Charles University in Prague

Faculty of Science

Department of Physical Chemistry

Ph.D. Study Program: Physical Chemistry

Structural studies of an abasic site DNA damage repair and DNA interstrand cross-link formation

Strukturní studie opravy abazického místa a tvorba kovalentního spojení opačných řetězců DNA

Mgr. Barbora Landová

Dissertation Thesis

Scientific supervisor: Mrg. et Mgr. Evžen Bouřa, PhD.

Prague, 2023

This dissertation describes only my work except where acknowledgement is made in the text. It is not substantially the same as any work that has been, or is being, submitted to any other university for any degree, diploma, or any other qualification.

Prague date ….……

Mgr. Barbora Landova

Acknowledgement

I would like to express my gratitude to Dr. Bouřa, my supervisor, for his guidance, support, and insightful feedback throughout the course of my PhD thesis. His expertise and encouragement have been instrumental in shaping the direction of my research.

I am also deeply thankful to my family for their encouragement and support during my academic journey. Your belief in me has been a constant source of motivation.

I extend my appreciation to my colleagues for their collaborative spirit, stimulating discussions, and shared insights that have enriched my understanding of the subject matter.

I would like to acknowledge Dr. Šilhán, my formal supervisor and consultant, for providing valuable advice, thoughtful suggestions and a wealth of knowledge that significantly contributed to the quality of my work.

Lastly, I would like to thank all those who have played a part, no matter how small, in helping me reach this milestone.

This research was funded by the project the National Institute Virology and Bacteriology (Programme EXCELES, Project No. LX22NPO5103) - Funded by the European Union - Next Generation EU.

Table of Contents

1 Abstract

DNA damage refers to any alteration or modification in the DNA structure that deviates from its natural state. Abasic site (Ap site) is one of the most common DNA lesions resulting from spontaneous depurination/depyrimidination or enzymatic base excision. When left unrepaired it can lead to a cascade of genetic mutations, potentially causing diseases like cancer. Understanding DNA repair mechanisms is vital for medical research and applications.

Bacterial MutM is a DNA repair glycosylase, removing DNA damage generated by oxidative stress and preventing mutations and genomic instability. MutM belongs to the Fpg/Nei family of procaryotic enzymes, sharing structural and functional similarities with their eukaryotic counterparts, such as NEIL1-NEIL3. Here, I present two crystal structures of MutM from pathogenic *Neisseria meningitidis*: MutM holoenzyme and MutM bound to DNA. The free enzyme exists in an open conformation, while upon binding to DNA, both the enzyme and DNA undergo substantial structural changes and domain rearrangement.

One of the DNA lesion repaired by MutM is the Ap site, which, if not repaired, may spontaneously lead to the formation of an abasic site interstrand crosslink (Ap-ICL) with an adjacent adenine in the opposite strand. NEIL3 glycosylase is known to remove Ap-ICL. With a panel of different oligonucleotides, we investigated the rates of formation, the yields, and the stability of Ap-ICL. Our findings demonstrate how different bases in the vicinity of the AP site change crosslink formation *in vitro*. Based on our experimental data on Ap-ICL formation and known biochemistry of the Ap site we have estimated the number of Ap-ICLs within the cell.

2 Abstrakt

Poškozením DNA se rozumí jakákoli změna nebo modifikace struktury DNA, která se odchyluje od jejího přirozeného stavu. Abasické místo (Ap místo) je jedním z nejčastějších poškození DNA, které vzniká spontánní depurinací/depyrimidinací nebo enzymatickým odstraněním báze. Pokud se neopraví, může vést ke genetické mutaci a potenciálně způsobit onemocnění, jako je například rakovina. Pochopení mechanismu opravy DNA je zásadní pro lékařský výzkum a aplikaci.

Bakteriální MutM je glykosyláza opravující DNA, která odstraňuje poškození DNA vzniklé oxidačním stresem a zabraňuje mutacím a genomové nestabilitě. MutM patří do rodiny prokaryotických enzymů Fpg/Nei a je strukturně i funkčně podobná se svým eukaryotickým protějškům, jako jsou NEIL1-NEIL3. Zde prezentuji dvě krystalové struktury MutM z patogenní *Neisseria meningitidis*: MutM holoenzym a MutM vázaný na DNA. Volný enzym existuje v otevřené konformaci, zatímco po vazbě na DNA, dochází k podstatným strukturním změnám a přeskupení domén enzymu i ohybu DNA.

Jednou z poškození DNA opravovaných MutM je Ap místo, které, pokud není opraveno, může spontánně vést k vytvoření mezivláknového kovanetního prokřízení DNA (Ap-ICL) se sousedním adeninem na opačném vlákně DNA. Je známo, že glykosyláza NEIL3 odstraňuje Ap-ICL. Pomocí série různých oligonukleotidů jsme zkoumali rychlost tvorby, výtěžky a stabilitu Ap-ICL. Naše zjištění ukazují, jak různé báze v blízkosti Ap místa mění rychlost a výtěžek tvorby Ap-ICL *in vitro*. Na základě našich experimentálních údajů o tvorbě Ap-ICL a známé biochemii Ap místa jsme odhadli počet Ap-ICL v buňce.

3 Introduction

DNA undergoes various spontaneous chemical reactions that could lead to DNA damage and loss of genome integrity. Moreover, highly reactive compounds constantly attack susceptible groups of DNA molecule, resulting in a vast number of DNA lesions. Both endogenous and exogenous compounds, such as ionization or oxidative stress, are the most often responsible for changes in DNA molecule that can lead to DNA lesions. If not repaired, these lesions may result in genetic instability and cancer [1-3].

Although the sources may differ for both multicellular and single-cellular organisms, the main agents responsible for oxidized DNA lesion are thought to be hydroxyl radical, singlet oxygen and hydrogen peroxide [4]. Both purine and pyrimidine nucleotides are damaged, leading to more than 100 different DNA lesions. To prevent the harmful effects of accumulating DNA damage, various DNA repair pathways have developed, each specialized in repairing specific types of lesions [5].

Unlike proteins, lipids and RNA, DNA cannot be replaced when damaged, making DNA repair crucial for maintaining genomic integrity. Failure to repair DNA damage can lead to cell death through apoptosis or necrosis, preventing the transmission of mutations to daughter cells and development of diseases, including cancer. Interestingly, certain cancer therapies use DNA-damaging agents to trigger apoptosis in tumor cells.

3.1 Abasic site (Ap site)

First line of defence against damaged DNA is the base excision repair pathway (BER), which is initiated by enzymatical or spontaneous cleavage of N-glycosidic bond of the damaged base, forming an abasic site (Ap - Apurinic/Apyrimidinