 pmol/µl), Gift from Prof. Magnar Bjørås (Rikshospitalet, Norway)	10 mM Tris-HCl, pH 7.0, 50 mM NaCl and 10 mM β-mercaptethanol and 20% glycerol
hNEIL3	C-term His tag; amino acid 1-301, 8.63 pmol/µl; Gift from Prof. Magnar Bjørås (Rikshospitalet, Norway)	10 mM MES, pH 6.0, 50 mM NaCl, and 10 µM β-mercaptoethanol
Ung	M0280S; 5000 units/ml (1.95 pmol/µl); lot No. 0121611; New England Biolabs	10 mM Tris-HCl, pH 7.4, 50 mM KCl, 1 mM DTT, 0.1 mM EDTA, 0.1 mg/ml BSA, 50% Glycerol
UGI (Uracil glycosylase inhibitor)	M0281L; 2000 units/ml; lot No. 0021203; New England Biolabs	10 mM Tris-HCl, pH 7.4, 50 mM KCl, 1 mM DTT, 0.1 mM EDTA, 200 µg/ml BSA, 50% Glycerol

2.3 Glycosylase Activity Assay

$m^{N4,5}C$-ssDNA, or with $m^{N4,5}C$-dsDNA where $m^{N4,5}C$ was placed opposite G, C, A or T; called ($m^{N4,5}C$:G-DNA, $m^{N4,5}C$:C-DNA, $m^{N4,5}C$:A-DNA or $m^{N4,5}C$:T-DNA), respectively (Table 3); were incubated with purified DNA glycosylases at varying concentrations at 37°C in an appropriate reaction buffer (Table 6) in a final volume of 20 µl, for the time indicated for each experiment (stated in legend to the figures). For the positive control to show the enzyme as active, enzymes were incubated with the control substrate DNA containing dUMP (S8, S9 or S10; Table 4) were also incubated with or without enzyme at the same condition with the reaction buffer specific for each enzyme (Table 6). Fpg, Nei, Nth, hOGG1, hNEIL1, hNEIL2, and hNEIL3, which are bifunctional enzyme were shown active by AP lyase activity (Tchou et al., 1994; Zharkov et al., 2002; Saito et al., 1997; Shinmura et al., 1997; Katafuchi et al., 2004; Krokeide et al., 2013). hSMUG1 and hTDG were shown active by the removal of uracil from the DNA followed by NaOH/heat mediated incision (Nilsen et al., 2001; Barrett et al., 1998). In the reaction for negative controls, 1× reaction buffer specific for each enzyme was added.

The reactions were terminated by the addition of (Stop solution) 20 mM EDTA, 0.5% (w/v) sodium dodecyl sulfate (SDS), and 10 µg proteinase K, incubated for 10 min at 37°C. EDTA present in the higher concentration chelates the divalent cations like Mg^{++} that is necessary for the DNAse I activity (Campbell et al., 1980). SDS binds strongly to the proteins causing the surfactant-induced denaturation and stripping them from the DNA and making them more accessible to the protease (Ananthapadmanabhan et al., 1975). Proteinase K, which is stable over a wide range of pH is a broad-spectrum serine protease which can inactive the proteins in the native form, digests the contaminating proteins and nucleases (Ebeling et al., 1974).

Subsequently, the DNA was ethanol precipitated at –20°C overnight, following the addition of 16 µg tRNA and 0.1 M Sodium acetate (CH_3COONa) (Leiros et al., 2007). The nucleic acids are soluble in water due to hydrate shells because of the high dielectric constant of water. Polar liquid with a low dielectric constant like ethanol was chosen to precipitate the DNA as it depletes the hydrate shells for dissolution, then the Coulomb force of attraction increases between the sodium ion and negatively charged DNA backbone, neutralizing the DNA and precipitating it out. tRNA was used as a carrier for the nucleic acid precipitation as it is insoluble

in ethanol forming the precipitate that helps to track the nucleic acid by bulk. Moreover, it forms the visible pellet during centrifugation helping to remove supernatant easily and makes the precipitation quantitatively more efficient (Zumbo, 2012). The DNA pellets were collected by centrifugation (Eppendorf, 13,000 rpm, 4°C, 30 min) and washed in 70% ethanol (–20°C).

Table 6: Reaction mix and buffers for the assays.

Enzyme	***Substrate DNA***	***Reaction buffer (1×) and mixture***	***Positive control***
Fpg	$m^{N4,5}$C-ssDNA, $m^{N4,5}$C-dsDNA	10 mM Bis-Tris-Propane-HCl, pH 7, 10 mM $MgCl_2$, 1 mM DTT, and 0.1 mg/ml BSA (NEB)	AP:G
Nei, hNEIL1	$m^{N4,5}$C-dsDNA	10 mM Tris-HCl, pH 8.0, 75 mM NaCl, and 1 mM EDTA (NEB)	AP:G
Nth	$m^{N4,5}$C-dsDNA	20 mM Tris-HCl, pH 8, 1 mM EDTA, 1 mM DTT, and 0.1 mg/ml BSA (NEB)	AP:G
hSMUG1	$m^{N4,5}$C-ssDNA, $m^{N4,5}$C-dsDNA	45 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.5, 0.4 mM EDTA, 2% (v/v) glycerol, 0.1 mg/ml BSA, 1 mM DTT, and 70 mM KCl (Laboratory protocol)	U:G, ssU
hTDG	$m^{N4,5}$C-dsDNA	50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM DTT, 0.1 mg/ml BSA and 0.5 unit UGI (Hardeland et al., 2000)	T:G, U:G
hOGG1	$m^{N4,5}$C-dsDNA	50 mM NaCl, 10 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 1 mM DTT, and 0.1 mg/ml BSA (NEB)	AP:G
hNEIL2	$m^{N4,5}$C-dsDNA	10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 50 mM NaCl, and 0.1 mg/ml BSA (Katafuchi et al., 2004)	AP:G
hNEIL3	$m^{N4,5}$C-ssDNA, $m^{N4,5}$C-dsDNA	50 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS), pH 7.5, 1 mM EDTA, 5% glycerol and 1 mM DTT (Krokeide et al., 2013)	ssAP

Note: NEB, the reaction buffer was prepared as mentioned in New England Biolabs catalog.

The dried pellets were resuspended in 10 µl of 0.1 M NaOH or dH_2O and heated at 90°C for 10 mins. For bifunctional enzymes, $m^{N4,5}$C-DNA glycosylase activity was determined by NaOH-

mediated (0.1 M final concentration) incision of the resulting abasic site (90°C for 10 min), whereas incision of the DNA without alkaline treatment measured the $m^{N4,5}C$-DNA incision activity. Following the alkaline/heat treatment of the abasic sites created by hSMUG1 and hTDG, $m^{N4,5}C$-DNA glycosylase activity was determined.

U:G-DNA (S9 or S10) was treated with 1 pmol of hUNG or 1.95 pmol of Ung (as indicated in figures) to create the AP site. In the case of hNEIL3, single-stranded AP site was created on U-DNA (Forward strand of S10) with the treatment of 1.95 pmol Ung and ssAP lyase activity was shown. hUNG or Ung was shown active by NaOH mediated incision and no enzyme was added in negative control. In the case of mono-functional enzymes like hSMUG1 and hTDG, the AP site generated by the same enzyme on U:G-DNA (S8 or S9) was determined by NaOH-mediated (0.1 M final concentration) incision of the resulting abasic site.

Before loading into the denaturing PAGE, the samples were mixed with formamide gel loading buffer (80% formamide, 1 mM EDTA and 1% (w/v) blue dextran), heated for 5 min at 90°C, and chilled on ice immediately. Samples (20 µl) were prepared for electrophoresis by the addition of 10 µl of the loading buffer solution referred above followed by incubation at 90°C for 5 min to denature the DNA. After cooling on ice, a volume of 5 µl was subjected to denaturing PAGE [20% (w/v) polyacrylamide gel with 8 M urea or 7M if indicated] for time indicated in figures. The detailed protocol for the glycosylase activity assay is added in the Appendix A.

2.3.1 Protein characterization

2.3.1.1 Time-dependent protein function

1 pmol of $m^{N4,5}C$:C-DNA was incubated with 17 pmol of Fpg (lot No. 0061405) for 2, 5, 10, 30, 60 min at 37°C under the same condition as described in section 2.3. The DNA incubated with Fpg was studied for the excision/incision coupled activity. Also, the sample was treated with the NaOH and heat to observe only the glycosylase activity. Similarly, 1 pmol of $m^{N4,5}C$:G-DNA was incubated with 8.32 pmol of Nei (lot No. 0091605) for 2, 5, 15, 30, 45, and 60 min at 37°C. The DNA incubated with Nei was studied for the excision/incision coupled activity. Subsequently, the sample was subjected to electrophoresis for 2 h at 200 V in denaturing PAGE and the time-dependent activity for the proteins were determined.

The data obtained were analyzed using the following equation (Leiros et al., 2007).

$$[P] = E_o[1-\exp(-k_{obs}t)]$$

Where E_o represents the amplitude of the exponential phase or total initial DNA concentration and k_{obs} is the observed rate constant associated with that process.

A proposed kinetic scheme for the enzyme is as follows:

$$\text{Enzyme} + (\text{DNA})_s \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{Enzyme} \cdot (\text{DNA})_s \xrightarrow{k_2} \text{Enzyme} + (\text{DNA})_p$$

The initial binding of substrate is described by k_d which is defined as k_{-1}/k_1 and k_2 defines the overall rate constant for total conversion. If the release of enzyme from product DNA and its contribution is disregarded under this condition then rate constant is under pseudo-first order conditions. When [Enzyme] > [DNA], with assumption that enzyme-substrate binding is in a rapid equilibrium *(i.e.* $k_{-1} >> k_2$*)*

$$k_{obs} = \frac{[E_o].k_2}{k_d + [E_o]}$$

since enzyme concentration is well above k_d, $k_{obs} = k_2$.

2.3.1.2 Protein concentration dependent function

1 pmol of m^{N4,5}C:C-DNA was incubated with 0.5, 1, 2, 5, 10, and 17 pmol of Fpg (lot No. 0061402) for 60 min at 37°C under the same condition as described in section 2.3. The DNA incubated with Fpg was treated with the NaOH and heat to observe only the glycosylase activity and without such treatment to study both excision/incision coupled activity. Similarly, 1 pmol of m^{N4,5}C:G-DNA was incubated with 2.08, 4.16, 6.24, 8.32, 10.4, and 12.48 pmol of Nei (lot No. 0091611) for 60 min at 37°C under the same condition as stated in 2.3. The DNA incubated with Nei was studied for the excision/incision coupled activity present. Subsequently, the samples were subjected to electrophoresis for 2 h at 200 V in denaturing PAGE and the protein concentration dependent activity of both protein were determined.

2.3.1.3 Substrate concentration dependent function

250 nM of Fpg (lot No. 0081610) and 416 nM of Nei (lot No. 0091605) were used to study the effects of substrate concentration on the activity of the protein. Different concentrations ranging from 6.25–500 nM of m^{N4,5}C substrate-DNA opposite C and G were incubated with Fpg and Nei, respectively, for 10 min at 37°C in the reaction conditions as mentioned above. After the reaction, processed and unprocessed DNA were separated by denaturing PAGE at 200 V for 2 h. All the reactions were triplicated for the validation of the data obtained.

Kinetic description of the activity of the enzyme is very important to understand how enzymes functions and enhances the rate of reaction in the organisms which defines the number of moles of product formed per second at a fixed concentration of enzyme. The rate of catalysis, V_o, which defines the number of moles of product formed per second at a fixed concentration of enzyme is almost linearly proportional to substrate concentration in small amount but is nearly independent if in a large amount (Berg et al., 2002).

Consider an enzyme that catalyzes the S to P by the following pathway:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E \cdot S \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P$$

The Velocity of the reaction is given by Michaelis-Menten equation,

$$V_o = (V_{max} \cdot [S])/(K_m + [S])$$

Where V_{max} is the maximum rate of reaction when the enzyme is saturated with the substrate and K_m is the concentration when V_{max} is half of the maximum velocity V_{max}. If an enzyme obeys the Michaelis-Menten kinetics, in the plot of initial velocity as a function of substrate concentration, V_{max} is approached asymptotically (Berg et al., 2002).

2.4 Denaturing Polyacrylamide Gel Electrophoresis with Urea and analysis

Gel electrophoresis is a method where a charged molecule is separated according to charge or mass when driven through the sieve-like gel matrix by electrical current (Smith, 1984). Polyacrylamide Gel Electrophoresis (PAGE) is mainly used for separating the protein whereas agarose gel electrophoresis for separating the DNA. However, to separate the DNA of short length up to 500 nucleotides, PAGE is preferred because of its low pore size (Summer et al., 2009). Polyacrylamide gels are made by crosslinking the polymer network of acrylamide and bisacrylamide by polymerizing agent ammonium persulfate (APS), where N,N,N′,N′-tetramethylenediamine (TEMED) is used as a catalyst in the polymerization reaction. The pore size formed by crosslinking in the polyacrylamide gel is inversely proportional to the amount of acrylamide used (Johnson, 1979). The DNA or RNA in urea PAGE or denaturing PAGE which employs 6–8 M urea, denatures the secondary structures of DNA and separates them on a gel matrix based on their molecular weight. The mobility of the sample is based on the acrylamide concentration where a higher percentage of polyacrylamide resolves lower molecular weight fragments. The DNA heated with the formamide are denatured into the single-strands and are driven by the electrical current into the positively charged electrode – anode. The DNA with high molecular mass are retained by gel matrix and moves slowly

whereas low molecular weight DNA travels far more quickly during separation (Summer et al., 2009).

The urea gel was prepared by mixing the 20% (w/v) polyacrylamide (Saveen Werner AB), 8 M urea (Invitrogen), 1× Tris-Borax-EDTA buffer. 10% (w/v) APS (Bio-Rad) and TEMED (Invitrogen) were used for the polymerization of the polyacrylamide. The mixture was then poured into the gel cassette with gap thickness 0.75 mm where the mixture gets inside by capillary action. The gels were polymerized for 30 mins and later assembled for Bio-Rad, Mini Protean Tetra cell system. The samples were separated in PAGE for 2 h at 200V or as indicated in the figures.

2.4.1 Denaturing condition optimization for urea PAGE

During the separation, we came across with various problems in denaturation and thereby creating the double bands on gel in some cases (Appendix Figures: B9, B10). When the $m^{N4,5}C$-DNA was hybridized with reverse strand opposite the canonical bases (S2, S5, S6, and S7), problem with the separation and DNA denaturation was observed, producing the double bands in the gel. Firstly, the question was whether the DNA was hybridized well and the double band observed in PAGE was either single or dsDNA. To figure this out, ss and dsDNA were electrophoresed in native PAGE and it was observed that the DNA was indeed hybridized into double strand (Appendix Figure B17). The movement of ss and dsDNA in native PAGE were different confirming the substrate DNA was hybridized well opposite canonical bases. Then, it was confirmed that the appearance of the double band was not the fault in hybridization but some other issues. In the experiments conducted with hSMUG1 and hTDG, there were the presence of double bands in separation, however, we used $m^{N4,5}C$ ssDNA as a control to show that the double bands seen were not the product of any enzymatic cleavage but the denatured and undenatured DNA in the gel (Appendix Figures B9, B10).

The denaturing gel used for the separation was Tris-Taurine-EDTA (TTE) buffer system with 7M urea for denaturation. So, it was questioned whether the denaturation capability of the PAGE was lower compromising the denaturation of DNA. Thus, the buffer system was changed to Tris-Borax-EDTA (TBE) system with an increase in the urea to 8M in order to increase the

denaturation. The change in the buffer system with increased urea concentration solved the denaturation problem of DNA, hence the absence of the double bands from the gel (Appendix Figure B18). Moreover, TBE system offers a high buffering capacity and suitable for separation of small DNA molecule (Miura et al.,1999) whereas TTE is a glycerol-tolerant buffer (Hirano et al., 2009) having high buffering capacity at a higher voltage.

2.5 Optimization of signal measurement

After electrophoresis, the gels were scanned using Typhoon Trio^{TM} (GE Healthcare), having three lasers, enabling excitation and visualization for a wide range of wavelengths. The fluorescently labeled DNA oligomers with Cy3 were excited at 532 nm and emission was measured at 580 nm using a band pass emission filter. The DNA labeled with Fluorescein was excited at 488 nm by an Argon ion laser and emission was measured at 520 nm using band pass emission filter. The emission produced from the cy3-tagged DNA in the gel was detected by the detector photomultiplier tube (PMT) and its voltage settings was adjusted to get the precise image with the desired grayscale for analysis. The calibration curve for the detection of the DNA was prepared to determine the optimal concentration that lies in the linear range of the detector for the quantitation process. The varying concentration of 0.125–5 pmol/µl of substrate DNA in the increasing order were prepared by serial dilution from 10 pmol/µl of the substrate. 1 µl of the sample was mixed with 9 µl of dH_2O and 10 µl of loading buffer and heated for 5 min. 5 µl of the sample was loaded on denaturing PAGE with urea. The loaded amount of analyte was ranging from 0.03125–1.25 pmol and this experiment was repeated for three parallels. Then, Electrophoresis was performed at 300V for 40 mins. The gel was scanned at various photomultiplier tube (PMT) voltage settings from 450 to 850 V. The images were saved and quantification was performed with ImageJ software (Schneider et al., 2012). The intensity of each concentration at various PMT voltage was quantified and plotted to fit a curve.

The PMT absorbs the photons from the emission through photocathode and releases the electrons due to the photoelectric effect. These electrons are amplified by the electrodes called metal channel dynodes and finally ending on anode or collection electrode to give a signal (Abramowitz et al., 2002). When there is the high flux of photons due to the increase in the concentration of an analyte, there is the tendency that the detector become saturated at high

fluorescence emissions. The calibration curve at various PMT voltage settings and different concentration of analyte showed that the signal was increasing linearly but saturation occurred at higher concentration. The signal saturated after 0.125 pmol at PMT 650 V whereas signal saturation occurred at 0.25 pmol when using PMT 450 and 550 V. This observation suggested that the measurement should be taken at 550V, as there is the largest range in linearity up to 0.25 pmol with minimal error (Appendix Figure C1A). As mentioned in the user guide, the high-intensity samples sometimes saturate the system, so the PMT voltage setting should be decreased to optimize the high-intensity signals into the linear range of the instrument. Likewise, at PMT voltage higher than 750 V, signal saturated even at a lower concentration with a high amount of variation (Appendix Figure C1B), however, below 550 V it was hard to detect the signal in lower concentrations.

The signal deviation was minimal in PMT setting at 550V with the linearity of detection that was optimal for our measurement and quantification of substrate DNA. Also, for the higher concentrations of samples, a strategy was made to dilute the samples to get the data at linear range of calibration curve. I made a dilution to make sure that the amount being quantified lie within the linear range of the detector. Higher concentrations of the reaction samples were diluted to 1 pmol (0.25 pmol in the gel as 5 µl out of 20 µl was loaded) and then used for denaturing PAGE while maintaining the linear detection limit (Appendix Figure B8: C, F and I). This strategy helped us to reduce a lot of instrumental errors in the measurements. The two-different curve fitting of enzyme kinetics on substrate were compared with and without dilution. The curves suggested that the samples with dilution reduced errors due to detector saturation (Appendix Figure C2).

3 Results

3.1 DNA glycosylase activity on m^{N4,5}C-DNA

The determination of the specificity of different *E. coli* and human DNA glycosylases were prepared using the oligonucleotide substrate containing a modified m^{N4,5}C base at a defined position in the same sequence context (Figure 8A). The oligonucleotide was annealed to the reverse complementary sequence containing G, C, A or T opposite the modified base lesion. Using the enzyme that was able to excise more than 50% of the U:G substrate or incise an AP-DNA was used as the reference. Mono-functional glycosylases like Ung and hUNG which excises the aberrant or damaged bases, which we used as a control results in the alkali liable apurinic/apyrimidinic (AP) site which is cleaved by NaOH in presence of heat by a β/δ-elimination reaction (Bailly et al., 1987). However, bifunctional DNA glycosylases like Nei, Fpg, hOGG1, hNEILs, and Nth cleaves the resulting AP-DNA following base excision, and these consecutive activities are thus monitored without NaOH/heat treatment. Cleaved is separated from uncleaved DNA by PAGE under denaturing conditions (Figure 8C). One-hour incubation with *E. coli* Fpg cleaved m^{N4,5}C opposite C very efficiently and marginally opposite T, however, no activity was detected opposite A or cognate G (Figure 8D). Fpg was incubated with the ssDNA containing m^{N4,5}C base lesion for an hour but was without any activity. Similarly, *E. coli* Nei incubated for an hour efficiently incised the m^{N4,5}C opposite cognate G but there was only marginal incision with opposite A and T. However, the activity of Nei against m^{N4,5}C opposite C was not detected (Figure 9A). The assays were triplicated for the verification and mean percentage of DNA processed relative to the bases and the standard deviation was plotted in the bar diagram (Figure 8B, 9B). This contrasts with *E. coli* Nth (Appendix Figure B12) where m^{N4,5}C excision activity was absent.

The human DNA glycosylase enzymes tested for the possible m^{N4,5}C cleavage activity *i.e* hSMUG1(Appendix Figure B9), hTDG (Appendix Figure B10), hOGG1 (Appendix Figure B11). Similarly, the human homolog of Nei, hNEIL1, hNEIL2, and hNEIL3 (Figure 9C) were also without detectable DNA glycosylase activity for m^{N4,5}C at the various concentrations and incubation conditions employed. The activities of all human and bacterial glycosylases on m^{N4,5}C are listed (Table 7).

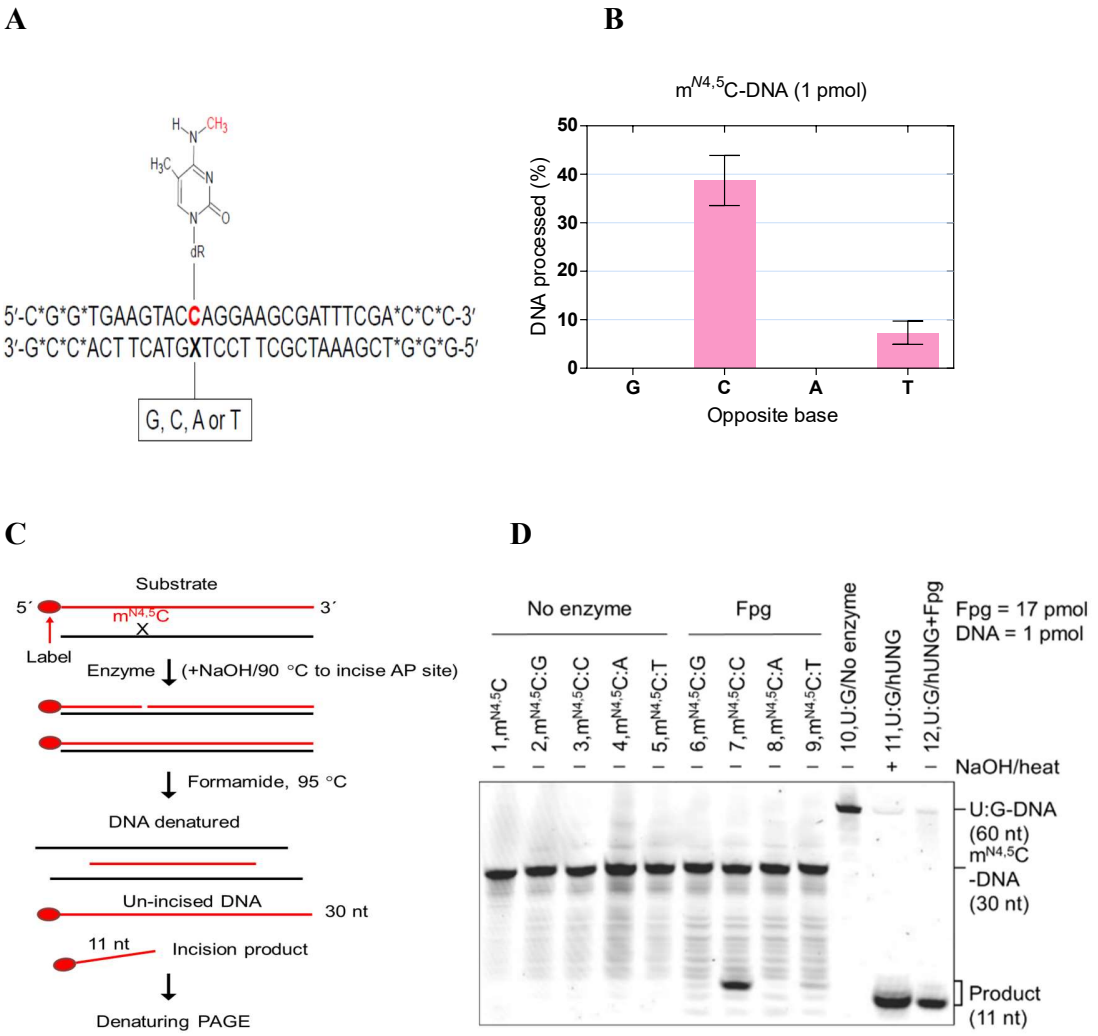


Figure 8: *E. coli* Fpg excises m^{N4,5}C base opposite cytosine in DNA.
(**A**) The nucleotide sequence of substrate oligomer. (**B**) The percentage incision of the m^{N4,5}C opposite canonical bases of three parallels where error bars represent SD. (**C**) The schematic representation of the glycosylase activity assays. (**D**) Substrate DNA (1 pmol, 30 nt) were incubated without enzyme (1–5), with Fpg 17 pmol (lane 6–9) at 37°C for 60 min in (10 mM Bis-Tris-Propane-HCl, pH 7, 10 mM $MgCl_2$), 1 mM DTT, 0.1 mg/ml BSA (final volume 20 µl). U:G-DNA (1 pmol, 60 nt) was incubated without enzyme (lane 11) and with hUNG 1 pmol, latter followed by treatment with NaOH and heat which were used as negative and positive control for hUNG activity, respectively to generate the AP-DNA to reveal Fpg activity (lane 12). The incision product was separated using denaturing PAGE at 200 V for 2 h.

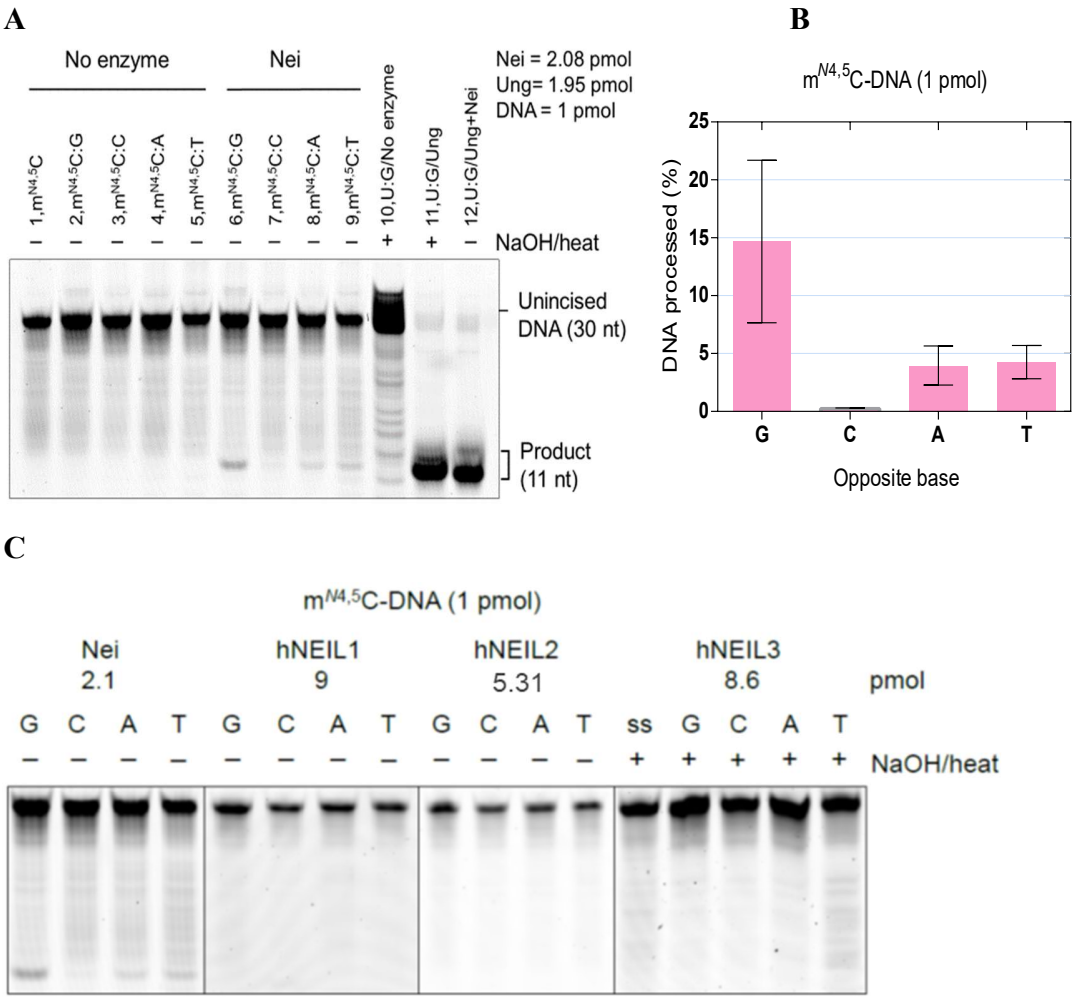


Figure 9: *E. coli* Nei incises m^{N4,5}C when placed opposite guanine in DNA.
(**A**) Substrate DNA (1 pmol, 30 nt) were incubated without enzyme (1–5), with Nei 2.08 pmol (lane 6–9) at 37°C for 60 min in 10 mM Tris-HCl, pH 8.0, 75 mM NaCl, and 1 mM EDTA (final volume 20 µl). U:G-DNA (1 pmol, 30 nt) was incubated without enzyme (lane 10) and with Ung 1.95 pmol (lane 11), latter followed by treatment with NaOH and heat which were used as negative and positive control for Ung activity, respectively to generate the AP-DNA to reveal Nei activity (lane 12). The incision product was separated using denaturing PAGE at 200 V for 2 h. (**B**) The extent of activity measured in different base pairing of substrate DNA and presented as mean±SD. (**C**) DNA substrate were incubated with the indicated amount of Nei, hNEIL1, hNEIL2 and hNEIL3 at 37°C for 1 h (final volume, 20 µl). See A**,** and appendix Figures B13, B14 and B15 for experimental details and activity controls for the different enzyme preparations. The m^{N4,5}C was placed opposite the four normal bases as indicated, in the 4th lane; ss**,** single-stranded DNA.

Table 7: Activity of different glycosylases on $m^{N4,5}C$ DNA paired with canonical bases

DNA pairings/ Glycosylases	**$m^{N4,5}C$**	**$m^{N4,5}C$:G**	**$m^{N4,5}C$:C**	**$m^{N4,5}C$:A**	**$m^{N4,5}C$:T**
Fpg	-	-	+++	-	+
Nei	NA	+++	-	+	+
Nth	NA	-	-	-	-
hOGG1	NA	-	-	-	-
hSMUG1	-	-	-	-	-
hTDG	NA	-	-	-	-
hNEIL1	NA	-	-	-	-
hNEIL2	NA	-	-	-	-
hNEIL3	-	-	-	-	-

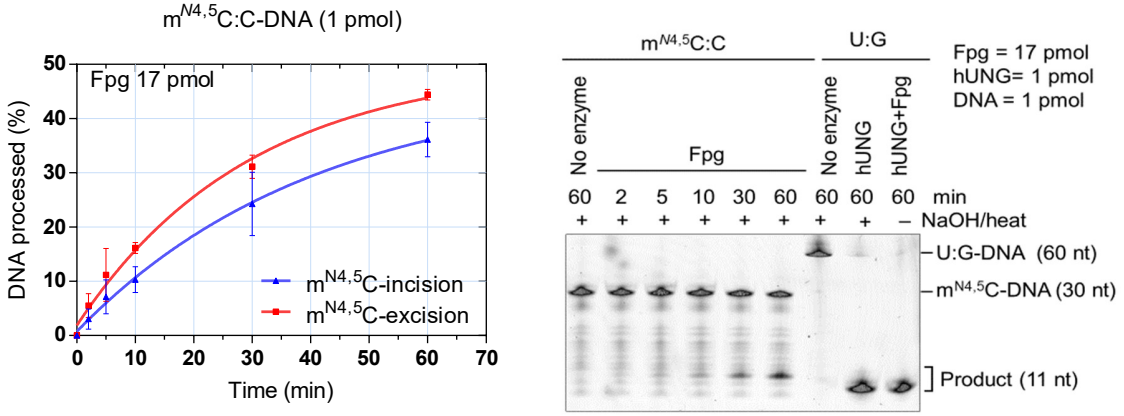
Note: +++ high activity; ++ medium activity; + low activity; - no activity; NA, not available

3.2 Protein characterization

3.2.1 Time-dependent activity of Fpg and Nei on $m^{N4,5}C$-DNA

Time dependence of excision was measured by incubating $m^{N4,5}C$:C-DNA and $m^{N4,5}C$:G-DNA with Fpg and Nei protein, respectively, for various time as stated in the methods. The excision of the lesion was assessed by the treatment of the samples with alkali/heat treatment and the incision/excision coupled activity was assessed without such treatment. Subsequently, the samples were subjected to electrophoresis for the separation of the incised and un-incised DNA. The data was obtained and fitted to one-phase association model for the calculation of single turnover rate for base removal using GraphPad Prism (Figure 10).

A



B

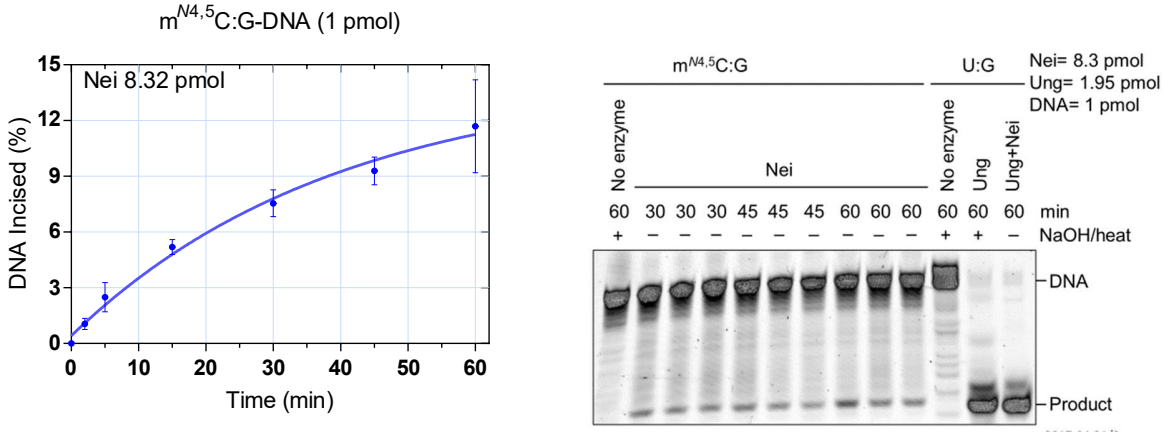


Figure 10: Excision and incision of m^{N4,5}C-DNA by Fpg and Nei protein as a function of time. (**A**) Fpg (17 pmol) was incubated with m^{N4,5}C:C DNA (1 pmol) over various time interval and its glycosylase activity (red) and coupled glycosylase/lyase activity (blue) was plotted (Appendix figure B2). A part of typical experiment for excision is shown in right panel (top). (**B**) Nei (8.32 pmol) was incubated with m^{N4,5}C:G substrate DNA (1 pmol) over various time interval for the plot of coupled glycosylase/lyase activity (Appendix figure B7). A part of typical experiment for incision is shown in right panel (bottom). The experiments were repeated three times for both proteins and each point represents mean value±SD. The values were fitted to a one phase association model for the calculation of k_{obs} (min^{-1}) using GraphPad Prism.

Table 8: Single turnover rate for the excision and incision of m^{N4,5}C in DNA

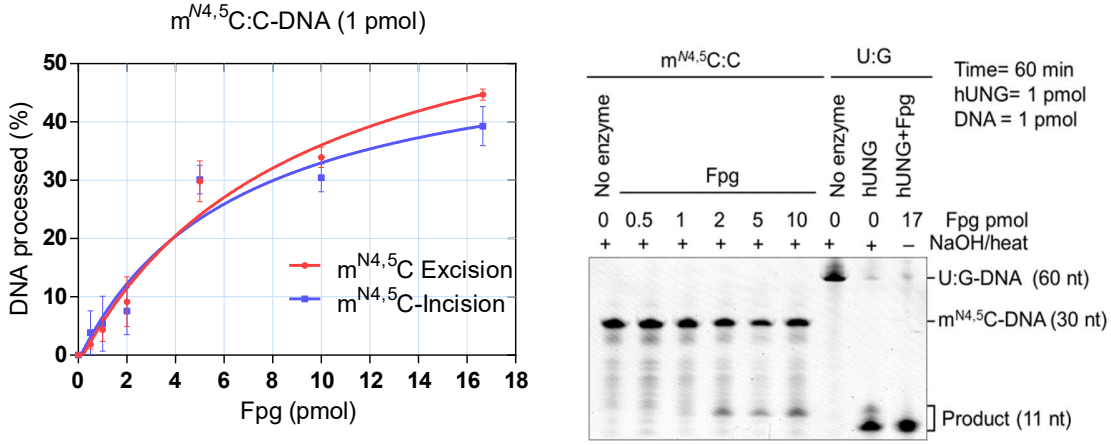
Enzyme	***k_{obs} (incision) min^{-1}***	***k_{obs} (excision) min^{-1}***
Fpg 17 pmol *m^{N4,5}C:C*	0.024±0.008	0.034±0.006
Nei 8 pmol *m^{N4,5}C:G*	0.025±0.008	-

For Fpg, the single turnover rate for base excision was 0.034±0.006 min^{-1}, and 0.024±0.008 min^{-1} for incision/excision coupled activity. The time dependence of excision as a function of the Fpg was greater than the coupled incision/excision (Figure 10A). Compared to previously determined k_{obs} for the coupled incision/excision of oxo^8G, Gh and Sp paired opposite C by Fpg showed: $k_{obs} > 7$ min^{-1}, 1.1–1.8 min^{-1} and 1.1–1.5 min^{-1} respectively. The rate of coupled incision/excision of $m^{N4,5}C$ opposite C measured by us is about 300 folds slower than oxo^8G and 60 times slower than Gh and Sp removal (Leipold et al., 2000). Likewise, the single turnover rate for incision/excision coupled activity of Nei was 0.025±0.008 min^{-1} which is very similar to the Fpg incision coupled activity indicating identical activity on dimethylated cytosine (Figure 10B). However, single turnover kinetics should be performed for different enzyme concentrations to validate the observed k_{obs}.

3.2.2 Protein concentration dependent Fpg and Nei activity on $m^{N4,5}C$

Excision and incision/excision dependence of two different base pairing of $m^{N4,5}C$-DNA as the function of the amount of enzyme is shown (Figure 11). The increasing amounts of the Fpg and Nei were incubated with $m^{N4,5}C$:C and $m^{N4,5}C$:G DNA, respectively, as stated in methods. The absolute amount of excision and incision products increased with the increasing amount of Fpg protein until 5 pmol of enzyme but at higher levels of enzyme concentration, excision activity exceeded the incision activity (Figure 11A) which accords with the time dependence graph of Fpg (Figure 10A). Similarly, the absolute amount of incision products increased with increasing amount of Nei up to 8.32 pmol but the activity decreased with the further increase in the enzyme concentration (Figure 11B). The reaction mixtures were prepared in three or more parallels and each data point on the graph represents mean value. The difference of the variance between the replicates is represented by the SD bars in the curve.

A



B

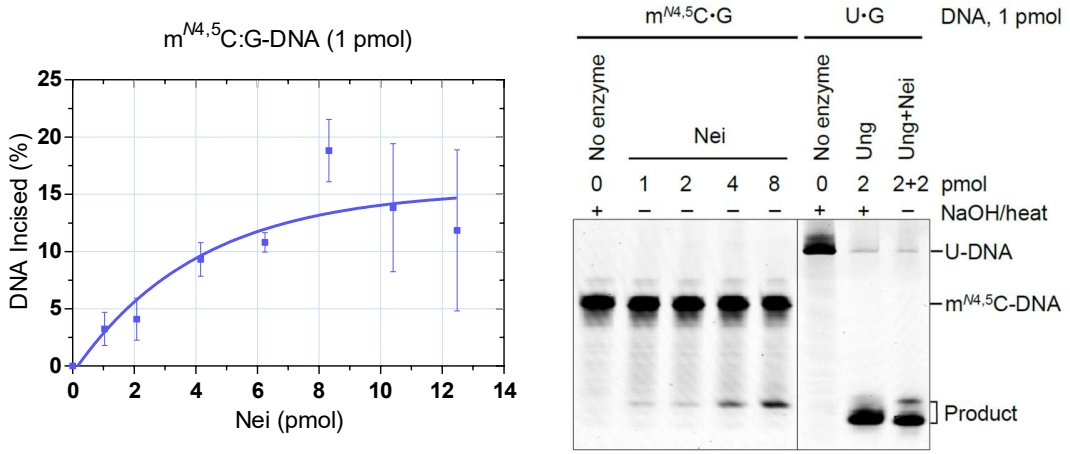


Figure 11: Excision and incision of m^{N4,5}C-DNA as a function of enzyme concentration.
(**A**) Various concentrations of Fpg was incubated with m^{N4,5}C:C-DNA (1 pmol) for 1h at 37°C as described in the methods. Excision and incision activity was measured for the activity of the protein (Appendix figure B3). A part of typical experiment for incision activity of Fpg is shown in the right panel. (**B**) Various concentrations of Nei was incubated with 1 pmol of m^{N4,5}C:G-DNA for 1h at 37°C as described in methods (Appendix figure B6). Incision activity was measured for the activity of the protein. A part of typical experiment for incision activity of Nei is shown in the right panel. Each data point represents the mean (±SD) of the values obtained from the analysis of three or more independently prepared samples.

3.2.3 Substrate dependent excision of m^{N4,5}C-DNA kinetics

The kinetic parameters were determined by measurement of excision for Fpg and incision for Nei at eight different concentrations of substrate DNA ranging from 6.25 –500 nM and the total amount of protein remaining constant in each assay. The optimal concentrations determined for performing the kinetics were 250 nM for Fpg and 416 nM for Nei. Excision and incision of the m^{N4,5}C from m^{N4,5}C:C by Fpg and from m^{N4,5}C:G by Nei, respectively, followed Michaelis-Menten kinetics (Figure 12). Kinetics constants were calculated using a computer program with a nonlinear least-square analysis of the data using GraphPad Prism. The initial velocities were estimated as the products incised or excised as a function of the incubation time. Since the time curve of Fpg and Nei showed the linearity around 10 min, initial velocity was measured at that time. The data taken at the higher time, the velocity was slower and attained saturation state, was not optimal for measuring the Michaelis-Menten kinetics (Appendix Figure C3).

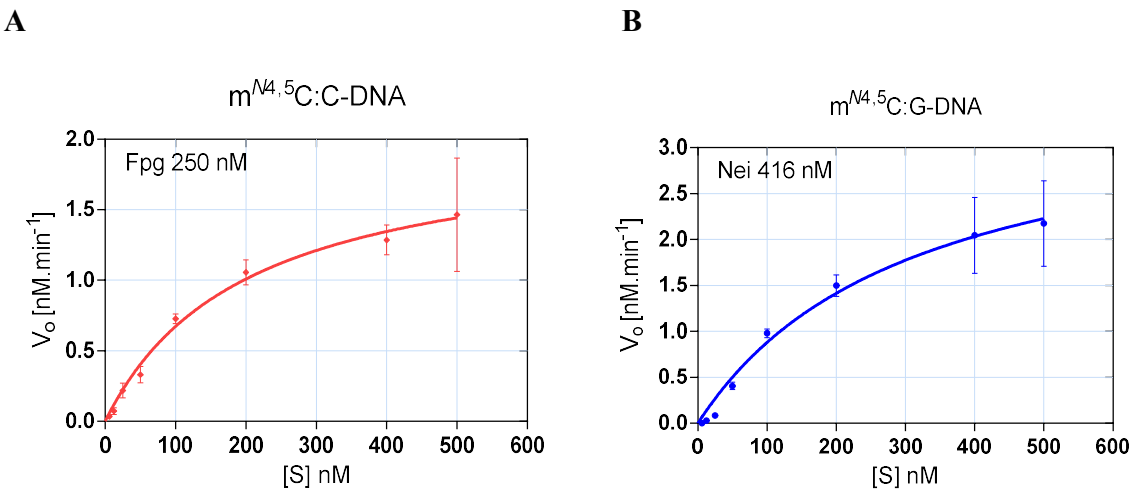


Figure 12: Michaelis-Menten kinetics of Fpg and Nei on m^{N4,5}C-DNA.
(**A**) Fpg (5 pmol, 250 nM final concentration) was incubated with 6.25 to 500 nM of m^{N4,5}C:C-DNA for 10 min and the initial velocity for excision at specific concentrations were determined (Appendix figure B4). (**B**) Nei (8.32 pmol, 416 nM final concentration) was incubated with 6.25 to 500 nM of m^{N4,5}C:G-DNA for 10 min and the initial velocity for incision at specific concentrations were determined (Appendix figure B8). The initial velocity is plotted as a function of substrate concentration and the curves were fitted for nonlinear least-square analysis using GraphPad Prism. Abbreviations: [S], substrate concentrations; V_o, initial velocity.

Table 9: Specificity constants of the DNA lesions excised by Fpg and Nei.

Lesion (**bold**, if aromatic)	Opposite base	K_m (nM)	k_{cat} (min^{-1})
Fpg protein			
$m^{N4,5}C$	C	201±47	0.0080±0.0007
fapyG		7	0.2
fapyA[i]		275–1632	0.026–0.106
fapyG[i]		1037–7063	0.067–0.625
oxo^8G[i]		861–2405	0.082–0.178
oxo^8dG	dC	14±6	0.13±0.01
	dT	2.5±0.7	0.68±0.04
	dG	5.7±1.4	0.60±0.04
	dA	190±74	0.10±0.02
AP site		4	2.5
Nei protein			
$m^{N4,5}C$	G	311±87	0.0087±0.0011
dhC	G	1010±142	0.031±0.004
fapyA	T	1828±172/635±77	0.0937±0.0083/0.0333±0.0033
Th^5	A	1838±61	0.037±0.001
Uh^5	G	1473±226	0.025±0.004
dHU	A	90±10, 160±120	10, 0.035±0.015
	C	6.6±2.5	0.49±0.05
	G	8.5±3.2	4.3±0.5, 21.0±1.2
	T	3.5±1.3	0.19±0.02
Tg	A	3115±550/1109±73	0.078±0.037/0.045±0.001
Tg_A	A	16±14	1.4±0.7
	C	4.7±2.9	0.69±0.18
	G	2.3±1.2	3.4±0.6
	T	4.6±2.6	0.67±0.16
Tg_C	A	3.2±1.6	2.1±0.5
	C	4.4±2.7	2.5±0.8
	G	2.3±0.8	2.0±0.3
	T	4.4±1.6	2.9±0.6
Ug	G	128±17/2098±154	0.0123±0.0007/0.1127±0.0077
Hmh[1]	A	2569±476/1707±442	0.29±0.05/0.141±0.036
oxo^8G	C	505±50, 680±210	1.1, 0.95±0.10
	A	415±25, 610±110	1.4, 0.47±0.04
	G	357±20, 580±200	2, 0.79±0.11
	T	1300±300	1.4±0.2
AP site	A	0.083±0.015	0.64±0.02
	G	0.11±0.03	0.53±0.13, 47.4±1.8

Note: [1], Ring-contracted; [i], irridated DNA; Tg_A, (5*S*,6*R*)-5,6-dihydroxydihydropyrimidine-2,4(1*H*,3*H*)-dione; Tg_C, (5*R*,6*S*)-5,6-dihydroxydihydropyrimidine-2,4(1*H*,3*H*)-dione; Abbreviations in Table 2; our observations for Nei in black, as well as (Hazra et al., 2000), (Kropachev et al., 2006), (Kladova et al., 2017), and (Tchou et al., 1994) employed assays which measured the consecutive base excision + DNA incision reaction; our observations for Fpg in black, as well as (Dizdaroglu et al., 2001), (Graves et al., 1992), and (Karakaya et al., 1997) used assay only determining the base excision reaction.

The Kinetic analysis of $m^{N4,5}C$ excision with Fpg indicated the Michaelis-Menten behavior with a K_m of 201±47 nM and V_{max} of 2.02±0.2 nM.min^{-1} (Figure 12A). The kinetic analysis of Nei, likewise Fpg, showed the Michaelis-Menten behavior (Figure 12B), indicating a K_m value of 311±87 nM and V_{max} of 3.61±0.5 nM.min^{-1}. k_{cat} was determined for Fpg with 0.0080±0.0007 min^{-1} and Nei with 0.0087±0.0011 min^{-1}, indicating Fpg is slightly slower in the conversion of the substrate into the product.

These values are compared to previously published values for Fpg activity towards other DNA substrates as presented in Table 9. The binding affinity of Fpg for $m^{N4,5}C$ is 29 folds lower compared to the excision of fapyG from a fapyG-poly[(dGC):(dGC)] substrate (Graves et al., 1992). However, another study determined these constraints for excision of oxo^{8}G, fapyG and fapyA from DNA treated by oxidative agents and irradiation, indicating the binding affinity for excision of $m^{N4,5}C$ from dsDNA are higher, within and near to the ranges determined for other Fpg substrates (Karakaya et al., 1997). The binding affinity for the oxo^{8}dG opposite dG, dC, and dT in 23 mer oligonucleotide is higher as compared with our substrate. However, opposite dA, the binding affinity is very close to our observed value for $m^{N4,5}C$ opposite C (Tchou et al., 1994).

Compared to the previously determined parameters for the incision of dHU and oxo^{8}G by Nei, the binding affinity for $m^{N4,5}C$ observed is 3 folds lower than for dHU; whereas within oxo^{8}G with oxo^{8}G:C, oxo^{8}G:A, oxo^{8}G:C binding affinity is within the range determined or higher. However, k_{cat} is lower compared with the enzyme activity with other known substrates (Hazra et al., 2000). Similarly, in another study of Nei activity on the fapyG and Tg from the DNA treated by γ-radiation showing the specificity constants observed by us are within the range determined for fapyG and Tg (Dizdaroglu et al., 2001). However, the turnover number for the incision and excision of both Nei and Fpg, respectively, are very low compared to the known substrates, indicating a slow reaction (Table 9). Nevertheless, there is so much variation in the published values for the known substrates, comparison of parameters determined in diverse experimental conditions have limited significance.

4 Discussion

4.1 m^5C damage, mutagenicity and repair

The major epigenetic mark in eukaryotes including mammalian DNA is m^5C, which mostly exists in CpG sequences throughout the genome (Klungland et al., 2016). In prokaryotic DNA m^5C shares importance with $m^{N4}C$ and $m^{N6}A$ as normal methyl base modifications (Ehrlich et al., 1987); participating in certain biological functions such as protection against DNA hydrolysis, DNA restriction and DNA repair. Recently, $m^{N6}A$ was discovered as the second methylated epigenetic base in eukaryotes including mammals (Wu et al., 2016). All three base modifications are products of enzyme-catalyzed transfer of a methyl group from SAM to the cognate unmodified base.

Most human promoters contain CpG islands, which are genomic regions that have a significant overrepresentation of the CpG sequence where methylation of cytosine takes place. There is a strong correlation between promoter methylation and gene repression, also called gene silencing, and vice versa (Klungland et al., 2016). The m^5Cs constitute 4–5% of the total cytosine content and are mutational hotspots in mammalian cells being damaged two-fold more frequently than cytosine in identical sequences (Madugundu et al., 2014). The disposition of m^5C is highly regulated in the cells and any aberration in m^5C profiles could lead to deregulation of gene expression resulting in tumorigenesis or improper development besides cytotoxicity and mutagenicity as for the common unmethylated bases (Klungland et al., 2016).

As stated previously, CpG sequences are hotspots of mutations induction as m^5C are particularly subjected to damage by spontaneous deamination to thymine and is deaminated three folds faster than cytosine (Ehrlich et al., 1986; Wang et al., 1982). Despite being one of the four genuine bases, thymine has been reported to be one of the major base damages responsible for one third transition mutations accountable for genetic diseases and cancers in humans (Um et al., 1998). During DNA replication, if there is a G:m^5C base pair deaminated to a G:T base pair, an A would be introduced opposite T during the next replication event resulting in a G:C $\rightarrow$ A:T transition. The thymine generated through the deamination of m^5C is correctly replaced with the cytosine before the replication by species-specific repair systems.

In *E. coli,* very short patch repair (Vsr) endonuclease cleaves the phosphodiester backbone of the mismatched thymine (Tsutakawa et al.,1999) whereas in mammals the base excision repair system initiated by the TDG enzyme is involved (Hardeland et al., 2000; Hashimoto, 2014). Contrary to the hydrolytic damage to m^5C, there are few reports on the damage induced by reactive oxygen species (Bjelland et al., 2003; Cao et al., 2007). The damages caused by the exposure to methylating agents or spontaneous methylations in cells due to erroneous reaction with SAM is almost unknown (*Manuscript i.*).

4.1.1 Fpg and Nei protein share activity on $m^{N4,5}C$-DNA

A broad range of base lesions in DNA is removed by the Fpg/Nei family of DNA glycosylases (Bjelland et al., 2003; Dizdaroglu et al., 2001). We report for the first time a repair activity for a further methylated m^5C residue in DNA. We demonstrate that the two *E. coli* DNA glycosylase/lyases remove $m^{N4,5}C$ from DNA, as well as incise the resulting AP site *in vitro*. *E. coli* Fpg removes the $m^{N4,5}C$ lesion when placed opposite non-cognate C (Figure 8D), where the excision activity is more efficient than the coupled excision/incision activity. There was a much lower incision at $m^{N4,5}C$ in DNA if placed opposite non-cognate T while no incision was detected when $m^{N4,5}C$ was placed opposite cognate G as well as A. Likewise, *E. coli* Nei removes the $m^{N4,5}C$ lesion more effectively when placed opposite cognate G, whereas the removal of the lesion opposite A and G was lower as well as no activity was detected opposite non-cognate C (Figure 9A). Our discovery of repair activity for $m^{N4,5}C$ explains the difficulty in detecting $m^{N4,5}C$ in *E. coli* DNA of cells able to synthesize this dimethylated base (Klimasauskas et al., 2002). As opposed to the *E. coli* enzymes Fpg and Nei with the ability to accommodate $m^{N4,5}C$ in their active sites, several other *E. coli* and human glycosylases like Nth and hSMUG1, hTDG, hOGG1, hNEIL1, hNEIL2 and hNEIL3 were without such ability at the incubation conditions employed.

The characterization of the biochemical properties of the Fpg and Nei single turnover rates for the base removal (k_{obs}) were calculated from experiments on $m^{N4,5}C$:C and $m^{N4,5}C$:G, respectively (Figure 10). The turnover number for the excision of $m^{N4,5}C$ by Fpg was slightly higher than the incision activity. Moreover, the turnover number for incision at the $m^{N4,5}C$ lesion was approximately the same for Fpg and Nei. The K_m and k_{cat} values determined for Fpg and

Nei lie within close range to the values determined for the known substrates (Table 9). The multiple turnover experiments determined that the binding affinity of Nei for the $m^{N4,5}C$ substrate is two folds lower than for Fpg, however, Nei has higher catalytic efficiency. Furthermore, the presence of a 3′-phosphate following the incision of $m^{N4,5}C$:C-DNA and $m^{N4,5}C$:G-DNA by Fpg and Nei, respectively, was confirmed by exposing the DNA to T4 polynucleotide kinase (PseT) which specifically removed the phosphate from the 3′end (*Manuscript ii & iii*).

Although Fpg and Nei share a similar mechanism of action and sequence homology, they differ in substrate specificity where Nei exhibits similar substrate specificity as Nth (Lee et al., 2017), although Nth was not able to process $m^{N4,5}C$-DNA (Appendix Figure B12). hSMUG1 is a human DNA glycosylase which excises uracil from DNA but has a broad substrate specificity and the only enzyme with the ability to excise thymine with an oxidized methyl group (Pettersen et al., 2007). However, hSMUG1 was inactive against $m^{N4,5}C$ placed opposite all canonical bases (Appendix Figure B9). Similarly, hTDG protein, which is a homolog of *E. coli* Mug in mammals and only recognizes thymine and uracil mispairs with guanine (Barrett et al., 1998), was also inactive against $m^{N4,5}C$ in all contexts of base pairing (Appendix Figure B10). Although hSMUG1 and hTDG belong to the same UDG superfamily and are involved in the removal of a wide array of pyrimidine damages (Pearl, 2000), they may have no role in the repair of further methylated m^5C.

4.1.2 Fpg preference for $m^{N4,5}C$ opposite C

The preference for $m^{N4,5}C$ opposite C by Fpg as opposed to other bases (Figure 9D) is somewhat unexpected although the enzyme has a strong preference for oxo^8G opposite C rather than opposite A. It is argued that the unusually high stability of the oxo^8G:A base pair in DNA deters the formation of a stable substrate recognition complex (Fromme et al., 2002). Like other DNA glycosylases, Fpg induces a bend in DNA that extrudes the damaged base into the substrate binding pocket by inserting the intercalation loop of Met74, Arg109, and Phe111 into the minor groove (Gilboa et al., 2002; Nelson et al., 2014). Arg109 fills the void left by the extruded oxo^8G and H-bonds to the Watson-Crick face of estranged cytosine O^2 and N3 with NεH and NηH of guanidium side chain group. The exclusion of oxo^8G:A as a substrate is due to poor

binding of enzyme, ineffective recognition of the estranged A and a chemical mismatch with H-bonding functionality of arginine (Fromme et al., 2002). Thus, the favorable interaction of Arg109 with the estranged C might explain why Fpg is more likely to interact with $m^{N4,5}C$:C as compared to the other contexts of base pairing. There are some cases where Fpg removes $oxo^{8}G$ opposite other bases than C (Bjelland et al., 2003; Tchou et al., 1994), which accords with our result that Fpg also accepts T opposite $m^{N4,5}C$ (Figure 8B). Consequently, there is some flexibility for the estranged base conformation in the active site of Fpg also with $m^{N4,5}C$ as substrate. The kinetic data for Fpg determined for known substrates is compared with the value for $m^{N4,5}C$ determined by us (Table 9), showing that the binding affinity of Fpg for $m^{N4,5}C$ is greater than for fapyA, fapyG and $oxo^{8}G$ in irradiated DNA. However, another experiment indicated higher binding affinity of Fpg for fapyG and $oxo^{8}G$ than for $m^{N4,5}C$. Because of large fluctuations, these parameters should be determined using similar or identical experimental conditions. Nevertheless, the similarity of the specific constants between the known substrates of Fpg and $m^{N4,5}C$ suggests the presence of $m^{N4,5}C$ in *E. coli* cells.

The Fpg structural reports published in the last decades indicate that the enzyme contains a flexible active site pocket (Coste et al., 2004; Coste et al., 2008; Fromme et al., 2002, 2003), very distinct from hOGG1 which we found inactive against $m^{N4,5}C$ (Appendix Figure B11). The flexible αF-β9 loop in the base recognition pocket, in the α-helical C-terminal domain of Fpg, is found to be structurally ordered in some structures whereas disordered in others (Fromme et al., 2002). The flexible loop works in diverse ways so it can recognize and accommodate various damaged bases unlike the pre-formed and small substrate binding pocket of hOGG1 (Coste et al., 2004). The critical functional groups in oxidized purines as the N7 and C8-keto groups of the imidazole moiety, as well as analogous groups of pyrimidines, are argued to be involved in lesion recognition before nucleobase flipping where the amino acids Met75 and Tyr238, which are strictly conserved in true Fpg proteins, mediate contacts through water-mediated interactions (Coste et al., 2004). As shown for fapyG, succeeding the nucleobase flipping, the water molecule-mediated contact is established between the receptor amino acids and donor functional groups in the active site pocket (Figure 13C, left; water 198 facilitating contact between C8-keto and water 421 mediating contact between N7 and Met75/Glu76) for *L. lactis* Fpg protein. Similarly, a tentative model has been hypothesized for $m^{N4,5}C$ with *E. coli*

Fpg protein (Figure 13C, right; a water molecule facilitating the contact between O2 and Tyr236, another one between N3 and Met73/Ser74, and the third one between N1 and Met73/Ser74) contributing to damaged base binding and positioning (*Manuscript ii*). In future, the structural study of the Fpg/m^{N4,5}C crystal complex might help to test this hypothesis and could provide a better understanding of the molecular mechanisms. In brief, Fpg can bind the diverse lesions due to the flexible and open active site pocket with the αF-β9 loop affording many possible atomic contacts, either by the amino acid and critical functional group in the substrate or through water-mediated contact in the active site.

The unexpected result that the highest activity of Fpg for m^{N4,5}C is opposite non-cognate C followed by T rather than cognate G raises a question on its biological relevance. Perhaps the recognition of m^{N4,5}C by Fpg might be due to the structural resemblance to oxidized base lesions being Fpg substrates (Figure 13B), to fit into its quite promiscuous active site (Zharkov et al., 2003) evolved to initiate repair of guanine lesions. The highest activity opposite C accords with specific interactions of the active site arginine (Figure 13C, Arg109 in *E. coli* enzyme) which stabilizes this C in an intrahelical conformation by hydrogen bonding with O2 and N3 (Fromme et al., 2002; Serre et al., 2002; Zharkov et al., 2003). However, there is a likelihood that m^{N4,5}C opposite C and perhaps T are indeed *in vivo* substrates. Therefore, a working model is suggested to explain the emergence of a m^{N4,5}C:C base pair in DNA (Figure 14). Briefly, if the aromatic ring structure of m^{N4,5}C in m^{N4,5}C:G context happens to rotate 180^0 relative to N1–dR and N4–C4 bonds during strand displacement *e.g* during replication, a sterical clash should occur with cognate G. The latter could promote the insertion of a non-cognate C, probably by the recruitment of a trans-lesion bypass polymerase. Similar to the insertion of a base opposite the un-instructional AP site, the energetically unfavorable m^{N4,5}C:C base pair would form as a relief. Consequently, this trans-lesion synthesis relief-excision scenario could cause induction of a G:C → C:G transversion in the cell (*Manuscript ii*). However, to confirm or falsify this working model, the behavior of certain DNA polymerases on reaching the m^{N4,5}C lesion in DNA should be investigated.

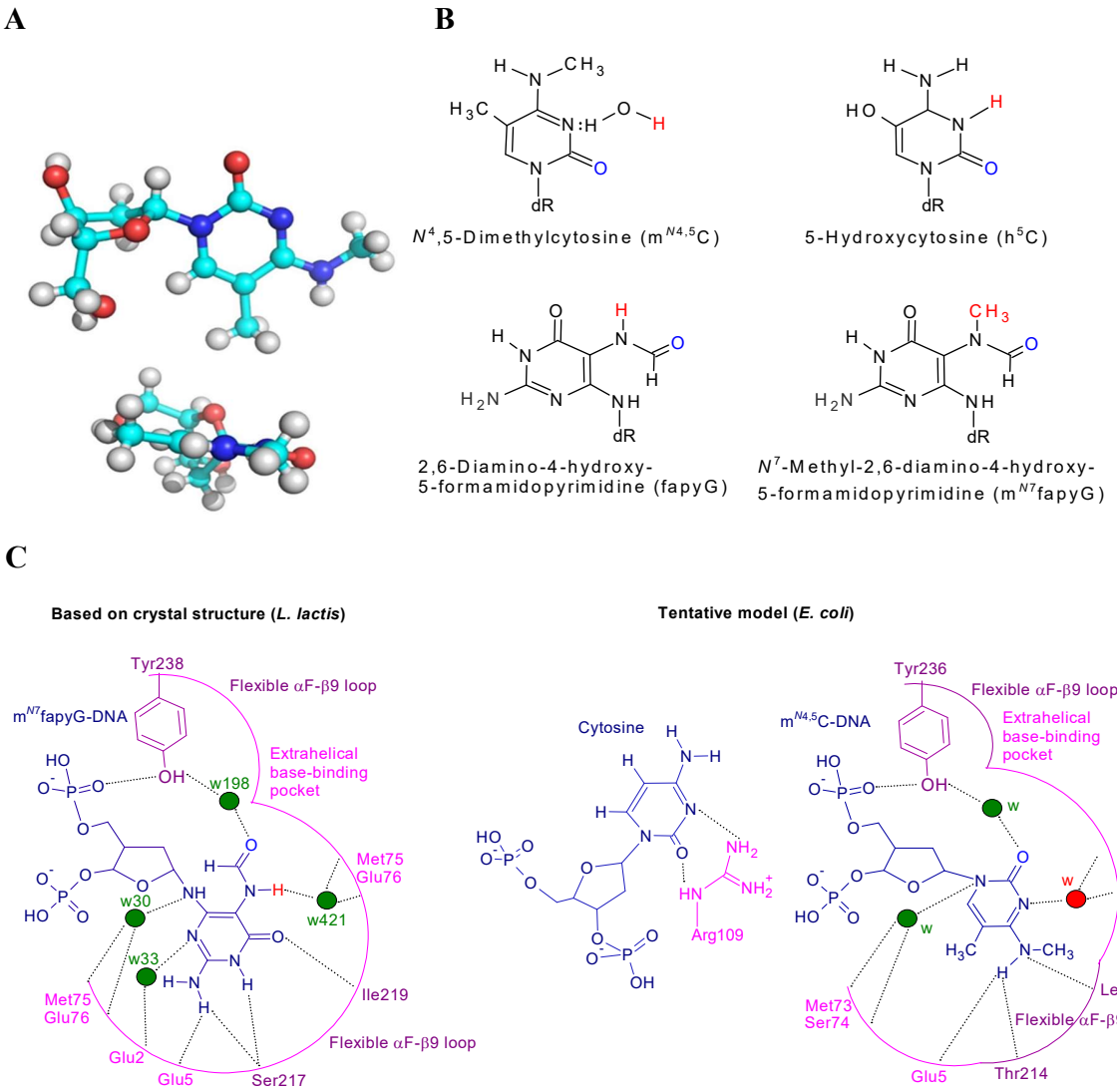


Figure 13: Structure and putative targets for the recognition of m^{N4,5}C in DNA by Fpg.
(**A**) Structure of m^{N4,5}dC. Top view, upper panel; lower panel, view from N4-C4-C1 axis. Light blue, C; blue, N; red, O; grey, H. (**B**) The structural similarities between m^{N4,5}C and other pyrimidine ring containing substrates like h^{5}C, fapyG, and m^{N7}fapyG; the critical functional groups for recognition by Fpg (Coste et al., 2004) are indicated in red (H, CH_3) and blue (O), (**C**) The tentative working model hypothesized for accommodation of m^{N4,5}C on *E. coli* Fpg (right drawings) based on the recognition of fapyG residue as described by Coste *et. al.*, where Fpg/fapyG-DNA interactions are based on the crystal structure determination of *L. lactis* (left drawing) and redrawn based on the *L. lactis* numbering. DNA is shown in dark blue, amino acids residues involved in recognition are shown in pink; for H and O determinants see **B**. Dark green spheres represents the water mediated interactions that are involved in fapyG positioning and maybe m^{N4,5}C positioning. For *E. coli* Fpg one water molecule is shown in red (right) to underline a proposed different positioning of molecule to form H-bond to a lone pair N (see m^{N4,5}C in **B**) rather than the NH as for w421to fapyG (left). Residues Ser217, Ile219 and Tyr238 in *L. lactis* Fpg, and the corresponding Thr214, Leu216 and Tyr236 in *E. coli* Fpg are part of flexible αF-β9 loop (violet) with the ability to partly define the substrate binding pocket upon the accommodation of the flipped-out base. Abbreviation: w, water. Adapted from (*Manuscript ii.)*.

4.1.3 m^{N4,5}C residue and its stereochemistry

The chemical synthesis of N^4,5-dimethylcytosine 2′-deoxynucleoside and its 5′-mono triphosphate 50 years ago demonstrated the stability of the m^{N4,5}C–deoxyribose bond (Kulikowski et al., 1969), indicating the persistence of the m^{N4,5}C residue if present in DNA. However, this dimethylated base was not further studied until the start of the new millennium, when m^{N4,5}C was introduced into the sequence 5′-C<u>C</u>AGG-3′/3′-GGG<u>C</u>C-5′ (at the nucleosides underlined of forward and complementary strand) in DNA *in vitro* by the consecutive action of the *E. coli* Dcm and *Micrococcus varians* MvaI methyltransferases (Klimasauskas et al., 2002). This double methylation of cytosine was only achieved when the C5-methyltransferase acts first followed by the N4-specific enzyme, which is also true for other cytosine-N4 methylases like *Bacillus centrosporus Bcn*IB methyltransferase, which methylates at the nucleosides underlined (red) in the sequence 5′-C<u>C</u>CGG-3′/3′-GGG<u>C</u>C-5′ (Klimasauskas et al., 2002; Vilkaitis et al., 2002).

The crystal structure of m^{N4,5}C in DNA has not been determined yet, however, the N^4,5-dimethyl-2′-deoxycytidine (m^{N4,5}dC) crystal structure is known (Audette et al., 1998) and it indicates that N4-methyl group has a *trans* orientation with respect to the C5-methyl group (Figure 14). The N4-amino group is sp^2-hybridized and planar (the p system in pyrimidine ring overlaps with lone pair p) and consequently the C4, C5 and N4 carbon and the C5 and *N*4 methyl group all lie in the same plane. In contrast, if the *N*4 methyl group lies in the *cis* conformation with respect to the C5 methyl group, there would be a steric clash between the two methyl groups which energetically disfavours the conformation (Figure 14 upper left). Thus, it requires a certain distortion in the β-helical structure of DNA to accommodate this base (Klimasauskas et al., 2002) (Figure 14 upper right), which might initiate nucleotide excision repair which recognizes distortions in DNA.

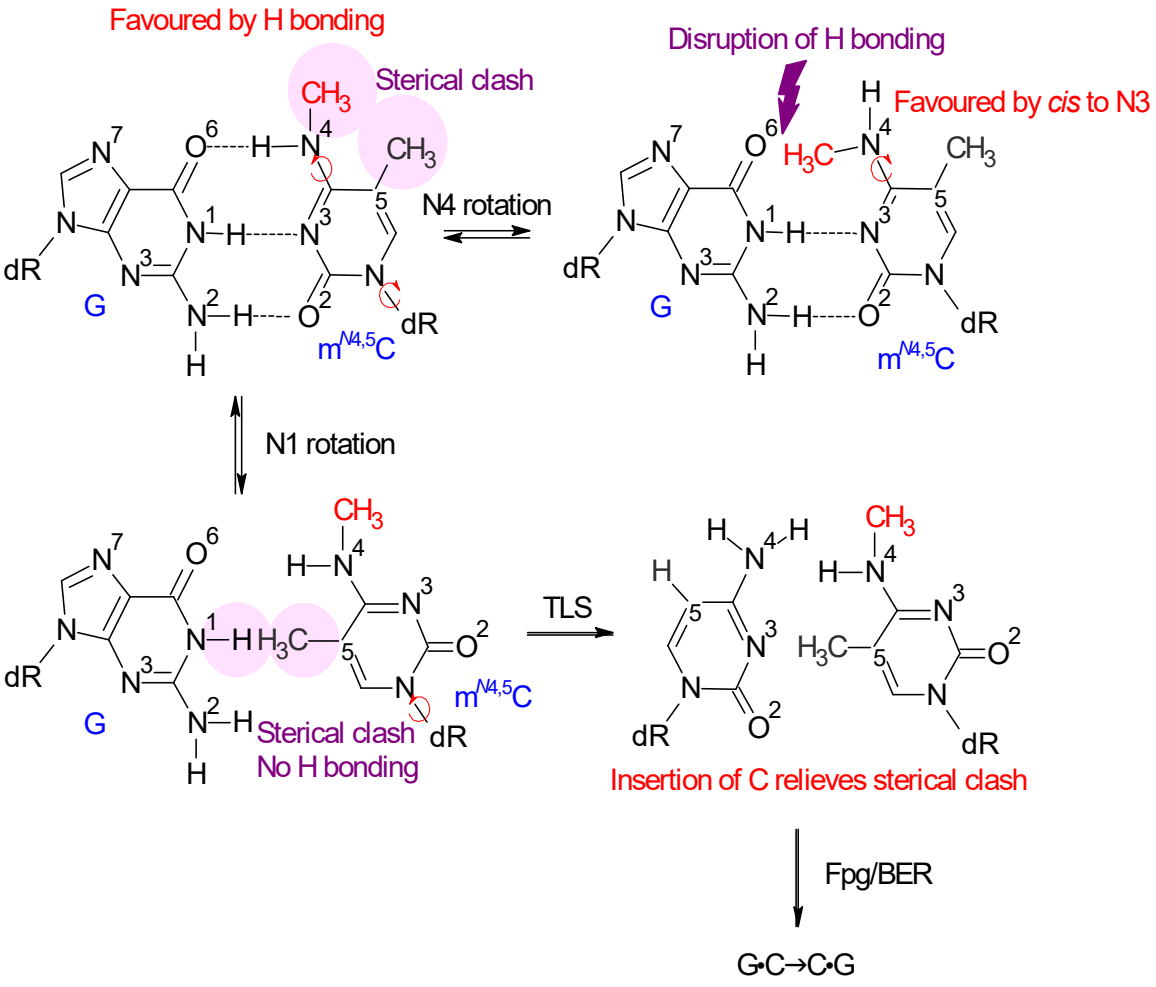


Figure 14: Working model for the conversion of cognate $m^{N4,5}C$:G to non-cognate $m^{N4,5}C$:C pair in DNA.

There are some destabilizing features in DNA structure due to the presence of $m^{N4,5}C$ residue (Klimasauskas et al., 2002). The upper panel displays that the putative sterical clash between the opposing methyl groups (left) is relieved by the rotation N4-C4 bond but in the same time disrupts the disrupts the G:C hydrogen bonding pattern (right), to exclude $m^{N4,5}C$ from accommodating it into normal Watson-Crick DNA structure. Rotation of the $m^{N4,5}C$ aromatic ring relative to N1-dR and N4-C4 bonds, which might take place during strand displacement, should result in even more unfavourable pairing with cognate G (lower panel left), which might promote insertion of non-cognate C opposite $m^{N4,5}C$ during replication to relieve this sterical clash (right). Abbreviations: BER, base excision repair; dR, deoxyribose; TLS, trans-lesion synthesis. Adapted from (*Manuscript ii.)*

4.1.4 Nei preference for $m^{N4,5}C$:G and its putative interactions in active site

In contrast to Fpg, *E. coli* Nei showed the most prominent activity towards $m^{N4,5}C$ opposite cognate G as well as opposite A and T, while there was no activity opposite non-cognate C. This result fully accords with a repair function *in vivo* and the possibility of the presence of this lesion in cellular DNA. Among substrates of Nei are oxidized pyrimidines like Tg and h^5U (Figure 15A), which are damages to m^5C and cytosine, respectively and show some structural resemblance to $m^{N4,5}C$ (Imamura et al., 2012). Although Nei has a structural resemblance with Fpg, there are variations in motifs/domains interacting with DNA. One key alteration is in the void filling intercalation triad of Nei, where the consecutive Gln69, Leu70 and Tyr71 residues are located in the same β4-β5 loop; in Fpg these triad residues are in separate positions (Prakash et al., 2012).

The crystal structure solved for *E. coli* Nei showed that, upon the DNA binding, there is an interdomain global conformational change in the protein structure that fits the DNA duplex better (Golan et al., 2005). The structure of *E. coli* Nei covalently crosslinked to an AP site-containing 13 mer DNA following Tg excision and $NaBH_4$ reduction was published in 2002 (Zharkov et al., 2002). The absence of excised Tg in the Nei-DNA complex and the inability to introduce Tg by soaking or co-crystallization, suggests that regeneration of the enzyme conformation by damaged DNA binding is necessary for subsequent catalytic cycle and thus affinity for Tg, which is destroyed after base excision. Molecular modeling simulations helped to solve the problem of how Tg is accommodated in the active site pocket, which indicated several hydrogen bonding interactions between Tg and Nei amino acid residues. The aromatic side chains of Phe227, Phe230 and Tyr 169 form the pocket enclosing the Tg. In addition to Tg, dhC is similar to $m^{N4,5}C$ by being excised by Nei and by its retained aromatic ring structure. $m^{N4,5}C$ and dhC may share 2 and 4 residues, respectively, of the total 5 residues of Tg which form hydrogen bonds with the protein if one of them is allowed to be water-mediated. The stronger binding of Tg compared to $m^{N4,5}C$ to the active site pocket maybe due to the higher number of hydrogen bonding interactions with amino acid residues (Figure 15B). However, the higher hydrophobicity of $m^{N4,5}C$ compared to Tg might give a better adaption to the hydrophobic pocket (Figure 15B right side). The K_m values reported for the different Nei-processed damaged bases vary considerably (Table 9). The combination of 3–4 hydrogen bond

interactions with the aromaticity of dhC proposes higher binding affinity to the binding pocket than for Tg, which offers 5 interactions without aromaticity (Figure 15A). The affinity for $m^{N4,5}C$ measured by us lies in between the values for other lesions processed by Nei. The binding affinity was higher than for dhC, fapyA, Th^5, $Uh^{5,}$ and Tg, while lower than for Tg_A, Tg_C and AP sites. The binding affinity for oxo^8G and Ug was very close to that for $m^{N4,5}C$. However, due to the large fluctuations the same experimental system should be used for such comparisons.

The crystal structure of the Nei-DNA covalent complex was determined with the damaged residue opposite A, where the Nε2 of Gln69 in the intercalation triad forms a hydrogen bond with N3 of estranged A (Zharkov et al., 2002). However, it was later found that Nei discriminates against an opposite A in the case of some lesions and has a preference for an opposite G. Possibly, it can form an additional hydrogen bond between Oε1 of Gln69 and N2 of estranged G (Kropachev et al., 2006). The latter accords with our results showing the highest preference for $m^{N4,5}C$ opposite G. Furthermore, in the case of hNEIL1, a human homolog of Nei, Tg is shown to occupy the wobble position when mispaired with G and is excised more quickly than opposite A (Imamura et al., 2012). However, the molecular mechanism behind the specificity of the enzyme preference opposite G is not clear. Thus, further studies should be performed on the interaction of Nei with $m^{N4,5}C$ in DNA to more exactly elucidate the contacts made by the enzyme, *e.g.* by using proteins with certain amino acid residues replaced. Since Fpg has even a higher affinity for the substrate $m^{N4,5}C$:C than Nei has for the substrate $m^{N4,5}C$:G, it is reasonable to suggest that Fpg and Nei have complementary roles in the repair process of this lesion *in vivo*.

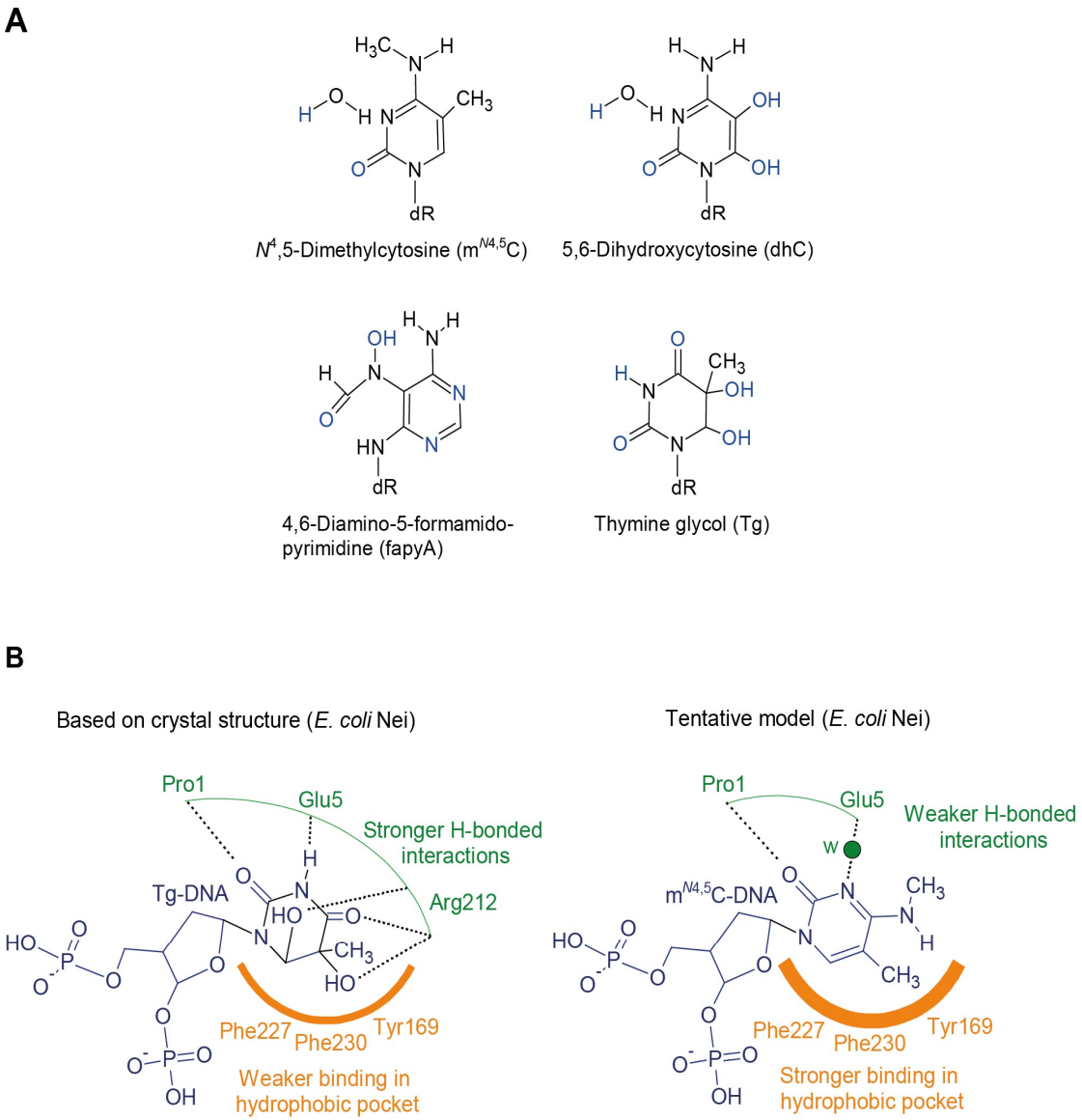


Figure 15: Structure and putative targets for recognition of $m^{N4,5}C$ in DNA by Nei.
(**A**) Structural similarities between $m^{N4,5}C$ and other pyrimidine ring-containing Nei substrates. Analogous determinants for recognition by Nei residues are indicated in blue. The determinants for the thymine glycol (Tg) residue is based on reported crystal structure of Nei together with Tg substrate (Zharkov et al., 2002) while tentative for the other damaged bases. (**B**) Tentative working model for accommodation of $m^{N4,5}C$ in *E. coli* Nei active site (right drawing) based on recognition of the Tg residue (left drawing) as described by Zharkov *et al.*, where the Nei/Tg-DNA interactions are based upon crystal structure determination and are redrawn after their report. Amino acid residues involved in recognition are shown in dark green and the DNA is shown in dark blue. A dark green sphere represents tentative water-mediated interactions that may be involved in $m^{N4,5}C$ positioning. Abbreviation: w, water. Adapted from (*Manuscript iii*).

In an *E. coli dcm*$^+$ strain carrying the MvaI methyltransferase gene, both m^5C and $m^{N4}C$ were present in cellular DNA as analyzed by HPLC, as opposed to $m^{N4,5}C$ which was hardly detectable. Thus, no solid evidence was provided that $m^{N4,5}C$ can be formed *in vivo*, but a reasonable explanation suggested was efficient removal by DNA repair. Also, transformation with a plasmid carrying the MvaI methyltransferase gene induced the SOS response, which were absent in mutations in this or the chromosomal *dcm* gene, providing an indication of some lethal effects in the cells because of the persistence of $m^{N4,5}C$ residues in DNA (Klimasauskas et al., 2002). This explanation seems reasonable, also because the SOS response increases the nucleotide excision repair capacity significantly. Interestingly, $m^{N4,5}C$ is more stable than cytosine, m^5C and $m^{N4}C$ to both deamination and hydrolysis in its cytidine form (Kusmierek et al., 1989), which suggests considerable *in vivo* stability and need for active repair of this putative lethal lesion (*Manuscript ii*). Our detection of removal of $m^{N4,5}C$ residue by the two DNA glycosylases/lyases Fpg and Nei in all contexts of base pairing accords with the *in vivo* observations cited above by confirming its repair for the first time.

There is no presence of cytosine-*N*4 methyltransferase in *E. coli* to methylate Dcm-methylated cytosines, however, other bacterial species or strains contain both m^5C and $m^{N4}C$ in their DNA (Ehrlich et al., 1987) and thus both types of methylases suggesting the formation of the $m^{N4,5}C$ damage as a relatively frequent event. Nevertheless, in the case of *E. coli*, we cannot totally rule out the possibility of acquisition of cytosine-*N*4 methyltransferase possibly by horizontal gene transfer or viral infection as well as $m^{N4,5}C$ may arise as a consequence of chemical insults to the cellular DNA, explaining the need for its repair. Mutated versions of methyltransferase genes and thus proteins may also be a source of erroneous methylations (Figure 16). Indeed, our finding that two enzymes complement each other in the initiation of BER of $m^{N4,5}C$ strengthen the notion that this lesion is a challenge to the genomic integrity of *E. coli*. Likewise, the lack of activity for $m^{N4,5}C$ exhibited by the human Nei orthologues (Figure 9C) and other human glycosylases (*Manuscript ii.*) might suggest that $m^{N4,5}C$ is a primarily "prokaryotic" lesion. However, the much higher abundance of m^5C in eukaryotic including mammalian DNA argues against that notion and suggests a thorough search for human or mammalian repair activities for $m^{N4,5}C$ and to possibly detect the lesion in biological samples.

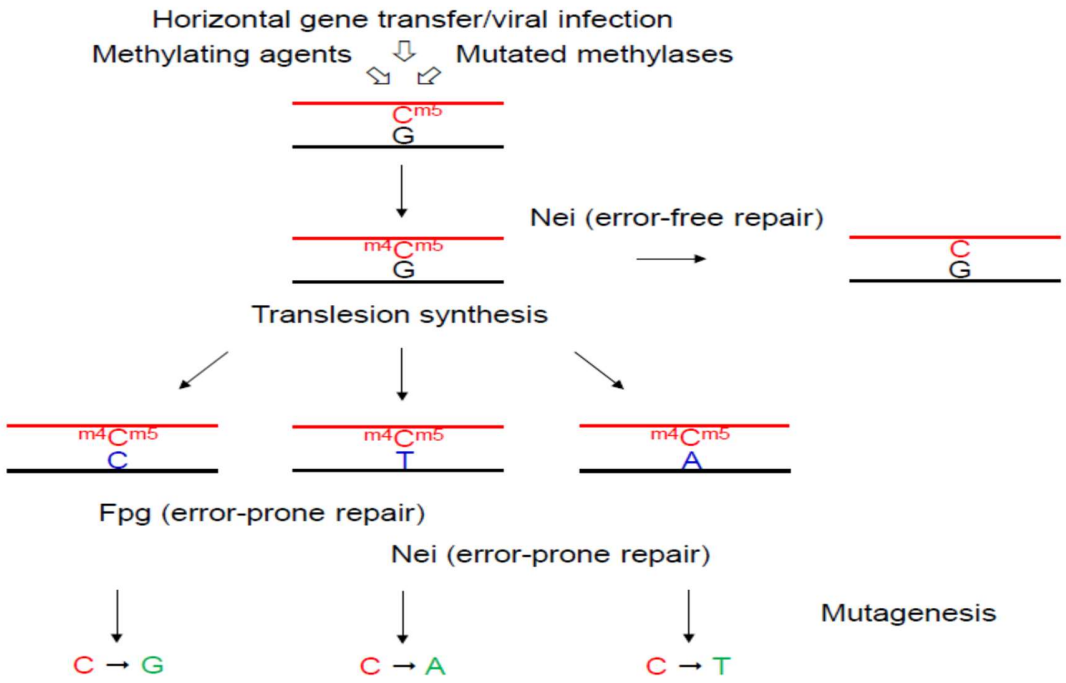


Figure 16: Origin and biological consequences of m^{N4,5}C in *E. coli* DNA.
Working model based on our present and the previous description of *in vitro* repair of m^{N4,5}C in DNA. Tentatively, the m^{N4,5}C damage may arise in *E. coli* DNA by exposure to a methylating agent, acquisition of a cytosine-*N*4 methylase gene by a horizontal gene transfer or other events or as a consequence of a mutated methylase gene able to produce a protein capable of executing erroneous methyl transfer reactions. Primarily, we expect the Nei glycosylase to initiate error-free BER of m^{N4,5}C in DNA. However, if m^{N4,5}C escape such repair, it may be bypassed by a replicative or translesion bypass polymerase to insert a non-cognate base during the first round of replication (blue) followed by error-prone repair of m^{N4,5}C initiated by Nei or Fpg resulting in substitution of C (original unmethylated and methylated base/base lesion in red) for another base (green) during repair replication. Adapted from (*Manuscript iii*.).

According to our working model (Figure 16), we primarily expect the Nei glycosylase to initiate error-free BER of m^{N4,5}C in DNA opposite its cognate G. However, if m^{N4,5}C escapes such repair, it may be bypassed by a replicative or translesion bypass polymerase to insert a non-cognate base opposite m^{N4,5}C during the first round of replication. That may be C, T or A, since all those three bases can be inserted opposite oxidized cytosines having similarity in structure to m^{N4,5}C such as h^{5}C and h^{5}U (Bjelland et al., 2003). This may be followed by error-prone removal of m^{N4,5}C by Nei or Fpg resulting in the substitution of the original C for another base during the second round of repair replication. The biological rationale for such error prone repair of the m^{N4,5}C lesion maybe that the genotoxicity of m^{N4,5}C is so devastating to the cell that mutagenesis is an acceptable prize to pay for its removal.

5 Conclusion

The DNA is always in threat of damage by various endogenous and exogenous agents causing cytotoxicity, mutagenicity, carcinogenicity and aging to cells. Methylation damage to the epigenetic mark m^5C residing in CpG dinucleotides/islands could be even more damaging to the cells than to the other bases in DNA since m^5C in addition to its informational function is important for gene regulation. One such damage is $m^{N4,5}C$.

Certain methylases can convert m^5C into $m^{N4,5}C$ in DNA *in vitro*, and there is a possibility of the presence of $m^{N4,5}C$ *in vivo*. We have found that the *E. coli* Fpg and Nei proteins initiate BER of $m^{N4,5}C$ in DNA *in vitro*. Fpg removes $m^{N4,5}C$ most efficiently opposite non-cognate C followed by T, and Nei incises at $m^{N4,5}C$ in DNA most efficiently opposite cognate G followed by A and T. Thus, the DNA might be protected from the possible harmful effect of $m^{N4,5}C$ by at least two DNA glycosylases complementing each other. This is the first description of repair functions for a further methylated m^5C, as well as the first alkylated base lesions found to be repaired by Fpg and Nei, which are involved in the repair of oxidized bases. We primarily expect the Nei glycosylase to initiate error-free BER of $m^{N4,5}C$ in DNA opposite its cognate G. If $m^{N4,5}C$ escapes such repair, it may be bypassed by a replicative or translesion bypass polymerase to insert a non-cognate base opposite $m^{N4,5}C$ during the first round of replication followed by error-prone removal of $m^{N4,5}C$ by Nei or Fpg. This will result in the substitution of the original C for another base during the second round of replication. The biological rationale for such error prone repair of the $m^{N4,5}C$ lesion maybe that the genotoxicity of $m^{N4,5}C$ is so devastating to the cell that mutagenesis is an acceptable prize to pay for its removal.

References

Abramowitz, M., & Davidson, M. W. (2002, 2016-02-12). Concepts in Digital Imaging Technology: Photomultiplier Tubes. Retrieved from https://micro.magnet.fsu.edu/primer/digitalimaging/concepts/photomultipliers.html

Ananthapadmanabhan, K., & Aronson, M. (1975). Spectroscopic probe analysis of protein-surfactant interactions: The BSNSDS system.

Audette, G., Kumar, S., Gupta, S., & Quail, J. (1998). N4, 5-Dimethyl-2'-deoxycytidine. *Acta Crystallographica Section C: Crystal Structure Communications, 54*(12), 1987-1990.

Bailly, V., & Verly, W. G. (1987). *Escherichia coli* endonuclease III is not an endonuclease but a β-elimination catalyst. *Biochem J, 242*(2), 565–572.

Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., & Pearl, L. H. (1998). Crystal Structure of a G:T/U Mismatch-Specific DNA Glycosylase: Mismatch Recognition by Complementary-Strand Interactions. *Cell, 92*(1), 117-129. doi:https://doi.org/10.1016/S0092-8674(00)80904-6

Barrows, L. R., & Magee, P. N. (1982). Nonenzymatic methylation of DNA by S-adenosylmethionine in vitro. *Carcinogenesis, 3*(3), 349-351.

Begley, T. J., & Samson, L. D. (2003). AlkB mystery solved: oxidative demethylation of N1-methyladenine and N3-methylcytosine adducts by a direct reversal mechanism. *Trends Biochem Sci, 28*(1), 2-5.

Berg, J., Tymoczko, J., & Stryer, L. (2002). The Michaelis-Menten model accounts for the kinetic properties of many enzymes. *Biochemistry*, 319-330.

Bjelland, S., Bjørås, M., & Seeberg, E. (1993). Excision of 3-methylguanine from alkylated DNA by 3-methyladenine DNA glycosylase I of Escherichia coli. *Nucleic acids research, 21*(9), 2045-2049.

Bjelland, S., & Seeberg, E. (2003). Mutagenicity, toxicity and repair of DNA base damage induced by oxidation. *Mutat Res, 531*, 37–80.

Bodell, W., & Singer, B. (1979). Influence of hydrogen bonding in DNA and polynucleotides on reaction of nitrogens and oxygens toward ethylnitrosourea. *Biochemistry, 18*(13), 2860-2863.

Breiling, A., & Lyko, F. (2015). Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics & chromatin, 8*(1), 1.

Campbell, V. W., & Jackson, D. A. (1980). The effect of divalent cations on the mode of action of DNase I. The initial reaction products produced from covalently closed circular DNA. *J Biol Chem, 255*(8), 3726-3735.

Cao, H., & Wang, Y. (2007). Quantification of oxidative single-base and intrastrand cross-link lesions in unmethylated and CpG-methylated DNA induced by Fenton-type reagents. *Nucleic acids research, 35*(14), 4833-4844. doi:10.1093/nar/gkm497

Chen, Q., Zhu, X., Li, Y., & Meng, Z. (2014). Epigenetic regulation and cancer (Review). *Oncology reports, 31*(2), 523-532.

Cooper, D. N., & Youssoufian, H. (1988). The CpG dinucleotide and human genetic disease. *Human genetics, 78*(2), 151-155. doi:10.1007/bf00278187

Coste, F., Ober, M., Carell, T., Boiteux, S., Zelwer, C., & Castaing, B. (2004). Structural basis for the recognition of the FapydG lesion (2, 6-diamino-4-hydroxy-5-formamidopyrimidine) by formamidopyrimidine-DNA glycosylase. *Journal of biological chemistry, 279*(42), 44074-44083.

Coste, F., Ober, M., Le Bihan, Y.-V., Izquierdo, M. A., Hervouet, N., Mueller, H., . . . Castaing, B. (2008). Bacterial base excision repair enzyme Fpg recognizes bulky N7-substituted-FapydG lesion via unproductive binding mode. *Chemistry & biology, 15*(7), 706-717.

Demple, B., Johnson, A., & Fung, D. (1986). Exonuclease III and endonuclease IV remove 3'blocks from DNA synthesis primers in H2O2-damaged Escherichia coli. *Proceedings of the National Academy of Sciences, 83*(20), 7731-7735.

Dizdaroglu, M., Burgess, S. M., Jaruga, P., Hazra, T. K., Rodriguez, H., & Lloyd, R. S. (2001). Substrate Specificity and Excision Kinetics of Escherichia coli Endonuclease VIII (Nei) for Modified Bases in DNA Damaged by Free Radicals. *Biochemistry, 40*(40), 12150-12156. doi:10.1021/bi015552o

Ebeling, W., Hennrich, N., Klockow, M., Metz, H., Orth, H. D., & Lang, H. (1974). Proteinase K from Tritirachium album Limber. *Eur J Biochem, 47*(1), 91-97.

Ehrlich, M., Norris, K. F., Wang, R. Y., Kuo, K. C., & Gehrke, C. W. (1986). DNA cytosine methylation and heat-induced deamination. *Biosci Rep, 6*(4), 387-393.

Ehrlich, M., Wilson, G. G., Kuo, K. C., & Gehrke, C. W. (1987). *N*4-methylcytosine as a minor base in bacterial DNA. *J Bacteriol, 169*(3), 939–943.

Fromme, J. C., & Verdine, G. L. (2002). Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nature Structural & Molecular Biology, 9*(7), 544-552.

Fromme, J. C., & Verdine, G. L. (2003). DNA lesion recognition by the bacterial repair enzyme MutM. *Journal of biological chemistry, 278*(51), 51543-51548.

Gilboa, R., Zharkov, D. O., Golan, G., Fernandes, A. S., Gerchman, S. E., Matz, E., . . . Shoham, G. (2002). Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *J Biol Chem, 277*(22), 19811-19816. doi:10.1074/jbc.M202058200

Golan, G., Zharkov, D. O., Feinberg, H., Fernandes, A. S., Zaika, E. I., Kycia, J. H., . . . Shoham, G. (2005). Structure of the uncomplexed DNA repair enzyme endonuclease VIII indicates significant interdomain flexibility. *Nucleic Acids Res, 33*(15), 5006-5016. doi:10.1093/nar/gki796

Graves, R. J., Felzenszwalb, I., Laval, J., & O'Connor, T. R. (1992). Excision of 5'-terminal deoxyribose phosphate from damaged DNA is catalyzed by the Fpg protein of Escherichia coli. *J Biol Chem, 267*(20), 14429-14435.

Grollman, A. P., & Moriya, M. (1993). Mutagenesis by 8-oxoguanine: an enemy within. *Trends in genetics, 9*(7), 246-249.

Guo, Y., Bandaru, V., Jaruga, P., Zhao, X., Burrows, C. J., Iwai, S., . . . Wallace, S. S. (2010). The oxidative DNA glycosylases of Mycobacterium tuberculosis exhibit different substrate preferences from their Escherichia coli counterparts. *DNA repair, 9*(2), 177-190. doi:https://doi.org/10.1016/j.dnarep.2009.11.008

Hardeland, U., Bentele, M., Jiricny, J., & Schär, P. (2000). Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *Journal of biological chemistry, 275*(43), 33449-33456.

Hashimoto, H. (2014). Structural and mutation studies of two DNA demethylation related glycosylases: MBD4 and TDG. *BIOPHYSICS, 10*, 63-68. doi:10.2142/biophysics.10.63

Hashimoto, H., Hong, S., Bhagwat, A. S., Zhang, X., & Cheng, X. (2012). Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acids Res, 40*(20), 10203-10214. doi:10.1093/nar/gks845

Hazra, T. K., Izumi, T., Venkataraman, R., Kow, Y. W., Dizdaroglu, M., & Mitra, S. (2000). Characterization of a novel 8-oxoguanine-DNA glycosylase activity in Escherichia coli and identification of the enzyme as endonuclease VIII. *Journal of biological chemistry, 275*(36), 27762-27767.

Heyn, H., & Esteller, M. (2015). An adenine code for DNA: a second life for N6-methyladenine. *Cell, 161*(4), 710-713.

Hirano, K., Ishido, T., & Ishikawa, M. (2009). Rapid sequencing gel electrophoresis using glycerol-tolerant sodium taurine medium. *Anal Biochem, 390*(1), 100-101. doi:10.1016/j.ab.2009.04.006

Imamura, K., Averill, A., Wallace, S. S., & Doublié, S. (2012). Structural characterization of viral ortholog of human DNA glycosylase NEIL1 bound to thymine glycol or 5-hydroxyuracil-containing DNA. *Journal of biological chemistry, 287*(6), 4288-4298.

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., . . . Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science, 333*(6047), 1300–1303.

Jiang, D., Hatahet, Z., Melamede, R. J., Kow, Y. W., & Wallace, S. S. (1997). Characterization of Escherichia coli endonuclease VIII. *Journal of biological chemistry, 272*(51), 32230-32239.

Johnson, G. (1979). Genetically controlled variation in the shapes of enzymes. *Progress in nucleic acid research and molecular biology, 22*, 293-326.

Karakaya, A., Jaruga, P., Bohr, V. A., Grollman, A. P., & Dizdaroglu, M. (1997). Kinetics of excision of purine lesions from DNA by Escherichia coli Fpg protein. *Nucleic acids research, 25*(3), 474-479.

Katafuchi, A., Nakano, T., Masaoka, A., Terato, H., Iwai, S., Hanaoka, F., & Ide, H. (2004). Differential specificity of human and Escherichia coli endonuclease III and VIII homologues for oxidative base lesions. *Journal of biological chemistry, 279*(14), 14464-14471.

Kim, Y.-J., & Wilson, D. M. (2012). Overview of Base Excision Repair Biochemistry. *Current molecular pharmacology, 5*(1), 3-13.

Kladova, O. A., Kuznetsova, A. A., Fedorova, O. S., & Kuznetsov, N. A. (2017). Mutational and kinetic analysis of lesion recognition by *Escherichia coli* endonuclease VIII. *Genes (Basel), 8*(5). doi:10.3390/genes8050140

Klimasauskas, S., Gerasimaite, R., Vilkaitis, G., & Kulakauskas, S. (2002). *N4, 5-dimethylcytosine, a novel hypermodified base in DNA.* Paper presented at the Nucleic acids symposium series.

Klungland, A., & Robertson, A. B. (2016). Oxidized C5-methyl cytosine bases in DNA: 5-Hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine. *Free Radical Biology and Medicine*.

Krokan, H. E., & Bjørås, M. (2013). Base excision repair. *Cold Spring Harbor perspectives in biology, 5*(4), a012583.

Krokan, H. E., Nilsen, H., Skorpen, F., Otterlei, M., & Slupphaug, G. (2000). Base excision repair of DNA in mammalian cells. *FEBS letters, 476*(1-2), 73-77.

Krokeide, S. Z., Laerdahl, J. K., Salah, M., Luna, L., Cederkvist, F. H., Fleming, A. M., . . . Bjørås, M. (2013). Human NEIL3 is mainly a monofunctional DNA glycosylase removing spiroimindiohydantoin and guanidinohydantoin. *DNA repair, 12*(12), 1159-1164. doi:https://doi.org/10.1016/j.dnarep.2013.04.026

Kropachev, K. Y., Zharkov, D. O., & Grollman, A. P. (2006). Catalytic mechanism of *Escherichia coli* endonuclease VIII: roles of the intercalation loop and the zinc finger. *Biochemistry, 45*(39), 12039–12049. doi:10.1021/bi060663e

Krwawicz, J., Arczewska, K. D., Speina, E., Maciejewska, A. M., & Grzesiuk, E. (2007). Bacterial DNA repair genes and their eukaryotic homologues: 1. Mutations in genes involved in base excision repair (BER) and DNA-end processors and their implication in mutagenesis and human disease. *Acta biochimica Polonica, 54*(3), 413-434.

Kulikowski, T., Żmudzka, B., & Shugar, D. (1969). Preparation and properties of N^4-alkyl analogues of 5-methyl-2′-deoxycytidine, their 5′-mono- and triphosphates, and N^4-alkyl derivatives of 1,5-dimethylcytosine. *Acta Biochim Pol, 16*(2), 201–217.

Kusmierek, J., Käppi, R., Neuvonen, K., Shugar, D., & Lönnberg, H. (1989). Kinetics and mechanisms of hydrolytic reactions of methylated cytidines under acidic and neutral conditions. *Acta Chem. Scand, 43*, 196-202.

Lee, A. J., & Wallace, S. S. (2017). Hide and seek: How do DNA glycosylases locate oxidatively damaged DNA bases amidst a sea of undamaged bases? *Free Radical Biology and Medicine, 107*, 170-178.

Lee, H.-W., Dominy, B. N., & Cao, W. (2011). New family of deamination repair enzymes in uracil-DNA glycosylase superfamily. *Journal of biological chemistry, 286*(36), 31282-31287.

Leipold, M. D., Muller, J. G., Burrows, C. J., & David, S. S. (2000). Removal of Hydantoin Products of 8-Oxoguanine Oxidation by the Escherichia coli DNA Repair Enzyme, FPG. *Biochemistry, 39*(48), 14984-14992. doi:10.1021/bi0017982

Leiros, I., Nabong, M. P., Grøsvik, K., Ringvoll, J., Haugland, G. T., Uldal, L., . . . Moe, E. (2007). Structural basis for enzymatic excision of N1-methyladenine and N3-methylcytosine from DNA. *The EMBO journal, 26*(8), 2206-2217.

Li, D., Fedeles, B. I., Shrivastav, N., Delaney, J. C., Yang, X., Wong, C., . . . Essigmann, J. M. (2013). Removal of N-alkyl modifications from N 2-alkylguanine and N 4-alkylcytosine in DNA by the adaptive response protein AlkB. *Chemical research in toxicology, 26*(8), 1182-1187.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *nature, 362*(6422), 709-715.

Madugundu, G. S., Cadet, J., & Wagner, J. R. (2014). Hydroxyl-radical-induced oxidation of 5-methylcytosine in isolated and cellular DNA. *Nucleic acids research*, gku334.

Miura, Y., Wake, H., & Kato, T. (1999). TBE, or not TBE; that is the question: Beneficial usage of tris-borate for obtaining a higher resolution of small DNA fragments by agarose gel electrophoresis. *Nagoya medical journal, 43*(1), 1-6.

Nelson, S. R., Dunn, A. R., Kathe, S. D., Warshaw, D. M., & Wallace, S. S. (2014). Two glycosylase families diffusively scan DNA using a wedge residue to probe for and identify oxidatively damaged bases. *Proc Natl Acad Sci U S A, 111*(20), E2091-2099. doi:10.1073/pnas.1400386111

Nilsen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdine, G. L., & Lindahl, T. (2001). Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil–DNA glycosylase. *The EMBO journal, 20*(15), 4278-4286. doi:10.1093/emboj/20.15.4278

Pearl, L. H. (2000). Structure and function in the uracil-DNA glycosylase superfamily. *Mutation Research/DNA Repair, 460*(3), 165-181.

Petit, C., & Sancar, A. (1999). Nucleotide excision repair: from E. coli to man. *Biochimie, 81*(1), 15-25.

Pettersen, H. S., Sundheim, O., Gilljam, K. M., Slupphaug, G., Krokan, H. E., & Kavli, B. (2007). Uracil–DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic acids research, 35*(12), 3879-3892. doi:10.1093/nar/gkm372

Pfeifer, G. P., Denissenko, M. F., Olivier, M., Tretyakova, N., Hecht, S. S., & Hainaut, P. (2002). Tobacco smoke carcinogens, DNA damage and p 53 mutations in smoking-associated cancers. *Oncogene, 21*(48), 7435-7451.

Pfeifer, G. P., You, Y.-H., & Besaratinia, A. (2005). Mutations induced by ultraviolet light. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 571*(1), 19-31.

Philip, P. A., Souliotis, V. L., Harris, A. L., Salisbury, A., Tates, A. D., Mitchell, K., . . . Kyrtopoulos, S. A. (1996). Methyl DNA adducts, DNA repair, and hypoxanthine-guanine phosphoribosyl transferase mutations in peripheral white blood cells from patients with malignant melanoma treated with dacarbazine and hydroxyurea. *Clinical cancer research, 2*(2), 303-310.

Prakash, A., Doublie, S., & Wallace, S. S. (2012). The Fpg/Nei family of DNA glycosylases: substrates, structures, and search for damage. *Prog Mol Biol Transl Sci, 110*, 71-91. doi:10.1016/b978-0-12-387665-2.00004-3

Rydberg, B., & Lindahl, T. (1982). Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *The EMBO journal, 1*(2), 211.

Saito, Y., Uraki, F., Nakajima, S., Asaeda, A., Ono, K., Kubo, K., & Yamamoto, K. (1997). Characterization of endonuclease III (nth) and endonuclease VIII (nei) mutants of Escherichia coli K-12. *Journal of bacteriology, 179*(11), 3783-3785.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Meth, 9*(7), 671-675.

Sedgwick, B. (2004). Repairing DNA-methylation damage. *Nature Reviews Molecular Cell Biology, 5*(2), 148-157.

Sedgwick, B., & Lindahl, T. (2002). Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene, 21*(58), 8886.

Seeberg, E., Eide, L., & Bjoras, M. (1995). The base excision repair pathway. *Trends Biochem Sci, 20*(10), 391-397.

Serre, L., de Jésus, K. P., Boiteux, S., Zelwer, C., & Castaing, B. (2002). Crystal structure of the Lactococcus lactis formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA. *The EMBO journal, 21*(12), 2854-2865.

Shinmura, K., Kasai, H., Sasaki, A., Sugimura, H., & Yokota, J. (1997). 8-hydroxyguanine (7,8-dihydro-8-oxoguanine) DNA glycosylase and AP lyase activities of hOGG1 protein and their substrate specificity. *Mutation Research/DNA Repair, 385*(1), 75-82. doi:http://dx.doi.org/10.1016/S0921-8777(97)00041-4

Smith, B. (1984). SDS polyacrylamide gel electrophoresis of proteins. *Proteins*, 41-55.

Subramanya, H. S., Doherty, A. J., Ashford, S. R., & Wigley, D. B. (1996). Crystal Structure of an ATP-Dependent DNA Ligase from Bacteriophage T7. *Cell, 85*(4), 607-615. doi:http://dx.doi.org/10.1016/S0092-8674(00)81260-X

Summer, H., Grämer, R., & Dröge, P. (2009). Denaturing Urea Polyacrylamide Gel Electrophoresis (Urea PAGE). *Journal of Visualized Experiments : JoVE*(32), 1485. doi:10.3791/1485

Takao, M., Kanno, S., Kobayashi, K., Zhang, Q. M., Yonei, S., van der Horst, G. T., & Yasui, A. (2002). A back-up glycosylase in Nth1 knock-out mice is a functional Nei (endonuclease VIII) homologue. *J Biol Chem, 277*(44), 42205-42213. doi:10.1074/jbc.M206884200

Tchou, J., Bodepudi, V., Shibutani, S., Antoshechkin, I., Miller, J., Grollman, A. P., & Johnson, F. (1994). Substrate specificity of Fpg protein. Recognition and cleavage of oxidatively damaged DNA. *Journal of biological chemistry, 269*(21), 15318-15324.

Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T., & Sedgwick, B. (2002). Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *nature, 419*(6903), 174-178.

Tsutakawa, S. E., Jingami, H., & Morikawa, K. (1999). Recognition of a TG mismatch: the crystal structure of very short patch repair endonuclease in complex with a DNA duplex. *Cell, 99*(6), 615-623.

Um, S., Harbers, M., Benecke, A., Pierrat, B., Losson, R., & Chambon, P. (1998). Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *J Biol Chem, 273*(33), 20728-20736.

Vilkaitis, G., Lubys, A., Merkienė, E., Timinskas, A., Janulaitis, A., & Klimašauskas, S. (2002). Circular permutation of DNA cytosine-N4 methyltransferases: in vivo coexistence in the BcnI system and in vitro probing by hybrid formation. *Nucleic acids research, 30*(7), 1547-1557.

Wang, R. Y., Kuo, K. C., Gehrke, C. W., Huang, L. H., & Ehrlich, M. (1982). Heat- and alkali-induced deamination of 5-methylcytosine and cytosine residues in DNA. *Biochim Biophys Acta, 697*(3), 371-377.

Wu, T. P., Wang, T., Seetin, M. G., Lai, Y., Zhu, S., Lin, K., . . . Xiao, A. Z. (2016). DNA methylation on N^6-adenine in mammalian embryonic stem cells. *nature, 532*(7599), 329–333. doi:10.1038/nature17640

Zharkov, D. O., Golan, G., Gilboa, R., Fernandes, A. S., Gerchman, S. E., Kycia, J. H., . . . Shoham, G. (2002). Structural analysis of an *Escherichia coli* endonuclease VIII covalent reaction intermediate. *EMBO J, 21*(4), 789–800. doi:10.1093/emboj/21.4.789

Zharkov, D. O., Shoham, G., & Grollman, A. P. (2003). Structural characterization of the Fpg family of DNA glycosylases. *DNA repair, 2*(8), 839-862.

Zumbo, P. (2012). Ethanol Precipitation. Retrieved from http://physiology.med.cornell.edu/faculty/mason/lab/zumbo/files/ETHANOL_PRECIPITATION.pdf

APPENDICES

Appendix A

Protocols

Protocol 1: Base excision/DNA incision assays

Vials kept on ice and in darkness during the assay.

Reaction mixtures

Fpg			
Reagents	*Stock*	*Reaction mix*	*Reaction 1× (µl)*
NEB1 buffer	10×	1×	2
DTT	20 mM	1 mM	1
Bovine serum albumin (BSA)	10 mg/ml	0.1 mg/ml	1
Labelled oligo	1 pmol/ µl	1 pmol	1
Enzyme	Varying	Varying	1
Deionized H_2O			13
Total volume			20
Nei			
Reagents	*Stock*	*Reaction mix*	*Reaction 1× (µl)*
Nei buffer 5×	5×	1×	4
Labelled oligo	1 pmol/ µl	1 pmol	1
Enzyme	Varying	Varying	1
Deionized H_2O			14
Total volume			20
hSMUG1			
Reagents	*Stock*	*Reaction mix*	*Reaction 1× (µl)*
HEPES buffer 5× + 5 mM DTT	5×	1×	4
KCl	1 M	70 mM	1.4
BSA	10 mg/ml	0.1 mg/ml	1
Labelled DNA	1 pmol/µl	1 pmol	1
Enzyme	Varying	Varying	1
Deionized H_2O			11.6
Total volume			20
hTDG			
Reagents	*Stock*	*Reaction mix*	*Reaction 1× (µl)*
Nicking buffer 5× + 10 mM DTT	5×	1×	4
Uracil Glycosylase Inhibitor (UGI)	0.5 units/µl	0.5 units	1
BSA	10 mg/ml	0.1 mg/ml	1
Labelled oligo	1 pmol/µl	1 pmol	1
Enzyme	Varying	Varying	1
Deionized H_2O			12
Total volume			20

hOGG1			
Reagents	*Stock*	*Reaction mix*	*Reaction 1× (µl)*
NEB2 5× + 5 mM DTT	5×	1×	4
BSA	10 mg/ml	0.1 mg/ml	1
Labelled oligo	1 pmol/µl	1 pmol	1
Enzyme	Varying	Varying	1
Deionized H_2O			13
Total volume			20
Nth			
Reagents	*Stock*	*Reaction mix*	*Reaction 1× (µl)*
Nth buffer 5× + 5 mM DTT	5×	1×	4
BSA	10 mg/ml	0.1 mg/ml	1
Labelled oligo	1 pmol/µl	1 pmol	1
Enzyme	Varying	Varying	1
Deionized H_2O			13
Total volume			20
hNEIL1			
Reagents	*Stock*	*Reaction mix*	*Reaction 1× (µl)*
Nei buffer 5×	5×	1×	4
Labelled oligo	1 pmol/ µl	1 pmol	1
Enzyme	Varying	Varying	1
Deionized H_2O			14
Total volume			20
hNEIL2			
Reagents	*Stock*	*Reaction mix*	*Reaction 1× (µl)*
hNEIL2 buffer 5×	5×	1×	4
BSA	10 mg/ml	0.1 mg/ml	1
Labelled oligo	1 pmol/ µl	1 pmol	1
Enzyme	Varying	Varying	1
Deionized H_2O			13
Total volume			20
hNEIL3			
Reagents	*Stock*	*Reaction mix*	*Reaction 1× (µl)*
hNEIL3 buffer 5× + 5 mM DTT	5×	1×	4
Labelled oligo	1 pmol/ µl	1 pmol	1
Enzyme	Varying	Varying	1
Deionized H_2O			14
Total volume			20

DTT used was made freshly, the following ratio was used to adjust the concentration of DTT in the reaction mixture.

- Mix 199 µl of 5× HEPES buffer or 5× Tris-EDTA-glycerol buffer, pH 8.0, and freshly made 1 µl of 1 M DTT prior to the experiment. (For hSMUG1 assay/ hUNG assay)
    - Mix 199 µl of 5× Nicking buffer (50 mM Tris-HCl, 1 mM EDTA), pH 8.0, and freshly made 1 µl of 2 M DTT prior to the experiment. (For hTDG assay)

- Mix 199 µl of 5× NEB1 buffer and freshly made 1 µl of 1 M DTT prior to the experiment. (For hOGG1 assay)
- Mix 199 µl of 5× Nth buffer and freshly made 1 µl of 1 M DTT prior to the experiment. (For Nth assay)
- Mix 199 µl of 5× hNEIL3 buffer and freshly made 1 µl of 1 M DTT prior to the experiment. (For hNEIL3 assay)

1. Mix the reactants and add enzyme at last.
2. Spin down, 4000 rpm for 1 min at room temperature.
3. Incubate at 37°C for 60 min or as specified.
4. After incubation, keep the vials on ice and spin down. Terminate the reaction with 45 µl of stop solution and add 1 µl of Proteinase K *i.e.* 10 µg from stock 10 mg/ml. Mix up and down by the pipet and spin down.
5. Prepare cold 96% ethanol with 0.1 M sodium acetate (CH_3COONa) and add 150 µl to the mixture. Add 1.6 µl of tRNA *i.e.* 16 µg on each sample from stock 10 mg/ml. Invert the tubes several times.
6. Incubate the tubes at –70°C for 2 h, or overnight at –20°C in darkness.
7. Centrifuge the tubes at 13000 rpm for 15 min two times with different position of the tubes at 4°C. Remove the supernatant.
8. Add 300 µl cold 70% ethanol.
9. Centrifuge at 13000 rpm for 5 min at 4°C. Remove the supernatant.
10. Centrifuge at 13000 rpm for 1 min at 4°C. Remove the rest of the supernatant using pipette.
11. Dry the pellet for 20 min at room temperature in the fume hood on ice (4°C).
12. Dissolve pellet in 10 µl deionized water (for samples without NaOH) or 10 µl 0.1 M NaOH (for heating 10 min at 90°C).
13. Mix with 10 µl of loading buffer and heat all the samples for 5 min at 90°C, mix properly.
14. Load 5 µl of each sample in the prewashed wells of polyacrylamide gel with urea (after the gels are ready and in the chamber).
15. Run the gel at room temperature in darkness.

Protocol 2: Hybridization of DNA substrate

Vials kept on ice and in darkness during the assay.

1. Mix in PCR-tube:
    a. 2 µl forward ssDNA (100 pmol/µl).
    b. 2 µl of complementary strand (100 pmol/µl).
    c. 16 µl of 1× Salt-Tris-EDTA (STE) buffer.
2. Incubate at 95°C for 4 min in the thermocycler.
3. Leave the tube in the thermocycler to cool down at 1°C per min for 2 h.
4. Dilute with 180 µl 1× Tris-EDTA (TE) buffer to make 1 pmol /µl. Store in aliquots at –20°C in the dark.

Protocol 4: Assembly and preparation of gel for Bio-Rad Mini-Protean Tetra Cell system

1. Assemble a gel cassette assembly with one casting frame, a spacer Plate, and a short Plate. The gap between the two plates is 0.75 mm. Fix it on a casting stand using gel cassette against the silicon casting gaskets.
2. Prepare the mixture of 3.36 gm of urea, 3.5 ml of 40% acrylamide solution, 700 µl of 10X TBE buffer, 280 µl of deionized water.
3. Microwave for 8–10 seconds to heat the solution to dissolve the urea into the solution.
4. The solution was degassed for 10 minutes and left to cool down.
5. 3.5 µl of TEMED and 35 µl of 10% APS solution was added to the mixture and mixed (stored at 4 ºC)
6. Transfer the mixture with transfer pipette and fill the gap between glass plate, put the 10 or 15 wells comb into it. Leave it for 30 minutes to polymerize.
7. After the gel is polymerized, remove the gel cassette sandwich from casting strand and casting frame. Assemble the gel cassette sandwich into the Electrode Assembly with Buffer Dam (use when running 1 or 3 gels) and put it on the Mini Tank. Use companion assembly when running 3 or 4 gels.
8. Fill the gel chamber with running buffer and carefully remove the comb. Wash the gel wells carefully using the 1000 ml pipette to remove the residual gel particles. This part is very critical in making the DNA bands sharper and straight.

9. Load 5 µl of loading solution with sample DNA into the wells carefully and run the gels at 200V for 2 h in darkness.

Protocol 4: Gel analysis

1. After the run of gel is complete, disconnect the power supply.
2. Remove the gel from the chamber and the cartridge. Clean with water and dry with paper.
3. Scan the gel (together with the plates, do not dissemble). If there is more than one gel, scan together at the same time.
4. For scanning of gel choose the fluorescence setup in Typhoon scanner trio of GE healthcare. For cy3 tagged oligomers, choose the excitation wavelength of 532 nm in green region and select emission filter at 580BP. For FAM tagged oligomers, select excitation wavelength of 488 nm in blue region and emission filter at 520BP. Photomultiplier tube voltage settings can be changed to vary the sensitivity for taking different measurements depending on whether the gel is used for quantification analysis or presentation.

Table A1: DNA oligomers

Name	*Manufacturer*	*Label*	*Sequence*
F-Cy3U60	Sigma Aldrich	Cyanine 3 (Cy3) dye on 5' with uracil at 21st position	5'-TAGACATTGCCCTCGAGGTA UCATGGATCCGATTTCGACCTC AAACCTAGACGAATTCCG-3'
R-Gcomp60	Sigma Aldrich	Complementary strand for hybridization, opposite G	5'-CGGAATTCGTCTAGGTTTGA GGTCGAAATCGGATCCATGGTA CCTCGAGGGCAATGTCTA-3'
F-Cy3U60	Sigma Aldrich	Cy3 on 5' with uracil at 12th position	5'-CCCTCGAGGTAUCATGGATCC GATCGATCCGATTTCGACCTCAAA CCTAGACGAATTCCG-3'
R-Gcomp60	Sigma Aldrich	Complementary strand opposite G for hybridization	5'-CGGAATTCGTCTAGGTTTGAGG TCGAAATCGGATCGATCGGATCCA TGGTACCTCGAGGG-3'
F- Cy3U30	Sigma Aldrich	Cy3 on 5' with uracil at 12th position	5'-CCCTCGATGTAUCATGGAT CCGATCGATCC-3'
R- Gcomp	Sigma Aldrich	Opposite G for hybridization	5'-GGATCGATCGGATCCATGGT ACATCGAGGG -3'
F-m^{N4,5}C	Sigma Aldrich	Cy3 on 5' with m^{N4,5}C lesion	5'-C*G*G*TGAAGTAC[m^{N4,5}C]A GGAAGCGATTTCGA*C*C*C-3'
R-m^{N4,5}C	Sigma Aldrich	m^{N4,5}C lesion	5'-G*G*G*TCGAAATCGCTTC [m^{N4,5}C]TGGTACTTCA*C*C*G-3'
R-Gcomp	Sigma Aldrich	Unlabeled, opposite G for hybridization	5'-G*G*G*TCGAAATCGCTTCC TGGTACTTCA*C*C*G-3'
R-Acomp	Sigma Aldrich	Opposite A for hybridization	5'-G*G*G*TCGAAATCGCTTCCT AGTACTTCA*C*C*G-3'
R-Tcomp	Sigma Aldrich	Opposite T for hybridization	5'-G*G*G*TCGAAATCGCTTCCT TGTACTTCA*C*C*G -3'
R-Ccomp	Sigma Aldrich	Opposite C for hybridization	5'-G*G*G*TCGAAATCGCTTC CTCGTACTTCA*C*C*G-3'
Cy38oxoG30	Sigma Aldrich	Cy3 on 5' with oxo^{8}G lesion	5'- CCAGTGATGGG[**oxo^{8}G**]TTCC CATCACTAGTCGAC -3'
Rv8oxoGcomp	Sigma Aldrich	Opposite C for hybridization	5'- GTCGACTAGTGATGGGAAC CCCATCACTGG -3'
F-Gcomp60	Gift §	Forward G for hybridization	5'-TAGACATTGCCCTCGAGGTACC ATGGATCCGATGTCGACCTCAAAC CTAGACGAATTCCG -3'
R-T60FAM	Gift §	Fluorescein on 5' with T at 24th position	5'- CGGAATTCGTCTAGGTTTGAGG TTGACATCGGATCCATGGTACCTC GAGGGCAATGTCTA-3'

§ Gift from Prof. Primo Schär and David Schürmann (University of Basel, Switzerland)

Table A2: Buffers

Buffers	***Composition (for 5×)***	***Stock***	***Mix preparation***
HEPES buffer (5×) (hSMUG1)	225 mM HEPES, pH 7.5	1 M (lab stock)	22.5 ml
	10% glycerol	85% (Merck, Cat # 1.04094)	12 ml
	2 mM EDTA	0.5 M (lab stock)	400 µl
	Deionized H_2O		to 100 ml
Tris buffer (5×)	100 mM Tris, pH 7.5	1 M (lab stock)	10 ml
	300 mM NaOH	3 M (lab stock)	10 ml
	5 mM EDTA	0.5 M (lab stock)	1 ml
	Deionized H_2O		to 100 ml
Tris-HCl-EDTA-nicking-buffer (5×) (hTDG)	250 mM Tris, pH 8	1 M (lab stock)	6.25 ml
	5 mM EDTA	0.5 M (lab stock)	0.25 ml
	Deionized H_2O		to 25 ml
Buffer (1×) in 10% glycerol for hTDG enzyme dilution	50 mM Tris-HCl, pH 8.0	1 M Tris-HCl	50 µl
	10% glycerol	85% (Merck)	117.6 µl
	50 mM NaCl	5 M NaCl (lab stock)	10 µl
	5 mM β-mercaptoethanol	7.8 mM	0.68 µl
	Deionized H_2O		to 1 ml
Running buffer (1×) for uracil glycosylase inhibitor (UGI)	20 mM Tris-HCl, pH 8.0	1 M Tris-HCl	20 µl
	1 mM DTT	1 M DTT	1 µl
	1 mm EDTA	500 mM EDTA	2 µl
	Deionized H_2O		to 1 ml
Tris-EDTA-glycerol buffer (5×) (hSMUG1)	225 mM Tris-HCl, pH 7.5	1 M (lab stock)	5.625 ml
	10% glycerol	85% (Merck)	2.94 ml
	2 mM EDTA	500 mM EDTA	0.1 ml
	Deionized H_2O		to 25 ml
NEB2 (5×) (hOGG1)	250 mM NaCl	5 M NaCl	2.5 ml
	50 mM Tris-HCl, pH 7.9	1 M (lab stock)	2.5 ml
	50 mM $MgCl_2$	500 $MgCl_2$	5 ml
	Deionized H_2O		to 25 ml
Tris-HCl-EDTA (5×) Nth buffer	100 mM Tris-HCl, pH 8.0	1 M (lab stock)	5 ml
	5 mM EDTA	500 mM EDTA	0.5 ml
	Deionized H_2O		to 25 ml
NEB1 (10×) (Fpg) (used from commercial source)	100 mM Bis-Tris-Propane-HCl, pH 7.0	-	-
	100 mM $MgCl_2$	-	-
	Deionized H_2O	-	-
Nei Buffer (5×) (Nei and hNEIL1)	50 mM Tris HCl pH 8	1 M (lab stock)	1.25 ml
	375 mM NaCl	5M NaCl	1.875 ml
	5 mM EDTA	500 mM EDTA	0.25 ml
	Deionized H_2O		to 25 ml

hNEIL2 Buffer (5×)	50 mM Tris HCl pH 7.5	1 M (lab stock)	1.25 ml
	250 mM NaCl	5M NaCl	1.25 ml
	5 mM EDTA	500 mM EDTA	o.25 ml
	Deionized H_2O		to 25 ml
hNEIL3 Buffer (5×)	250 mM MOPS pH 7.5	1 M (lab stock)	6.25 ml
	25% glycerol	85% glycerol	7.35 ml
	5 mM EDTA	500 mM EDTA	0.25 ml
	Deionized H_2O		to 25 ml

Stored at –20°C in aliquots.

Table A3: 96% ethanol with 0.1 M CH_3COONa (for DNA precipitation)

Composition	***Stock***	***Mix preparation***
0.1 M NaOAc	Sigma, Cat. # S2889, 82.03 g/mol	0.82 g
96% EtOH	99.5%	225 µl

Stored at room temperature.

Table A4: 1 M KCl

Composition	***Stock***	***Mix preparation***
1 M KCl	Sigma, Cat. # 5405, 74.55 g/mol	3.72 g
Deionized H_2O	96%	50 ml

Filtrated and stored at –20°C in aliquots.

Table A5: Stop solution

Composition	***Stock***	***Mix preparation***
20 mM EDTA	0.5 M (lab stock)	2 ml
Sodium dodecyl sulphate (SDS), 0.5% (w/v)	99% (Sigma, Cat. # L3771, 288.38 g/mol)	252 mg
Deionized H_2O		to 50 ml

Stored at room temperature.

Table A6: Salt-TE (STE) buffer (for DNA oligomer hybridization)

Composition	***Stock***	***Mix preparation***
10 mM Tris, pH 8.0	1 M	200 µl
50 mM NaCl	Sigma, Cat. # S5886, 58.44 g/mol	58.44 mg
1 mM EDTA	0.5 M (lab stock)	40 µl
Deionized H_2O		to 20 ml

Filtrated and stored in aliquots at –20°C.

Table A7: Loading buffer (Note: no bromophenol blue)

Composition	*Stock*	*Mix preparation*
Formamide, 80%	99.5% (Sigma, Cat. # F9037)	40.20 ml
1 mM EDTA	0.5 M	100 µl
Blue dextran, 1% (w/v)	Sigma, Cat. # D5751	0.5 g
Deionized H_2O		to 50 ml

Stored at –20°C in aliquots.

Table A8: TE buffer (1×)

Composition	*Stock*	*Mix preparation*
10 mM Tris, pH 7.5	1 M	200 µl
1 mM EDTA, pH 8.0, sterile filtrated	0.5 M	1.6 ml
Deionized H_2O		to 20 ml

Filtrated and stored in aliquots at –20°C.

Table A9: Taurine (20×), running buffer

Composition	*Stock*	*Mix preparation*
Tris base, 1.78 M	Sigma, Cat. # T6066, 121.14 g/mol	216 g
Taurine, 0.58 M	Sigma, Cat. # T0625, 125.15 g/mol	72 g
$Na_2EDTA{\times}2H_2O$	Lab stock	4 g
Deionized H_2O		to 1000 ml

Autoclaved before use.

Table A10: TBE (10×), running buffer

Composition	*Stock*	*Mix preparation*
Tris base, 890 mM	Sigma, Cat. # T6066, 121.14 g/mol	108 g
Boric acid, 890 mM	MW 61.8	55 g
20 mM EDTA, pH 8	500 mM (lab stock)	40 ml
Deionized H_2O		to 1000 ml

Autoclaved before use.

Table A11: Denaturing 20% PAGE gel with urea 7 M (taurine buffer system)

Composition	*Stock*	*Big gel*	*Small gel*
Polyacrylamide, 20% (w/v)	40% (Saveen Werner AB, Cat. # BIAC21)	12.5 ml	4.166 ml
Taurine buffer (1×)	Taurine (20×) (lab stock)	1.25ml	416.6 µl
Urea	99.5% (Sigma, Cat # F9037)	10.5 g	3.5 g
Deionized H_2O		3.75 ml	1.25 ml
Ammonium persulfate (APS)	BioRad, Cat. # 161-0700, 228.2 g/mol	125 µl (from 10% prepared stock)	41.66 µl
Tetramethylethylenediamine (Temed)	Invitrogen Cat. # 15524-010	25 µl	8.33 µl

Table A12: Denaturing 20% PAGE gel with urea 8 M (TBE buffer system)

Composition	*Stock*	*Small gel*
Polyacrylamide, 20% (w/v)	40% (Saveen Werner AB, Cat. # BIAC21)	3.5 ml
TBE (1×)	10× TBE (lab stock)	700 µl
Urea	99.5% (Sigma, Cat. # F9037)	3.363 g
Deionized H_2O		280 µl
Ammonium persulfate (APS)	(BioRad, Cat. # 161-0700, 228.2 g/mol)	35 µl
Tetramethylethylenediamine (Temed)	(Invitrogen Cat. # 15524-010)	3.5 µl

Table A13: 20% acrylamide native PAGE (Taurine buffer system)

Composition	*Stock*	*Small gel*
Polyacrylamide, 20% (w/v)	40% (Saveen Werner AB, Cat. # BIAC21)	3.25 ml
Taurine (1×)	Taurine (20×) (lab stock)	375 µl
Deionized H_2O		3.375 ml
Ammonium persulfate (APS)	(BioRad, Cat. # 161-0700, 228.2 g/mol)	75 µl
Tetramethylethylenediamine (Temed)	(Invitrogen Cat. # 15524-010)	7.5 µl

Table A14: Enzyme concentration calculations

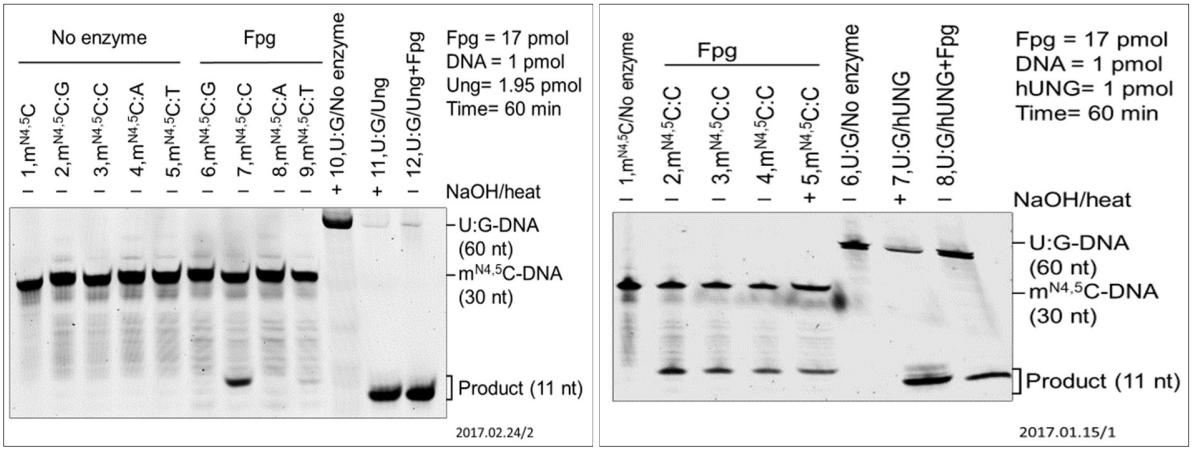
Enzyme	*MW (kDa)*	*Concentration (ng/µl)*	*Concentration (pmol/µl)*	*Concentration (units/ml)*
Fpg (Full length)	30	500	16.67	8000
Nei (Full length)	29.845	62	2.08	10000
Nth (Full length)	23	16	0.695	10000
hSMUG1 (Full length)	29.86	10	0.33	5000
hTDG (His tag + Full length)	46.876*	703.1	15	-
hOGG1 (His tag + AA 12-327)	36.399*	196.6	5.4	-
hNEIL1 (Full length + His tag)	44.524*	200	4.49	-
hNEIL2 (Full length + His tag)	37.666*	200	5.31	-
hNEIL3 (AA 1-301 + His tag)	34.754*	300	8.63	-

*Protein MW with the addition of the MW of His-tag.

Appendix B

Fpg experiments

A **B**



C

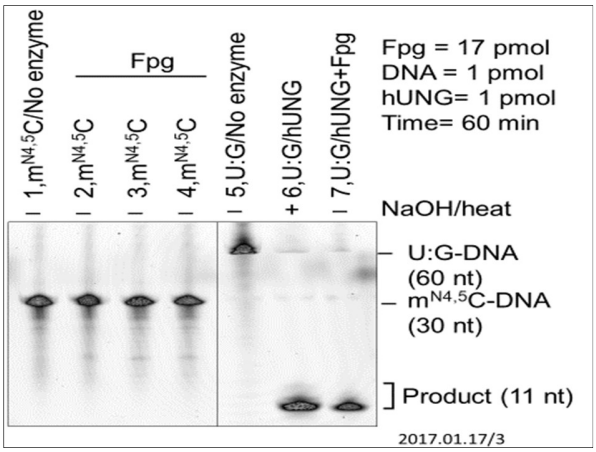
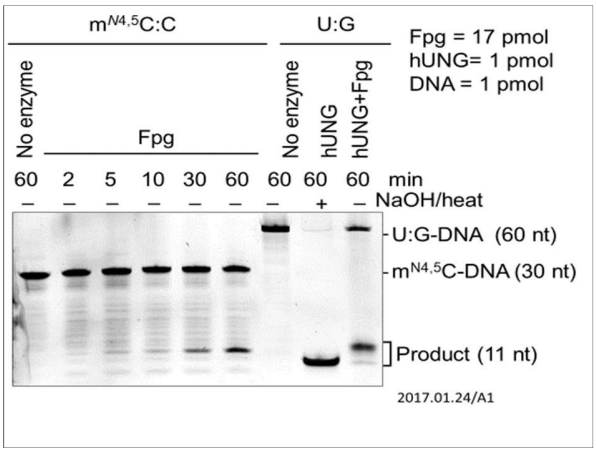


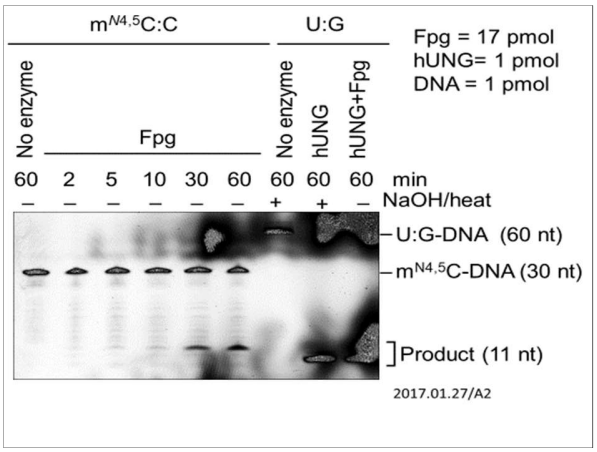
Figure B1: Activity on m^{N4,5}C-DNA by Fpg.
Fpg (17 pmol) was incubated with m^{N4,5}C or opposite G, C, A and T (1 pmol) in NEB1 buffer (10 mM Bis-Tris-Propane-HCl, pH 7, 10 mM $MgCl_2$), 1 mM DTT, 0.1 mg/ml BSA at 37°C for 1 h (final volume, 20 µl), followed by m^{N4,5}C-DNA glycosylase + AP lyase activity opposite G, C, A, and T. U:G-DNA (1 pmol) incubated without (lane 10) and with hUNG/Ung, the latter followed by NaOH/heat treatment (lane 11), was used as negative and positive control for active hUNG/Ung, respectively, which was used to convert U:G-DNA into AP-DNA to demonstrate active Fpg (*i.e.*, lyase activity; lane 12) (**A**); Three parallels of m^{N4,5}C-DNA glycosylase activity opposite C (**B**); Fpg activity on m^{N4,5}C ssDNA (**C**). Incised was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h. Gels were scanned at PMT 650V.

Fpg time curve

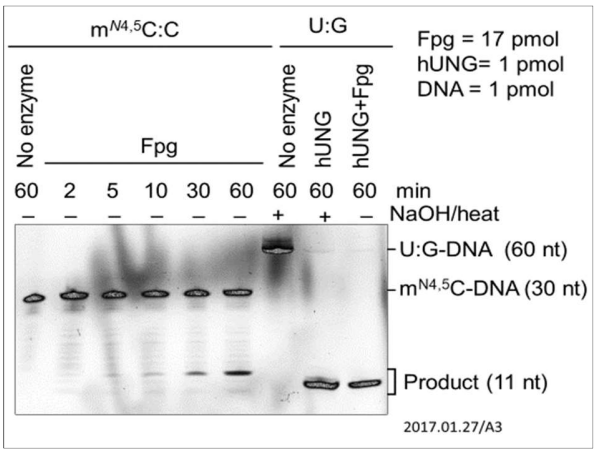
A



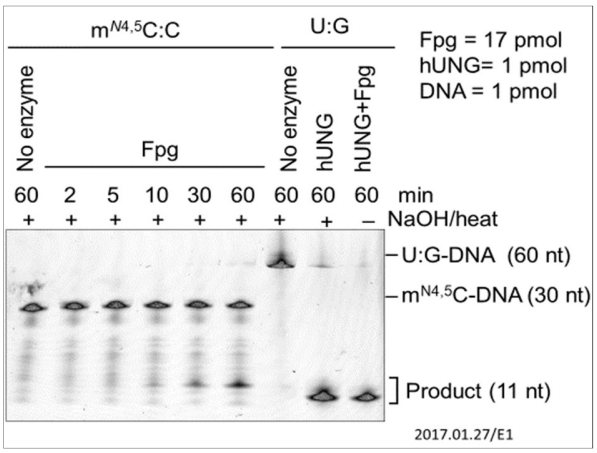
B



C



D



E

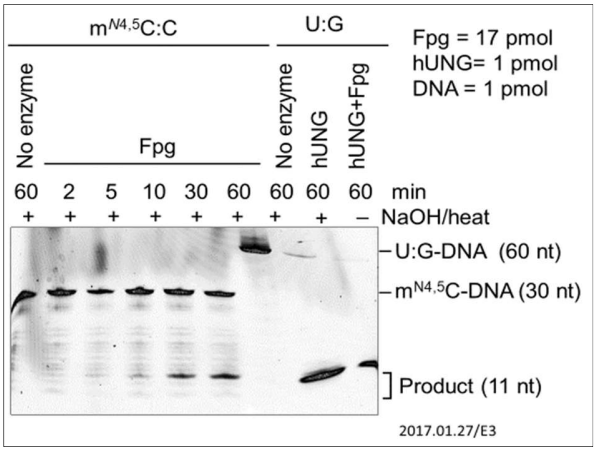
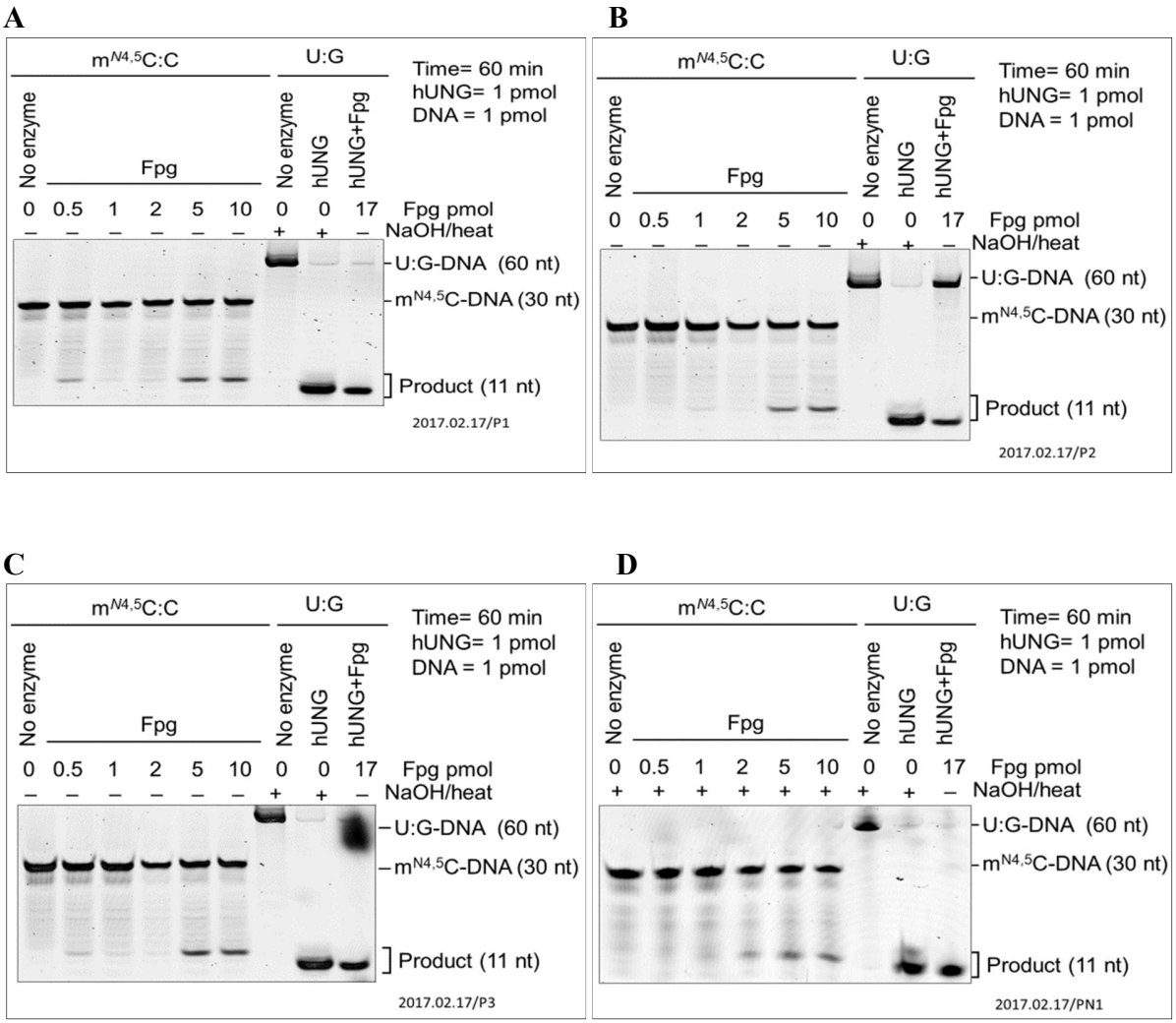
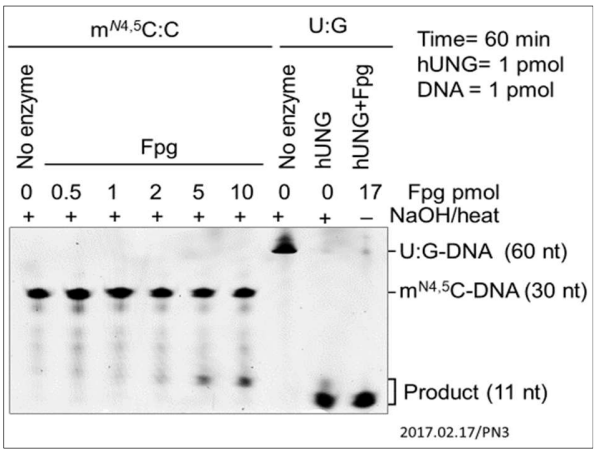


Figure B2: Time dependent activity of Fpg on m^{N4,5}C:C-DNA.
Fpg (17 pmol) was incubated with DNA substrate (1 pmol) in NEB1 buffer (10 mM Bis-Tris-Propane-HCl, pH 7, 10 mM $MgCl_2$), 1 mM DTT, 0.1 mg/ml BSA at 37°C for time periods indicated (final volume 20 µl), followed by NaOH/heat treatment **(D** and **E)** to measure glycosylase activity; **(A, B** and **C)** without alkali/ heat treatment to measure glycosylase + AP lyase activity for the construction of time curve. Gels were scanned at PMT 650 V and data was analyzed.

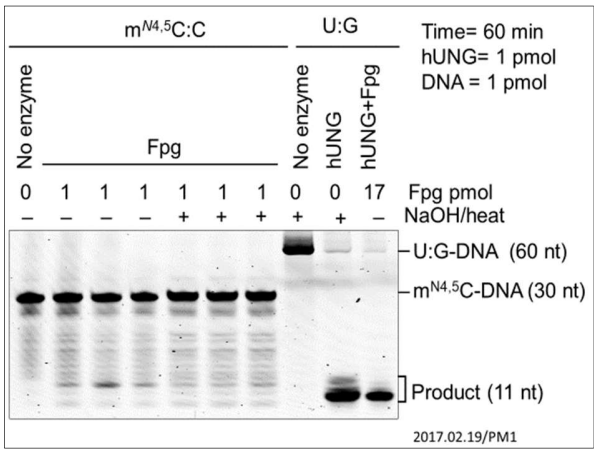
Fpg protein curve experiments



E



F



G

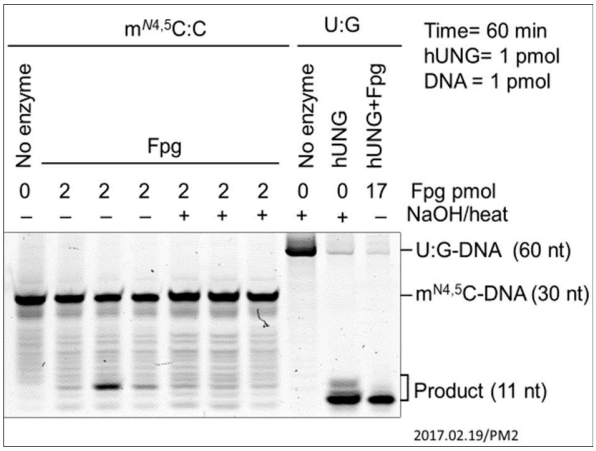
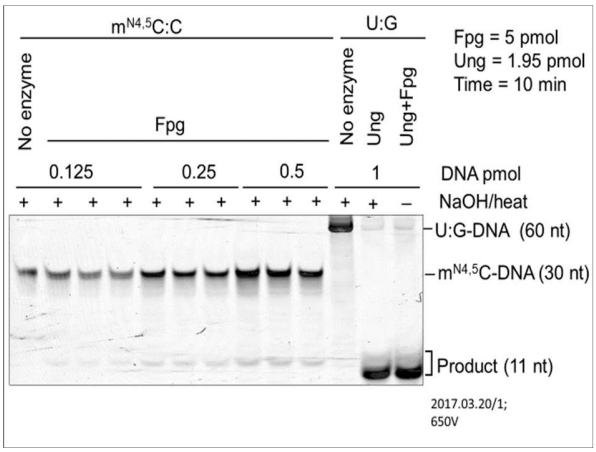


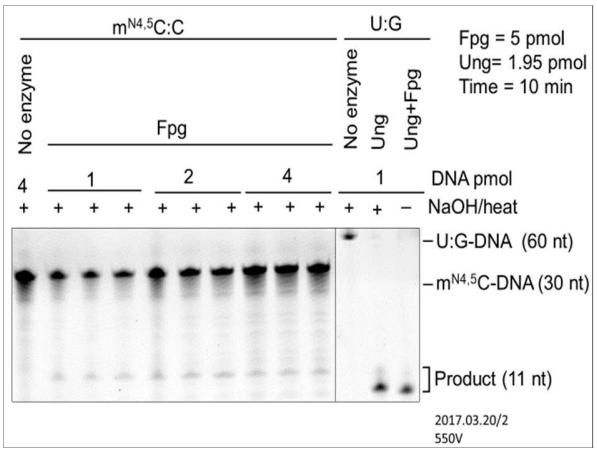
Figure B3: Concentration dependent activity of Fpg on the m^{N4,5}C:C-DNA.
Fpg protein of various concentrations as indicated were incubated with DNA substrate (1 pmol) in NEB1 buffer (10 mM Bis-Tris-Propane-HCl, pH 7, 10 mM $MgCl_2$), 1 mM DTT, 0.1 mg/ml BSA at 37°C for 1 h (final volume 20 µl). 0.5–10 pmol of Fpg protein was reacted with substrate DNA, followed by (**A, B** and **C**) without alkali and heat treatment to measure glycosylase + AP lyase activity; (**D** and **E**) NaOH/heat treatment to measure m^{N4,5}C-DNA glycosylase activity; (**F** and **G**) 1 and 2 pmol of Fpg was reacted with substrate DNA, followed by heat and alkali treatment or without such treatment. Incised DNA was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h. Gels were scanned at PMT 650 V and data was analyzed.

Fpg Enzyme kinetics experiments

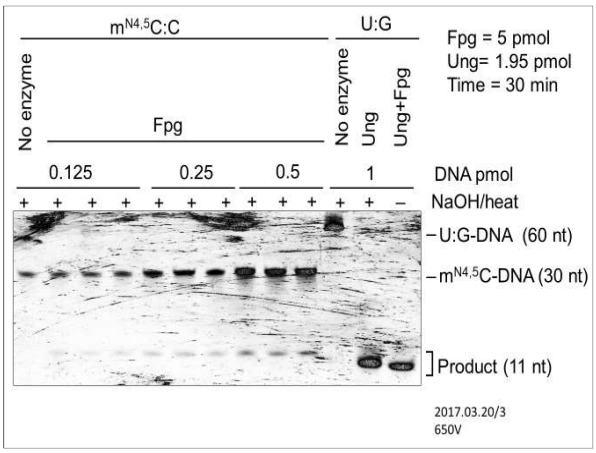
A



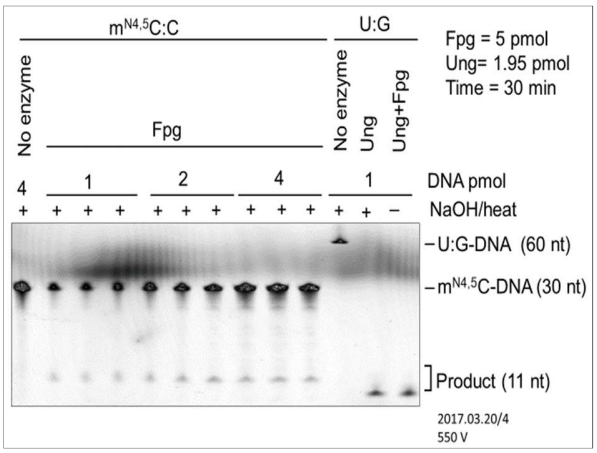
B



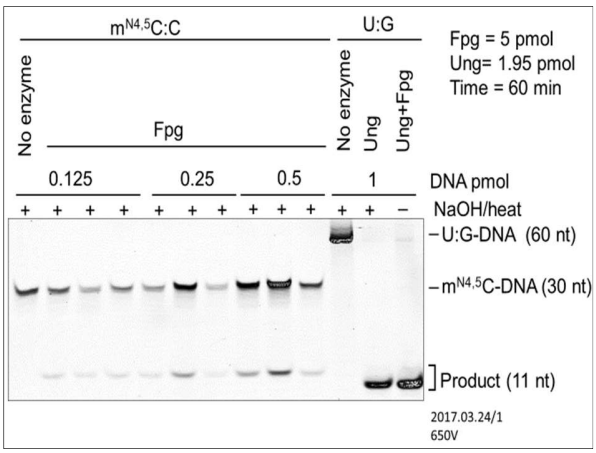
C



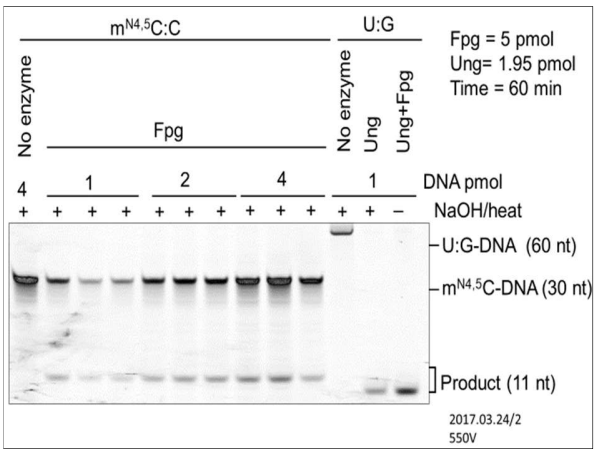
D



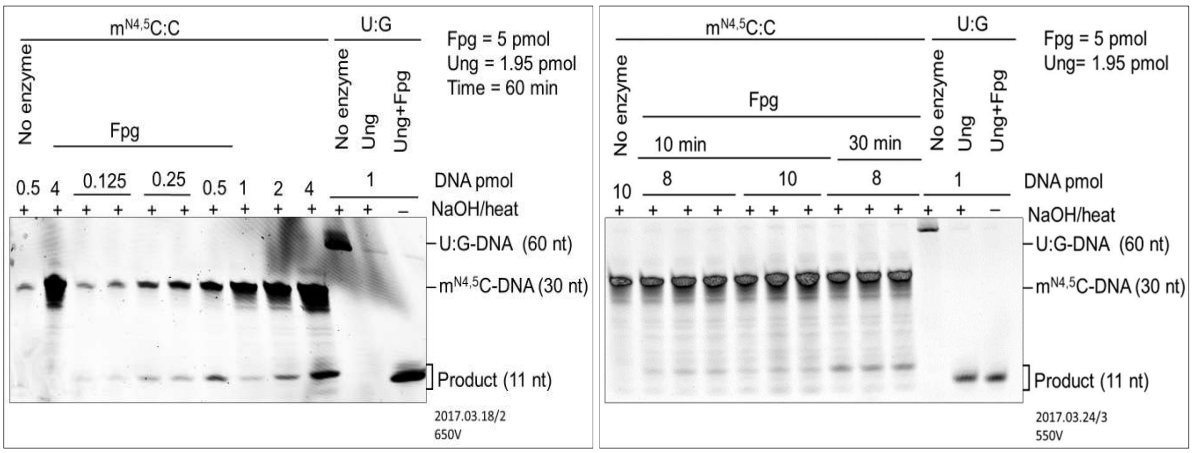
E



F



G **H**



I

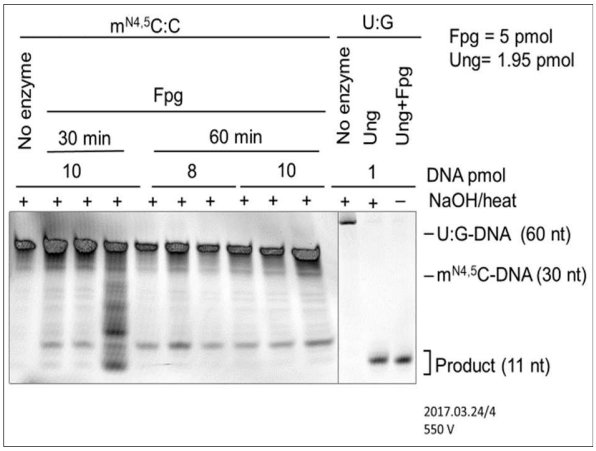


Figure B4: Excision kinetics of Fpg on the $m^{N4,5}C$:C-DNA.
Fpg (5 pmol; 250 nM final concentration) was incubated with various concentrations of DNA substrate as indicated in NEB1 buffer (10 mM Bis-Tris-Propane-HCl, pH 7, 10 mM $MgCl_2$), 1 mM DTT, 0.1 mg/ml BSA at 37°C for different time periods indicated (final volume 20 µl), for 10 min, **(A** and **B)**; 30 min**, (C** and **D)**; 60 min, **(E, F** and **G)**; time as indicated, **(H** and **I**). This was followed by NaOH/heat treatment to measure the excision (glycosylase activity) for the protein. Incised was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h. The PMT voltage is indicated in the figure, all of the data was obtained from the figure scanned at 550 V, however, for the presentation purpose the gels were scanned at 650 V (**A**, **C**, **G** and **E**).

Nei experiments

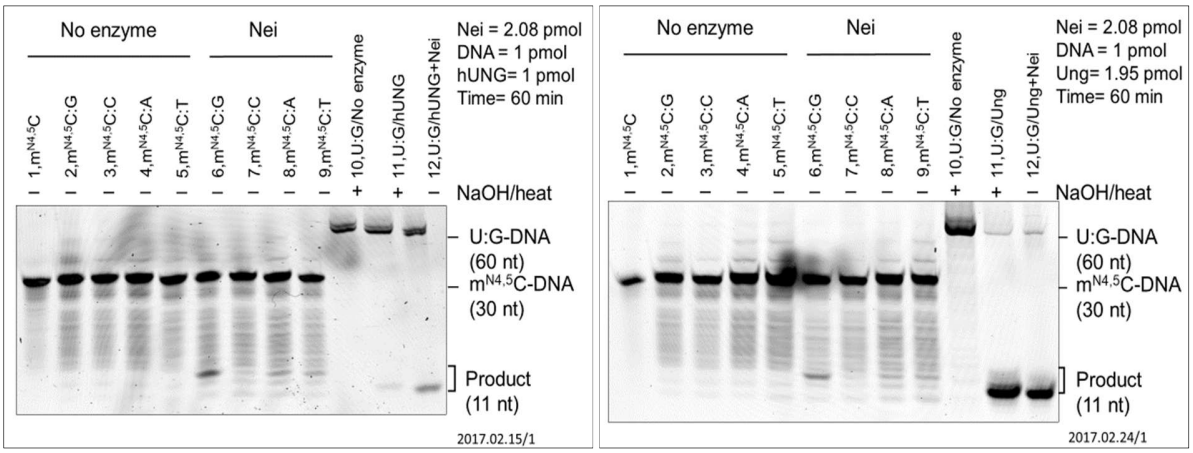
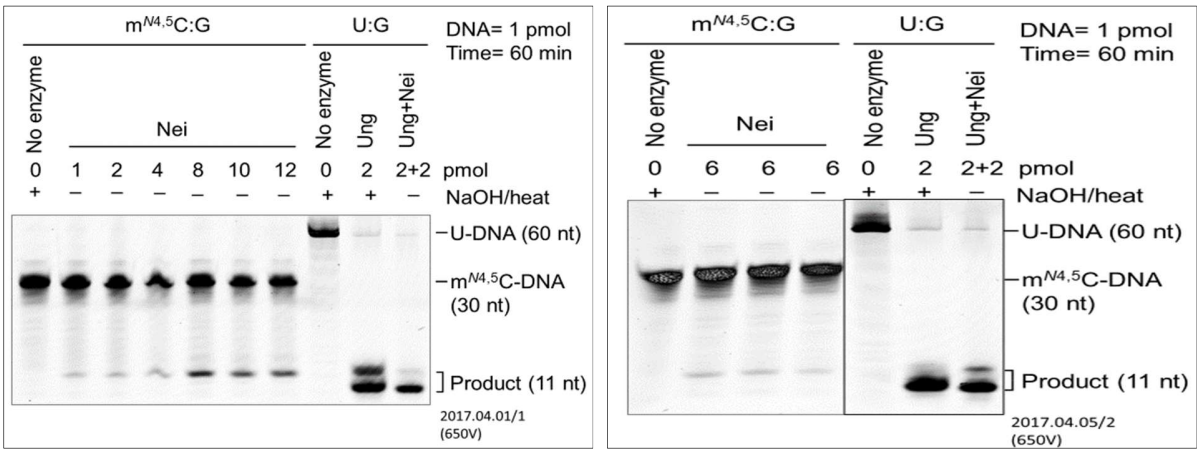


Figure B5: Activity on m^{N4,5}C-DNA by Nei.
Nei (2.08 pmol) was incubated with DNA substrate (1 pmol) in Nei buffer (10 mM Tris-HCl, pH 8.0, 75 mM NaCl, and 1 mM EDTA) at 37°C for 1 h (final volume, 20 µl) to observe m^{N4,5}C-DNA glycosylase activity opposite G, C, A, and T. U:G-DNA (1 pmol) incubated without (lane 10) and with hUNG/Ung, the latter followed by NaOH/heat treatment (lane 11), was used as negative and positive control for active hUNG/Ung, respectively, which was used to convert U:G-DNA into AP-DNA to demonstrate active Nei (*i.e.*, lyase activity; lane 12). Incised was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h. Gels were scanned at PMT 650 V.

Nei Protein curve

A **B**



C

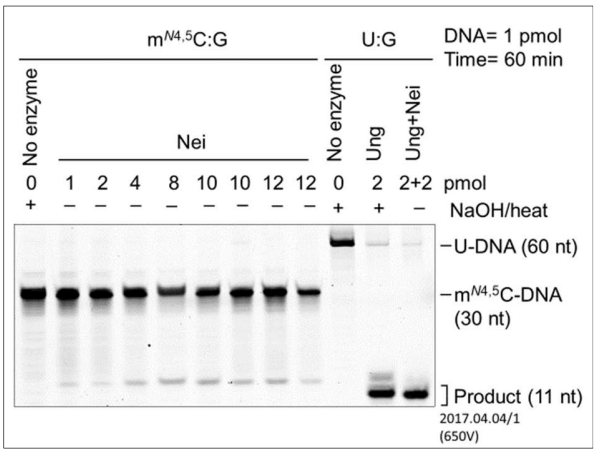


Figure B6: Concentration dependent activity of Nei on the m^{N4,5}C:G-DNA.
Nei protein of various concentrations as indicated were incubated with DNA substrate (1 pmol) in Nei buffer (10 mM Tris-HCl, pH 8.0, 75 mM NaCl, and 1 mM EDTA) at 37°C for 1 h (final volume, 20 µl). 1.04–12.48 pmol of Nei protein was incubated with substrate DNA to measure glycosylase + AP lyase activity. Incised DNA was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h. All of the data was obtained from the gels scanned at 550 V, however, for the presentation purpose the gels were scanned at 650 V.

Nei Time curve

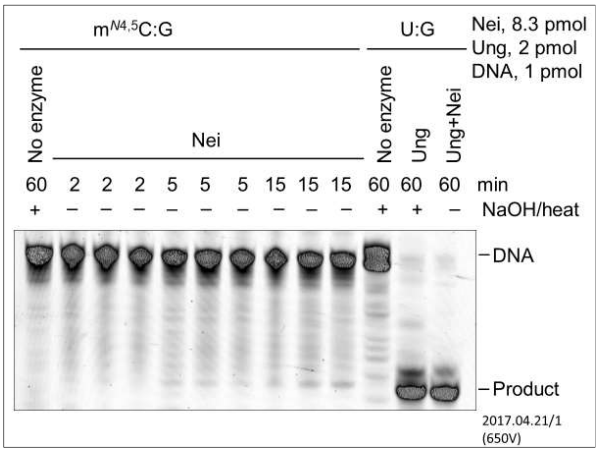
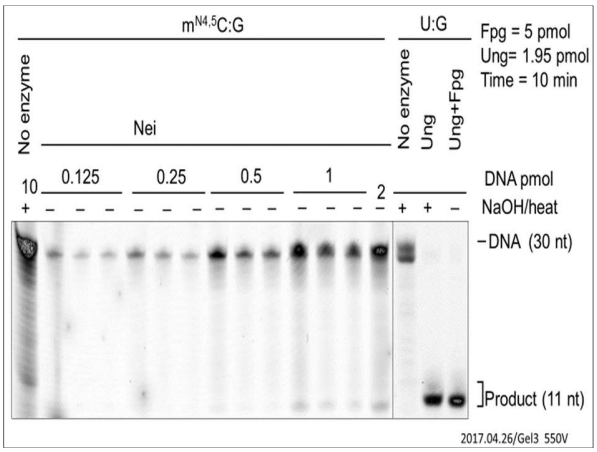


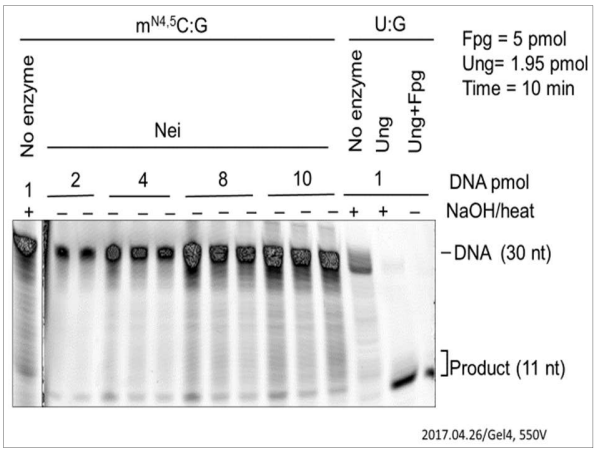
Figure B7: Time dependent activity of Nei on the m^{N4,5}C:G-DNA.
Nei (8.32 pmol) was incubated with DNA substrate (1 pmol) in Nei buffer (10 mM Tris-HCl, pH 8.0, 75 mM NaCl, and 1 mM EDTA) at 37°C for various time as indicated (final volume, 20 µl) to measure glycosylase + AP lyase activity. Incised DNA was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h. All the data was obtained from the gels were scanned at 550 V, however, for the presentation purpose the gels were scanned at 650 V.

Nei Michaelis-Menten experiments

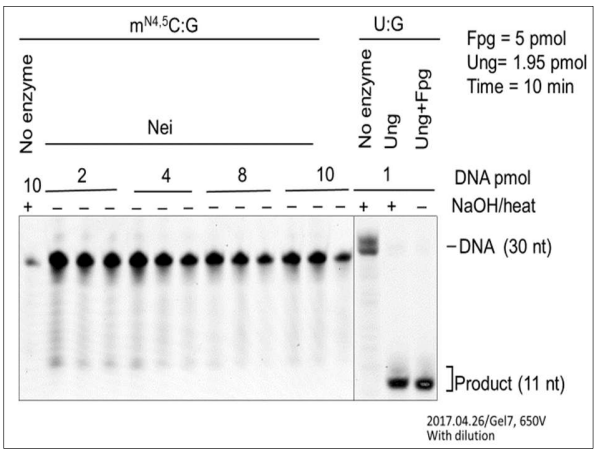
A



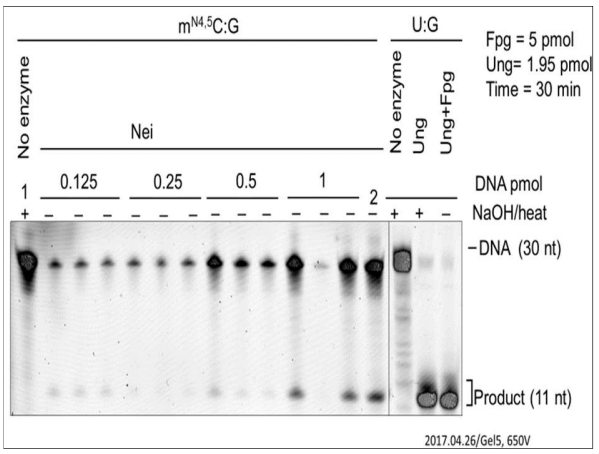
B



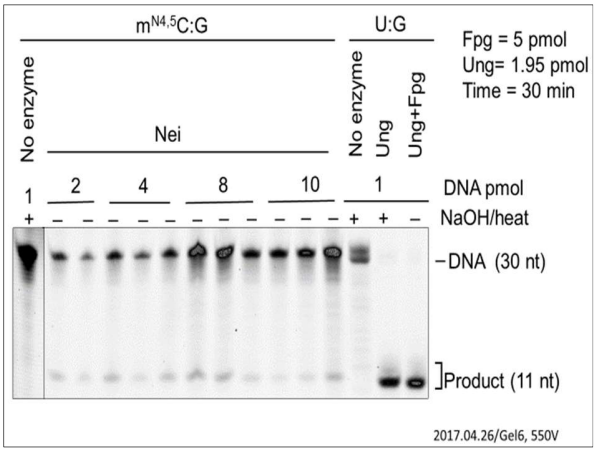
C



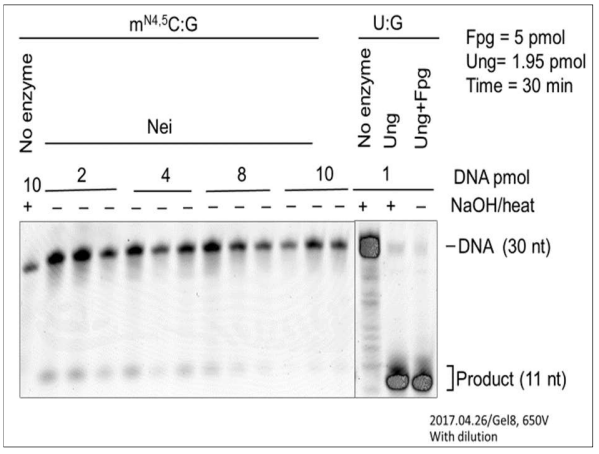
D



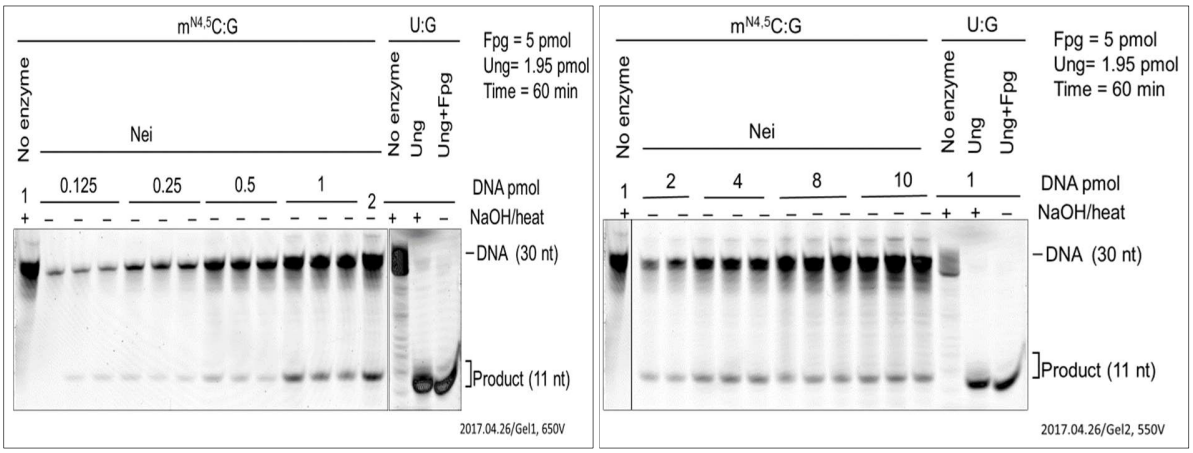
E



F



G **H**



I

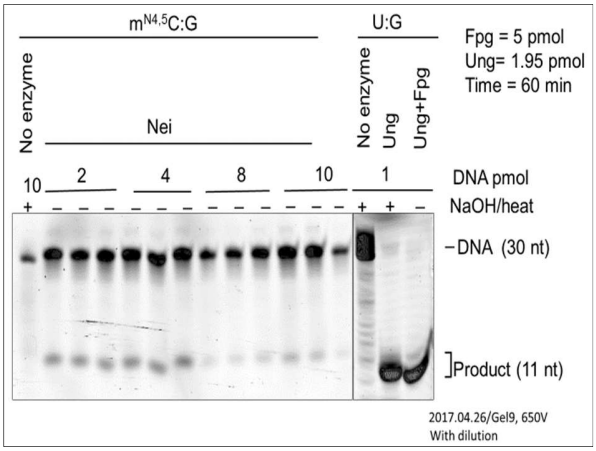
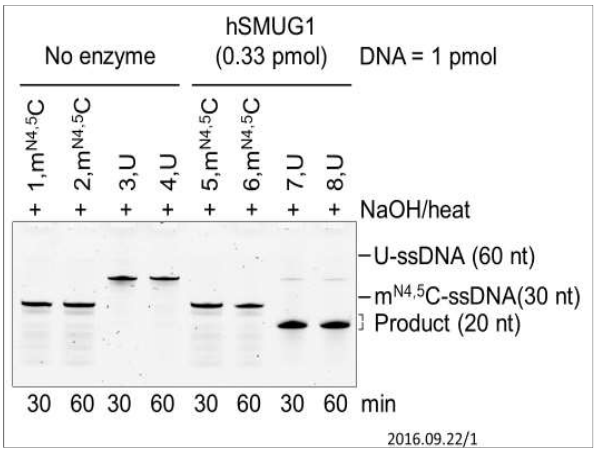


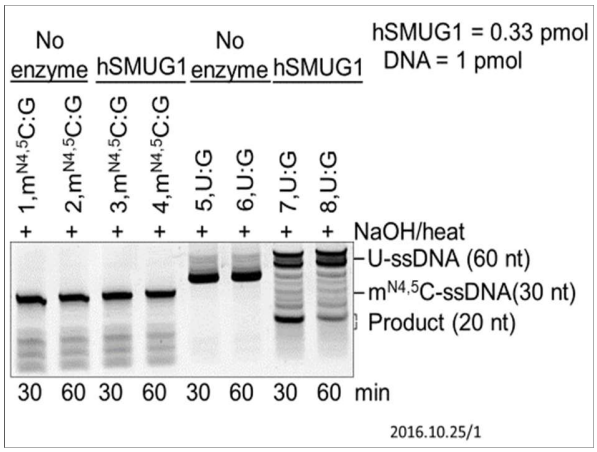
Figure B8: Incision kinetics of Nei on the m^{N4,5}C:G-DNA.
Nei (8.32 pmol; 416 nM final concentration) was incubated various concentrations of DNA substrate in Nei buffer (10 mM Tris-HCl, pH 8.0, 75 mM NaCl, and 1 mM EDTA) at 37°C for various time as indicated (final volume, 20 µl) for 10 min, (**A, B** and **C**); for 30 min, (**D, E** and **F**); for 60 min, (**G, H** and **I**). The incision of DNA was measured to analyze glycosylase + AP lyase activity. The samples having concentrations from 2–10 pmol were diluted to 1 pmol using loading buffer and loaded (**C, F** and **I**) to reduce the error due to detector saturation. Incised DNA was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h. The PMT voltage is indicated in the figure, all data were obtained from the gels scanned at 550 V, however, for the presentation purpose the gels were scanned at 650 V (**C**, **D**, **F**, **G** and **I**).

hSMUG1 experiments

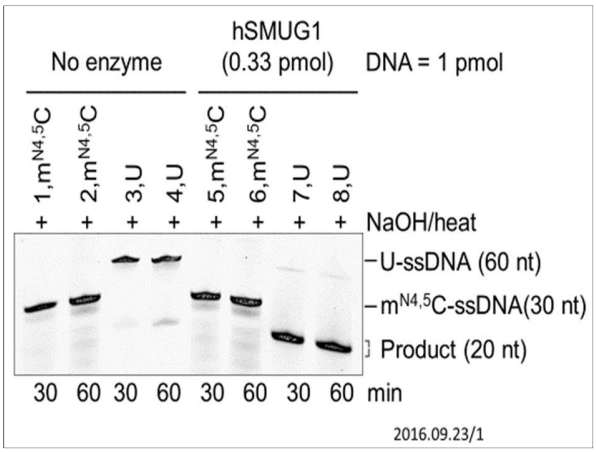
A



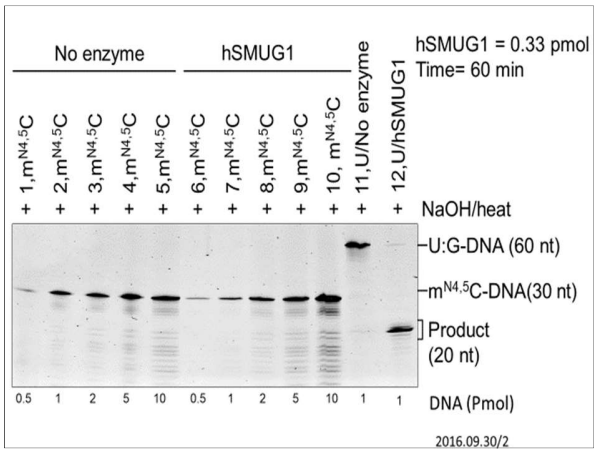
B



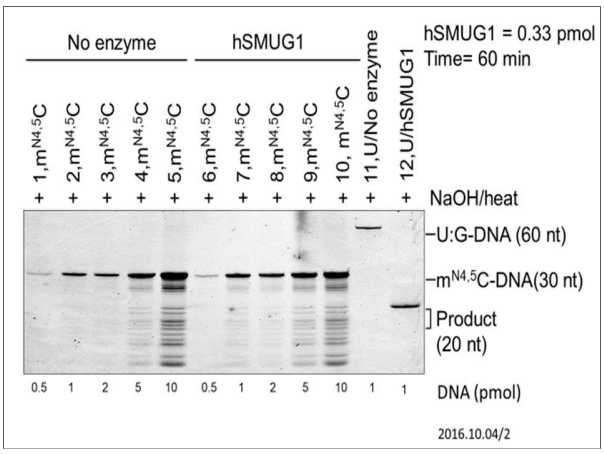
C



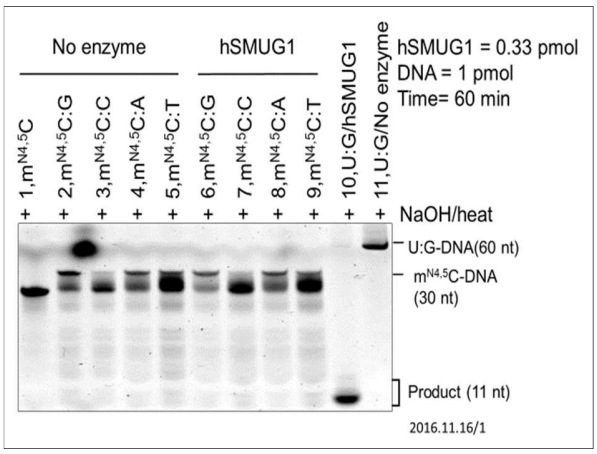
D



E



F



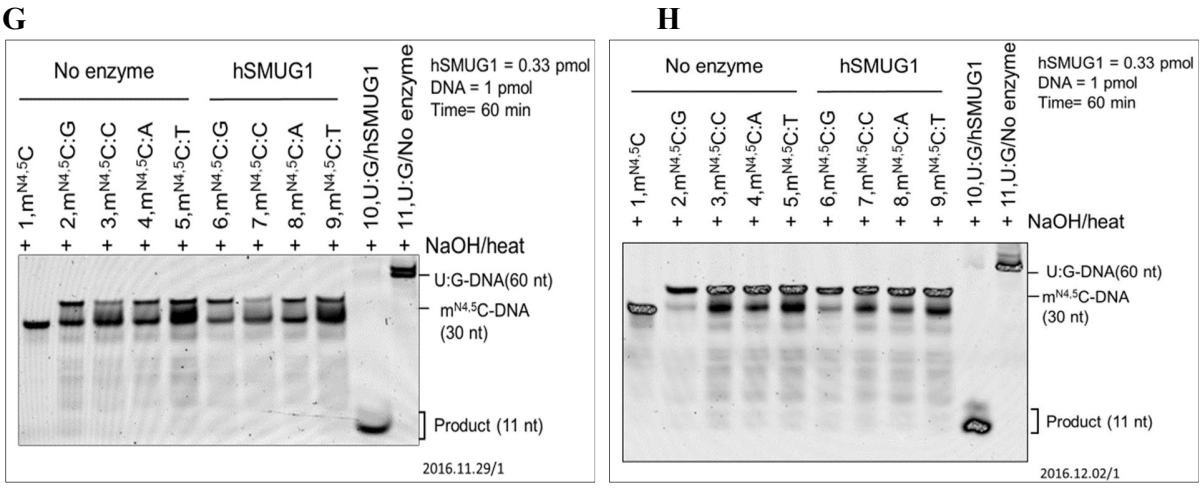


Figure B9: Activity on m^{N4,5}C-DNA by hSMUG1.
hSMUG1 (0.33 pmol) was incubated with DNA substrate (1 pmol) in HEPES buffer (45 mM HEPES, pH 7.5, 0.4 mM EDTA, 2% (v/v) glycerol, 0.1 mg/ml BSA, 1 mM DTT, and 70 mM KCl) at 37°C for 1 h or as indicated (final volume, 20 µl), followed by NaOH and heat treatment. hSMUG1 was shown active by removal of the uracil from U:G DNA following the alkali/heat treated incision. **(A, B, C)** to observe DNA glycosylase activity on m^{N4,5}C- ssDNA for 30 and 60 min incubation. Incised was separated from un-incised DNA by denaturing PAGE at 120 V for 2 h; **(D, E)** ssDNA substrate of various concentrations were reacted with hSMUG1 and incised were separated from un-incised DNA by denaturing PAGE at 600 V for 3h (**D**); and 400 V for 4 h (**E**); m^{N4,5}C-DNA glycosylase activity opposite G, C, A, and T was observed **(F)**; Incised was separated from un-incised DNA by denaturing PAGE at 180V for 2 h. hSMUG1 was incubated with DNA substrate (1 pmol) and electrophoresed at 150 V for 1 h 45 min **(G)**; hSMUG1 was incubated with DNA substrate (1 pmol) in 45 mM Tris-HCl, pH 7.5, 0.4 mM EDTA, 2% (v/v) glycerol, 0.1 mg/ml BSA, 1 mM DTT, and 70 mM KCl) at 37°C for 1 h or as indicated (final volume, 20 µl), followed by NaOH and heat treatment to observe m^{N4,5}C-DNA glycosylase activity opposite G, C, A, and T **(H)**. Incised was separated from un-incised DNA by denaturing PAGE at 150 V for 2 h. All the gels were scanned at PMT 650 V.

hTDG experiments

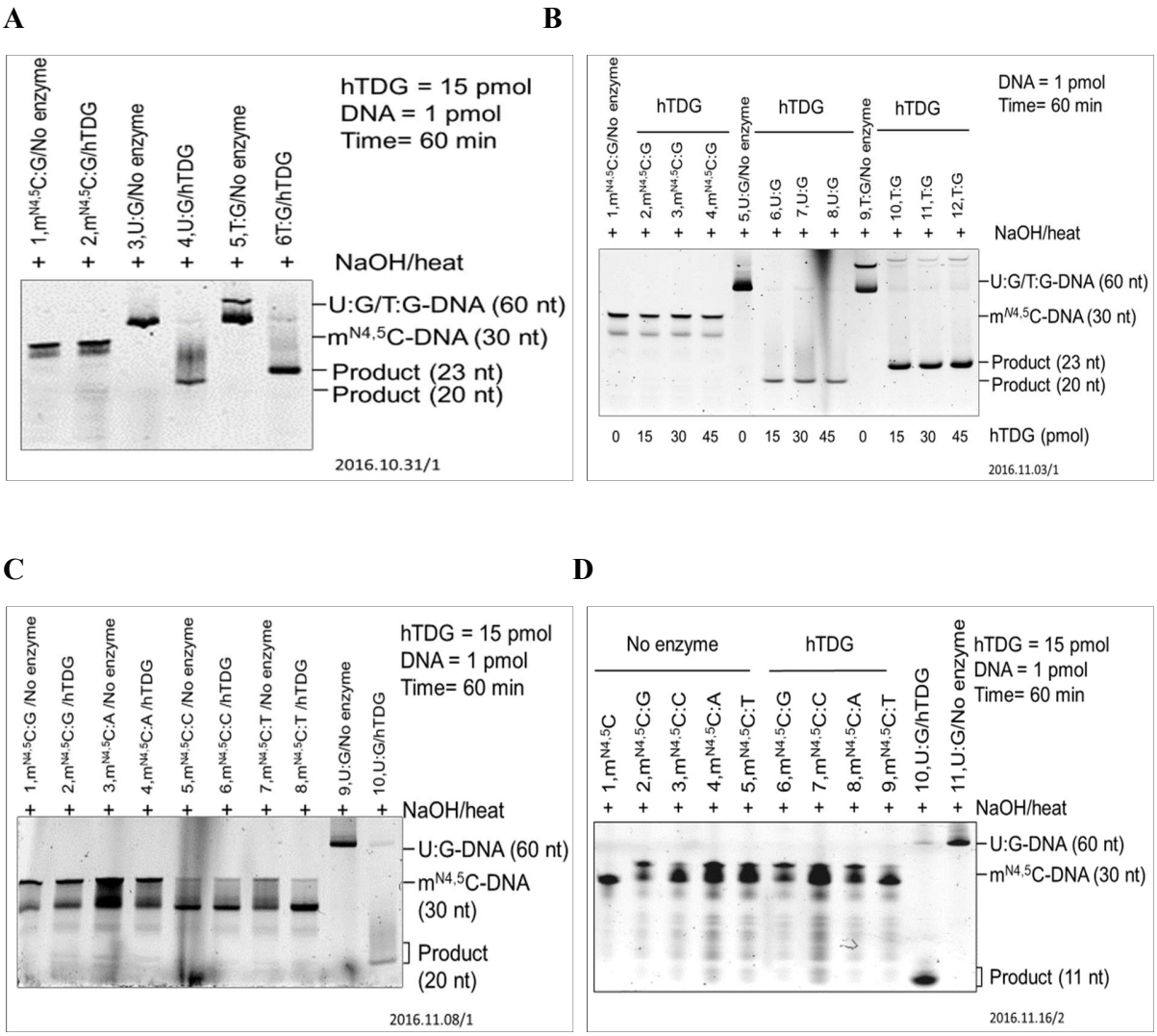


Figure B10: Activity on $m^{N4,5}C$-DNA by hTDG.
hTDG (15 pmol or as indicated) was incubated with DNA substrate (1 pmol) in 1× nicking buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 2 mM DTT), 0.1 mg/ml BSA and 0.5 unit of uracil DNA glycosylase inhibitor (final volume 20 µl) at 37°C for 1 h (final volume, 20 µl), followed by NaOH and heat treatment. **(A)** 15 pmol hTDG was incubated with substrate DNA and hTDG was shown active against cy3 tagged U:G-DNA and FAM tagged T:G-DNA (S11); **(B)** 15, 30, and 45 pmol of hTDG was incubated with 1 pmol of substrate DNA; **(C, D)** 15 pmol of hTDG was incubated with 1 pmol substrate $m^{N4,5}C$-DNA opposite G, C, A and T and examined for the glycosylase activity. Control used in **(A, B, C)** were U:G and T:G-DNA generating 20 and 23 nt incision products respectively. Incised was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h 30 min; whereas 11 nt product was incised in **(D)** and electrophoresed at 180 V for 2 h. Gels were scanned at PMT 650 V.

hOGG1 experiments

A **B**

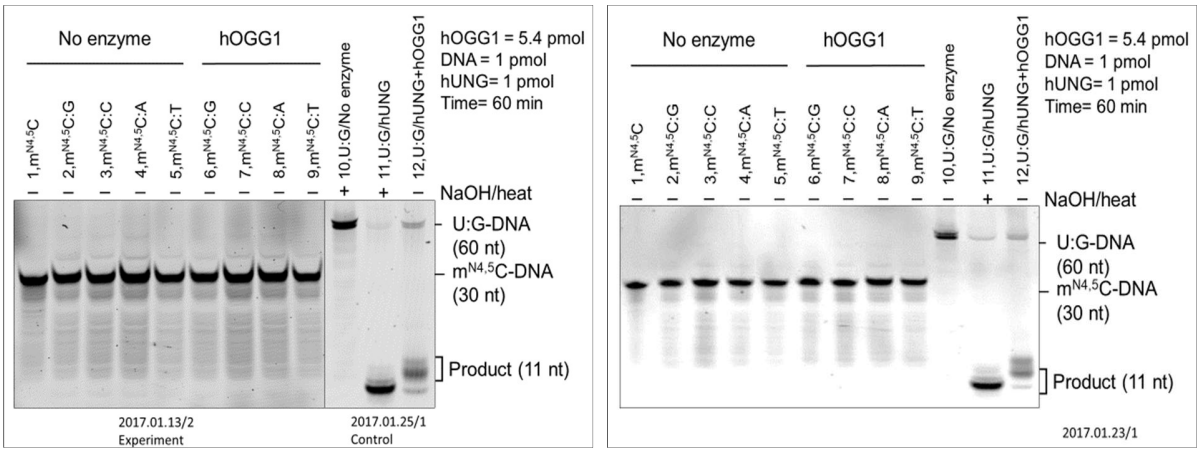


Figure B11: Activity on $m^{N4,5}C$-DNA by hOGG1
hOGG1 (5.4 pmol) was incubated with DNA substrate (1 pmol) in 1× NEB2 buffer (50 mM NaCl, 10 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 1 mM dithiothreitol), and 0.1 mg/ml BSA (final volume 20 µl) at 37°C for 1 h (final volume, 20 µl) to observe $m^{N4,5}C$-DNA glycosylase activity opposite G, C, A, and T. U:G-DNA (1 pmol) incubated without (lane 10) and with hUNG, the latter followed by NaOH/heat treatment (lane 11), was used as negative and positive control for active hUNG, respectively, which was used to convert U:G-DNA into AP-DNA to demonstrate active hOGG1 (*i.e.*, lyase activity; lane 12). Incised was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h. Gels were scanned at PMT 650 V.

Nth experiments

A **B**

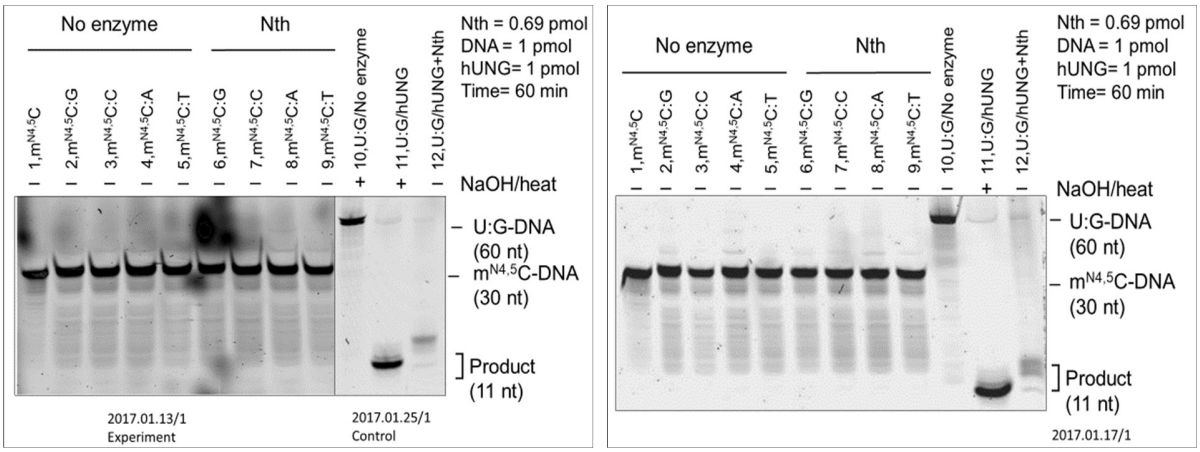


Figure B12: Activity on $m^{N4,5}C$-DNA by Nth.
Nth (0.69 pmol) was incubated with DNA substrate (1 pmol) in 1× TE-Nth buffer (20 mM Tris-HCl, pH 8, 1mM EDTA, 1 mM dithiothreitol), and 0.1 mg/ml BSA (final volume 20 µl) at 37°C for 1 h (final volume, 20 µl) to observe $m^{N4,5}C$-DNA glycosylase activity opposite G, C, A, and T. U:G-DNA (1 pmol) incubated without (lane 10) and with hUNG, the latter followed by NaOH/heat treatment (lane 11), was used as negative and positive control for active hUNG, respectively, which was used to convert U:G-DNA into AP-DNA to demonstrate active Nth (*i.e.*, lyase activity; lane 12). Incised was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h. Gels were scanned at PMT 650 V.

hNEIL1 experiments

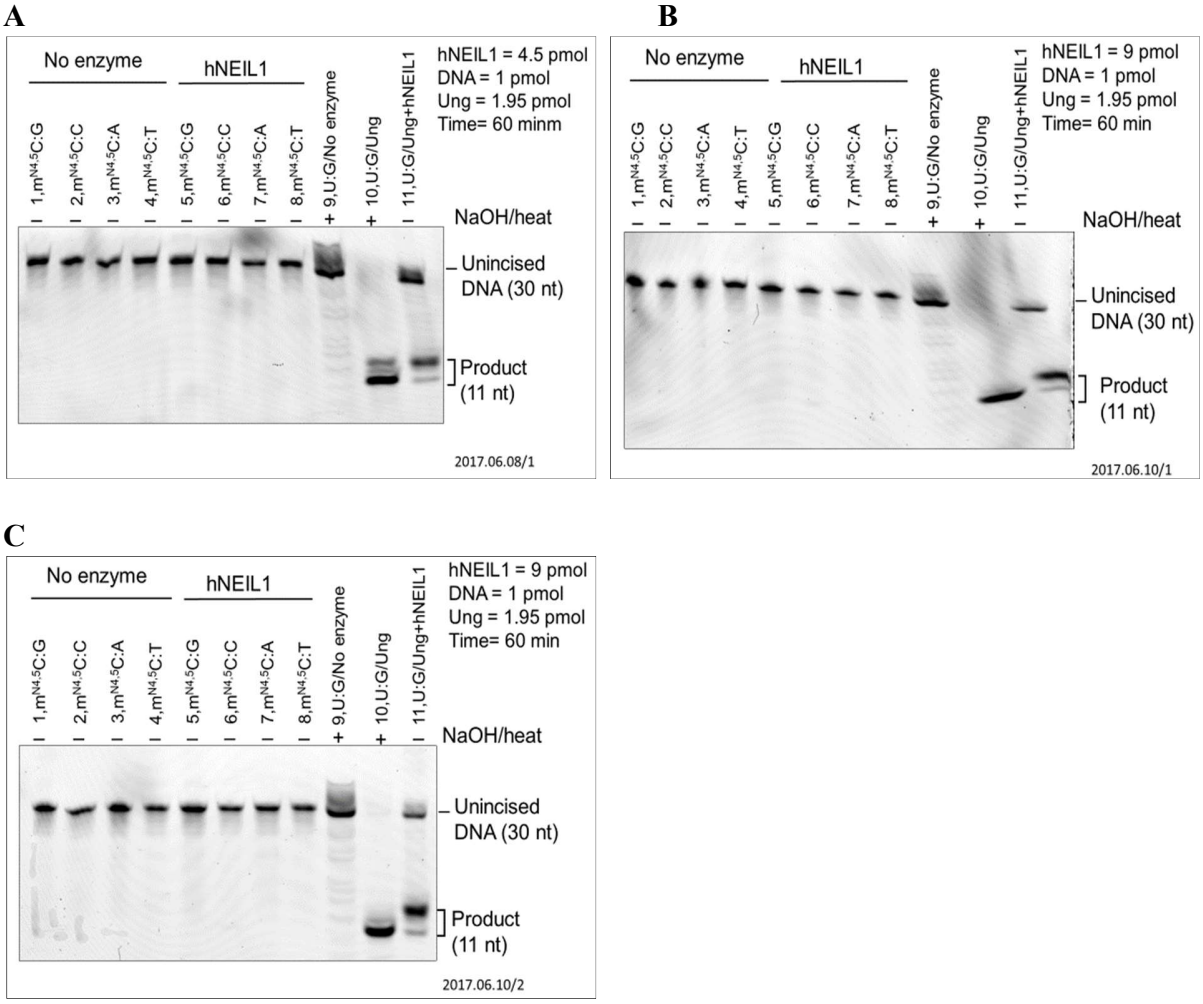


Figure B13: Activity on m^{N4,5}C-DNA by hNEIL1.
hNEIL1 4.49 pmol (**A**) and 9 pmol (**B, C**) was incubated with DNA substrate (1 pmol) in buffer (10 mM Tris-HCl, pH 8.0, 75 mM NaCl, and 1 mM EDTA) at 37°C for 1 h (final volume, 20 µl) to observe m^{N4,5}C-DNA glycosylase activity opposite G, C, A, and T. U:G-DNA (1 pmol) incubated without (lane 10) and with Ung, the latter followed by NaOH/heat treatment (lane 10), was used as negative and positive control for active Ung, respectively, which was used to convert U:G-DNA into AP-DNA to demonstrate active hNEIL1 (*i.e.*, lyase activity; lane 11). Incised was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h. Gels were scanned at PMT 550 V.

hNEIL2 experiments

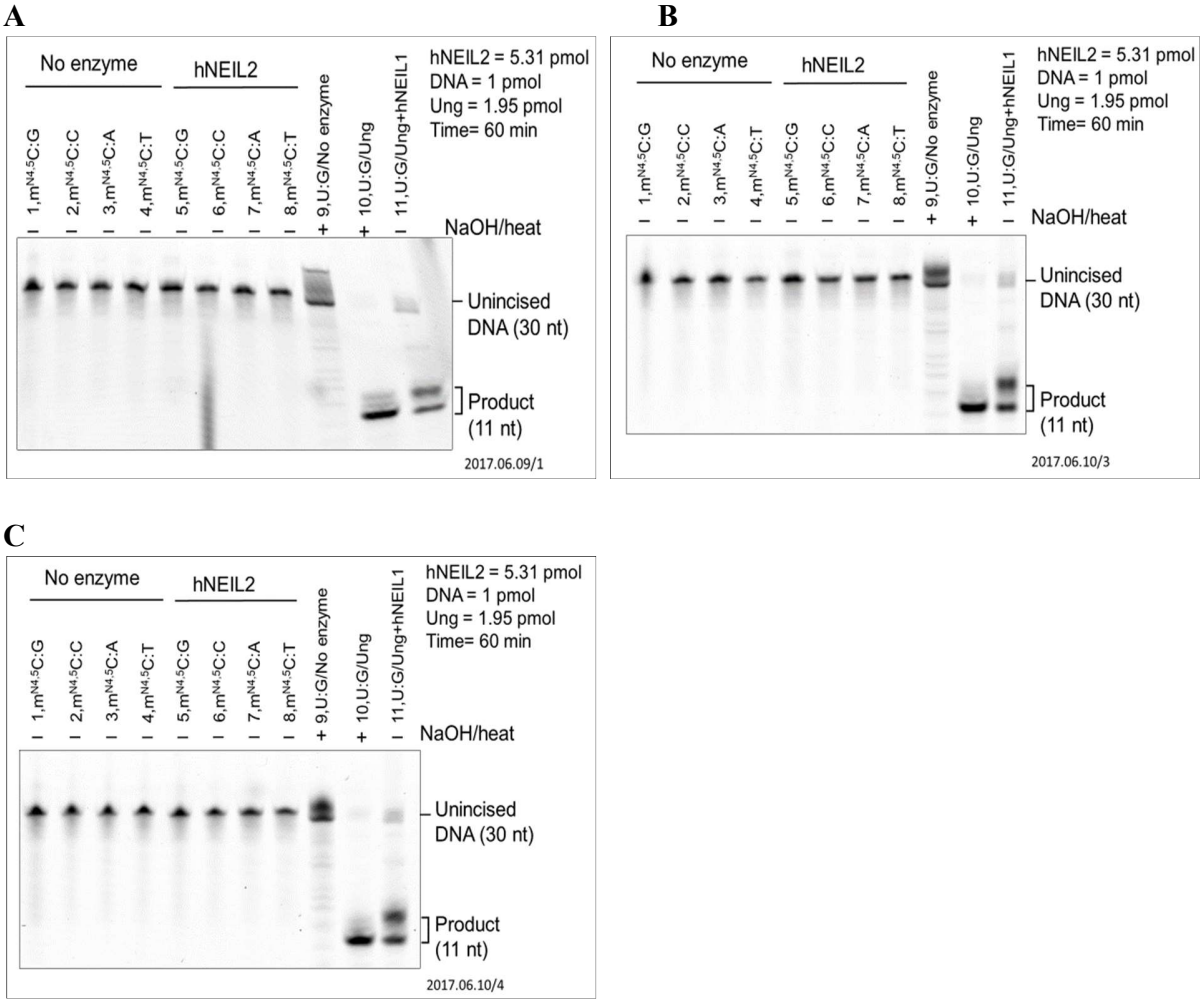


Figure B14: Activity on m^{N4,5}C-DNA by hNEIL2.
hNEIL2 (5.31 pmol) was incubated with DNA substrate (1 pmol) in buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, and 1 mM EDTA) and 0.1 mg/ml BSA at 37°C for 1 h (final volume, 20 µl) to observe m^{N4,5}C-DNA glycosylase activity opposite G, C, A, and T. U:G-DNA (1 pmol) incubated without (lane 10) and with Ung, the latter followed by NaOH/heat treatment (lane 10), was used as negative and positive control for active Ung, respectively, which was used to convert U:G-DNA into AP-DNA to demonstrate active hNEIL2 (*i.e.*, lyase activity; lane 11). Incised was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h. Gels were scanned at PMT 550 V.

hNEIL3 experiments

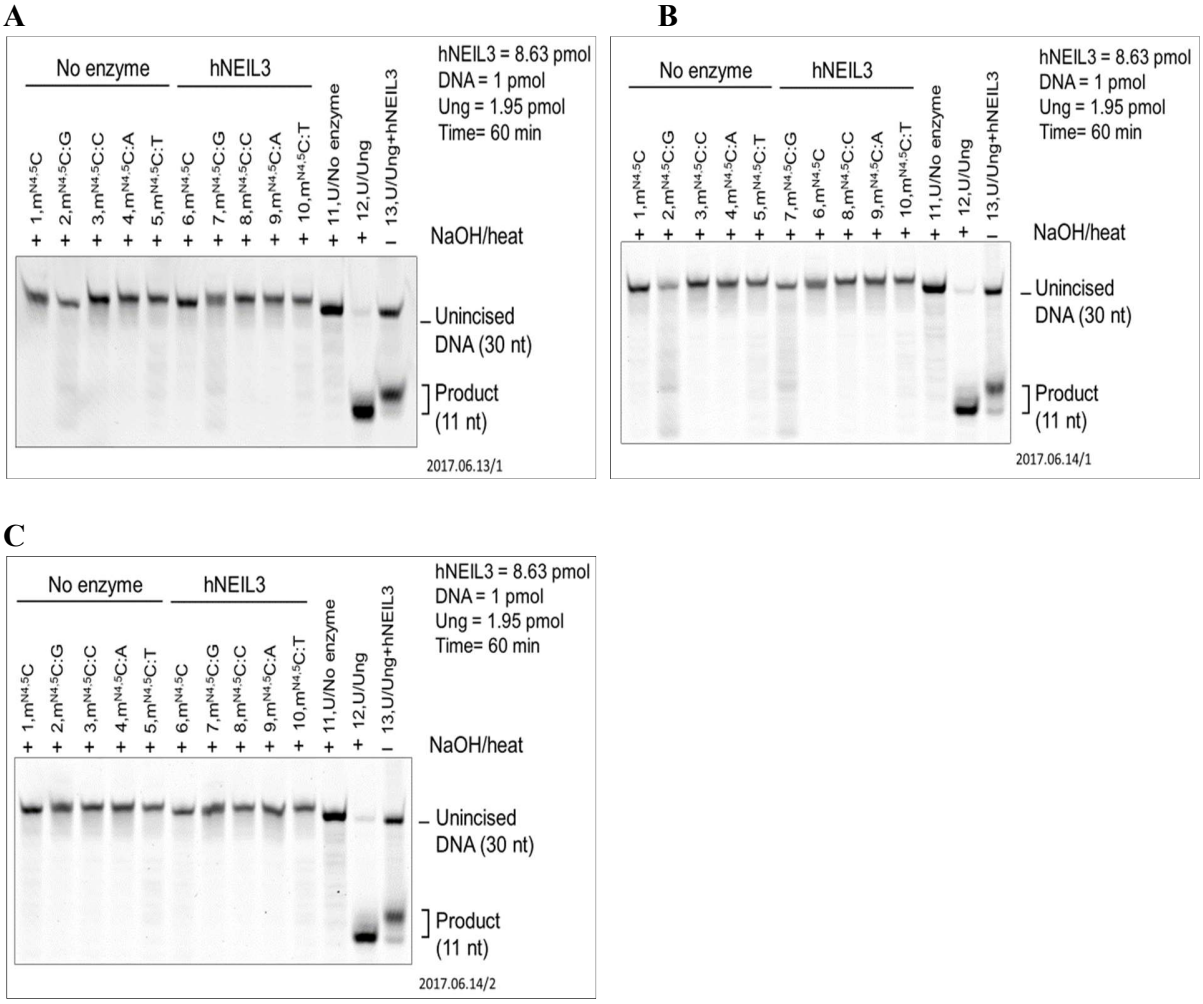


Figure B15: Activity on m^{N4,5}C-DNA by hNEIL3.
hNEIL3 (8.63 pmol) was incubated with DNA substrate (1 pmol) in buffer (50 mM MOPS, pH 7.5, 5% glycerol, 1 mM DTT, and 1 mM EDTA) at 37°C for 1 h (final volume, 20 µl) to observe m^{N4,5}C-DNA glycosylase activity opposite G, C, A, and T. U-DNA (1 pmol) incubated without (lane 11) and with Ung, the latter followed by NaOH/heat treatment (lane 12), was used as negative and positive control for active Ung, respectively, which was used to convert U:G-DNA into AP-DNA to demonstrate active hNEIL1 (*i.e.*, lyase activity; lane 13). Incised was separated from un-incised DNA by denaturing PAGE at 200 V for 2 h. Gels were scanned at PMT 550 V.

Calibration curve of PMT detector

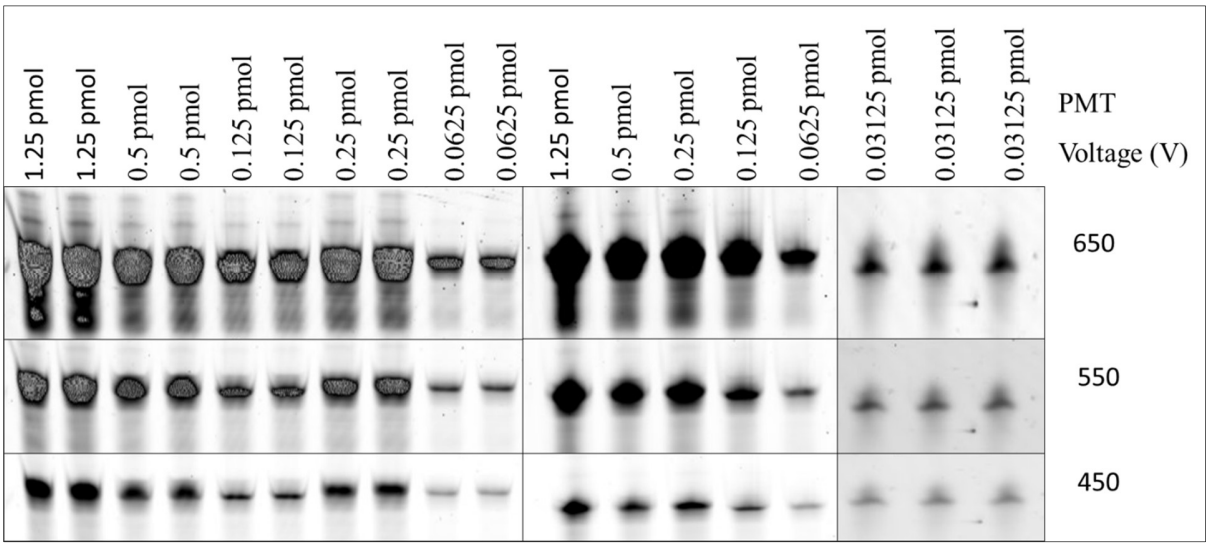


Figure B16: Calibration curve preparation
$m^{N4,5}C$-DNA at various amount ranging from 0.03125 pmol–1.25 pmol were electrophoresed on denaturing PAGE and quantified for the preparation of calibration curve at various PMT settings.

Optimization of DNA separation and denaturation

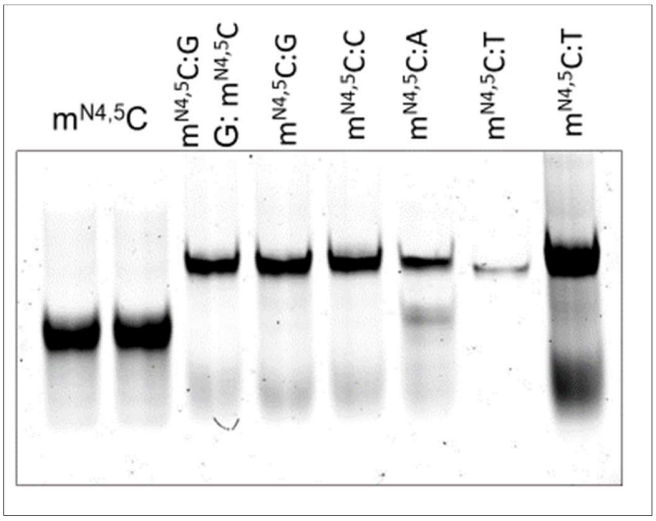


Figure B17: DNA hybridization studied on native-PAGE.
The oligonucleotides with $m^{N4,5}C$ lesion, single-stranded (lane 1 and 2) and duplex opposite different four bases were separated on 20% native-PAGE at 150 V for 1 h.

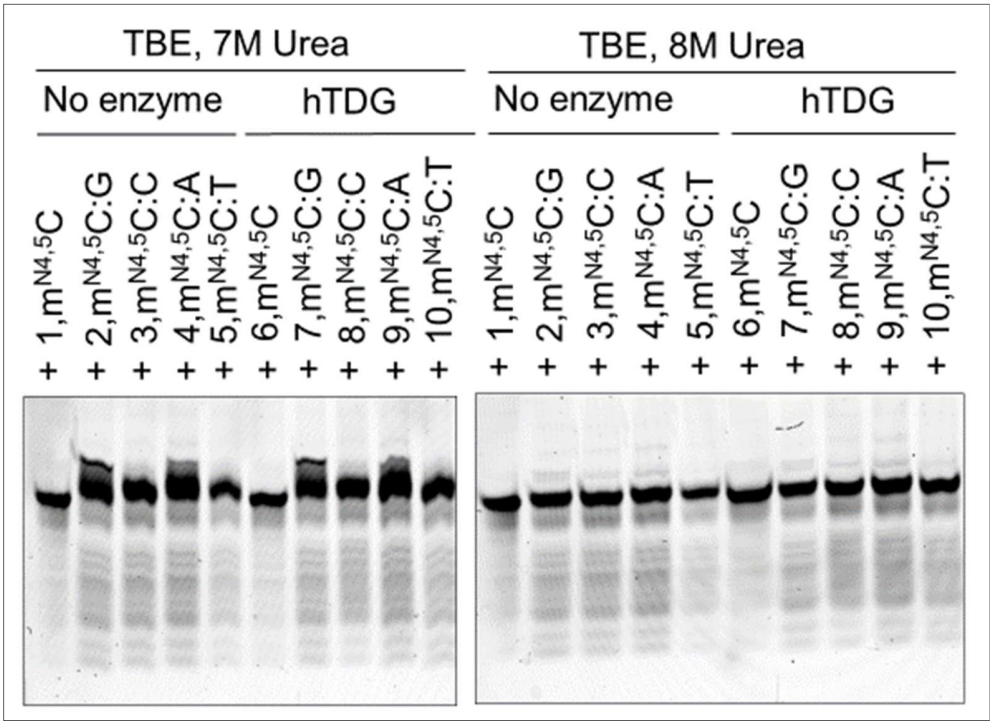


Figure B18: Comparison of DNA migration under various denaturing conditions.
The oligonucleotides with $m^{N4,5}C$ lesion, single-stranded (lane 1 and 6) and duplex opposite different four bases were migrated on denaturing-PAGE with urea 7 M (left) and 8 M (right) with Tris-Borax-EDTA (1×) buffer system. The samples from same reaction were loaded on two different gels with varying composition of urea and at higher concentration of urea (8 M) the denaturation is observed to be effective.

Appendix C

Table C1: Data of Fpg enzyme kinetics on $m^{N4,5}C$:C (measured after 10 min incubations).

Concentration			Velocity (min^{-1})			
pmol/20 µl	**nmol/20 µl**	**nM**	**Cleavage (%)**	**Cleavage (pmol/µl)**	**Cleavage (pM)**	**Cleavage (nM)**
S1						
0.125	0.000125	6.25	0.426	0.000027	26.68	0.026
0.25	0.00025	12.5	0.421	0.000053	52.63	0.052
0.5	0.0005	25	0.694	0.000174	173.54	0.173
1	0.001	50	0.528	0.000264	264.19	0.264
2	0.002	100	0.717	0.000718	717.67	0.717
4	0.004	200	0.501	0.001002	1002.20	1.00
8	0.008	400	0.291	0.001165	1165.21	1.16
10	0.01	500	0.295	0.001477	1477.27	1.47
S2						
0.125	0.000125	6.25	0.476	0.000030	29.75	0.029
0.25	0.00025	12.5	0.528	0.000066	66.03	0.06
0.5	0.0005	25	0.824	0.000206	206.02	0.20
1	0.001	50	0.714	0.000357	357.41	0.35
2	0.002	100	0.764	0.000764	764.10	0.76
4	0.004	200	0.579	0.001159	1159.42	1.15
8	0.008	400	0.333	0.001333	1332.83	1.33
10	0.01	500	0.3722	0.001861	1861.33	1.86
S3						
0.125	0.00012	6.25	0.722	0.000045	45.13	0.04
0.25	0.00025	12.5	0.782	0.000098	97.87	0.09
0.5	0.0005	25	1.10	0.000275	275.39	0.27
1	0.001	50	0.743	0.000372	371.83	0.37
2	0.002	100	0.698	0.000699	698.55	0.69
4	0.004	200	0.503	0.001007	1007.13	1.00
8	0.008	400	0.340	0.001361	1360.77	1.36
10	0.01	500	0.211	0.001059	1058.66	1.05

Note: Data obtained from figure B4

Table C3: Data of time curve for Fpg

Time (min)	Incision (%)			Excision (%)		
	Exp 1	Exp 2	Exp 3	Exp 1	Exp 2	Exp 3
0	0.00	0.00	0.00	0.00	0.00	0.00
2	0.93	4.49	3.71	4.74	3.83	8.01
5	10.52	6.65	4.31	8.04	8.73	16.81
10	12.60	10.52	7.85	15.15	16.09	17.17
30	28.22	27.04	17.56	29.04	31.04	33.33
60	39.32	36.10	33.00	43.63	44.12	45.50

Note: Data obtained figure B2

Table C2: Data of Protein curve for Fpg at 60 min

Fpg Concentration (pmol)	Incision (%)							
	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6	Exp 7	Exp 8
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.5	0.64	8.65	8.65	1.95	1.66	-	-	-
1	2.81	4.70	2.89	6.44	17.05	9.56	2.72	3.06
2	6.73	14.15	4.10	4.66	8.10	-	-	-
5	28.92	26.42	30.81	32.18	-	-	-	-
10	26.96	32.53	32.87	-	-	-	-	-
17	39.32	36.10	33.00	-	-	-	-	-
	Excision (%)							
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.5	2.40	0.21	1.98	1.26	2.59	2.33	-	-
1	6.14	3.18	4.19	4.87	5.22	6.96	-	-
2	18.82	11.02	7.84	6.31	4.89	7.84	8.01	8.59
5	32.37	23.62	31.00	-	-	-	-	-
10	30.93	33.88	34.95	-	-	-	-	-
17	43.63	44.12	45.50	-	-	-	-	-

Note: Data obtained from figure B3

Table C4: Data of Nei enzyme kinetics on $m^{N4,5}C$:G (measured after 10 min incubations).

Concentration			Velocity (min^{-1})			
pmol/20 µl	nmol/20 µl	nM	Cleavage (%)	Cleavage (pmol/µl)	Cleavage (pM)	Cleavage (nM)
S1						
0.125	0.000125	6.25	0.00	0.00	0.00	0.00
0.25	0.00025	12.5	0.148	0.00002	18.54	0.019
0.5	0.0005	25	0.377	0.00009	94.36	0.094
1	0.001	50	0.903	0.00045	451.29	0.451
2	0.002	100	1.027	0.00103	1026.73	1.03
4	0.004	200	0.815	0.00163	1630.35	1.63
8	0.008	400	0.624	0.00250	2495.92	2.50
10	0.01	500	0.530	0.00265	2650.15	2.65
S2						
0.125	0.000125	6.25	0.00	0.00	0.00	0.00
0.25	0.00025	12.5	0.3008	0.0000	37.60	0.0376
0.5	0.0005	25	0.3049	0.0001	76.23	0.0762
1	0.001	50	0.7688	0.0004	384.38	0.3844
2	0.002	100	0.9357	0.0009	935.67	0.9357
4	0.004	200	0.7008	0.0014	1401.53	1.40
8	0.008	400	0.4215	0.0017	1686.17	1.69
10	0.01	500	0.3436	0.0017	1718.24	1.72
S3						
0.125	0.00012	6.25	0.00	0.00	0.00	0.00
0.25	0.00025	12.5	0.3164	0.0000	39.55	0.0396
0.5	0.0005	25	0.3448	0.0001	86.21	0.0862
1	0.001	50	0.7592	0.0004	379.62	0.3796
2	0.002	100	0.9778	0.0010	977.81	0.9778
4	0.004	200	0.7333	0.0015	1466.67	1.47
8	0.008	400	0.4892	0.0020	1956.96	1.96
10	0.01	500	0.4316	0.0022	2157.79	2.16

Note: Data obtained from figure B8

Table C5: Data of Nei Protein curve.

Nei Concentration (pmol)	Incision (%)		
	Experiment 1	Experiment 2	Experiment 3
0	0.00	0.00	0.00
1.04	2.17	4.88	2.68
2.08	3.56	6.13	2.57
4.16	10.73	9.39	7.81
6.24	11.68	10.69	10.00
8.32	20.52	20.25	15.68
10.4	8.44	19.59	13.45
12.48	7.19	19.94	8.42

Note: Data obtained from figure B6

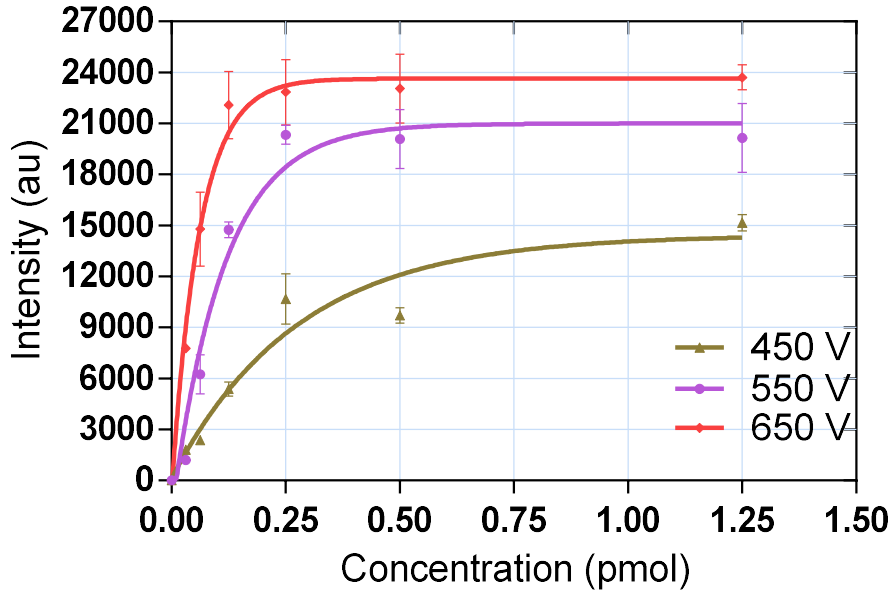
Table C5: Data of Nei Time curve.

Time (min)	Incision (%)		
	Experiment 1	Experiment 2	Experiment 3
0	0.00	0.00	0.00
2	0.96	0.82	1.40
5	2.22	1.87	3.38
15	5.34	5.51	4.75
30	8.37	7.04	7.24
45	9.04	8.71	10.13
60	14.57	10.01	10.51

Note: Data obtained from figure B7

Calibration curve for substrate DNA

A



B

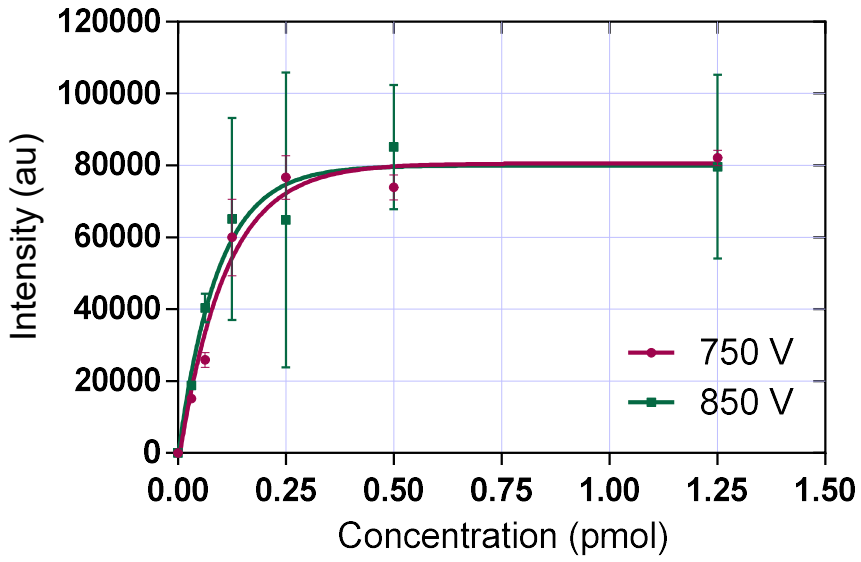


Figure C1: Calibration curve for the substrate-DNA.
Intensities measured at different PMT voltage are plotted for three parallels of the serially diluted substrate DNA **(A)** PMT settings at 450–650V; **(B)** PMT settings at 750 and 850V. The DNA migrated on denaturing PAGE and quantified by fluorescence imaging (Figure B16).

Comparison of the data obtained with and without dilution

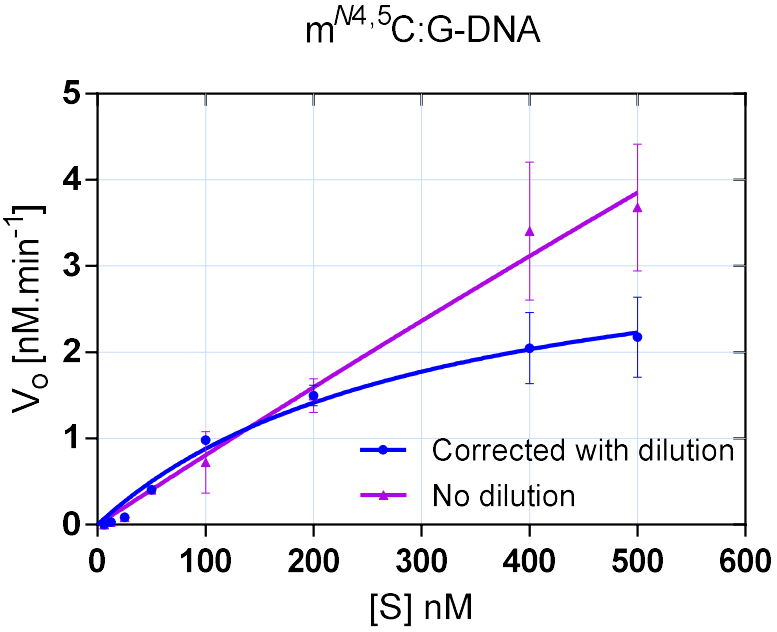


Figure C2: Excision of $m^{N4,5}C$:G-DNA by Nei as a function of substrate concentration.
Nei enzyme 416 nM was incubated with 6.25–500 nM of $m^{N4,5}C$:G-DNA over 10 min. The initial velocity is plotted as a function of substrate concentration and the curve was fitted for nonlinear least-square analysis using GraphPad Prism. The samples were electrophoresed on PAGE with and without the prior dilution. Due to signal saturation at higher concentration, undiluted DNA samples (purple) are not following the Michaelis-Menten kinetics due to high error. Abbreviations: V_o, initial velocity; [S], substrate concentrations.

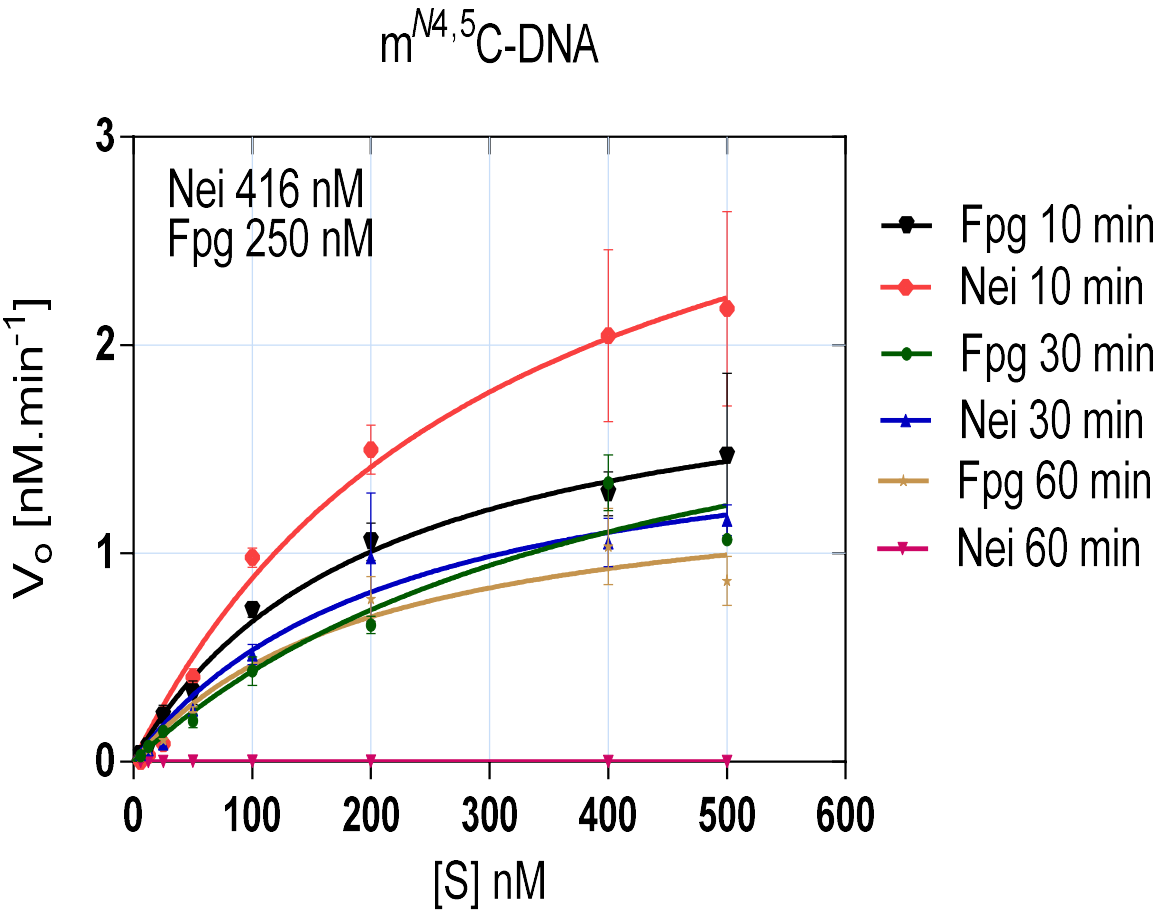


Figure C3: Excision and incision of m^{N4,5}C-DNA by Nei and Fpg as a function of substrate concentration.

Nei protein (416 nM) and Fpg (250 nM) was incubated with 6.25–500 nM of m^{N4,5}C:G and m^{N4,5}C:C respectively, over 10 min, 30 and 60 min. The velocity plotted as a function of substrate concentration and the curve was fitted for nonlinear least-square analysis. At 10 min, when the velocity was high and linear it indicated Michaelis-Menten kinetics. Due to saturation of enzyme at higher time, the velocity slowed down with the increase in the time. Abbreviations: V_o, initial velocity; [S], substrate concentrations.

Table C7: Units

˚C	Degrees Celsius
g	Grams
mg	Milligram (10^{-3} g)
µg	Microgram (10^{-6} g)
ng	Nanogram (10^{-9} g)
L	Liter
ml	Milliliter (10^{-3} L)
µl	Microliter (10^{-6} L)
M	Molar
mM	Millimolar (10^{-3} M)
µM	Micromolar (10^{-6} M)
nM	Nanomolar (10^{-9}M)
nmol	Nanomole (10^{-9} mole)
pmol	Picomole (10^{-12} mole)
fmol	Femtomole (10^{-15} mole)
nm Nanometer (wavelength of light)	Nanometer (wavelength of light)
kDa	Kilo Daltons (10^{3} Daltons)
RPM	Revolutions per minute
U	Units
V	Volts
MW	Molecular weight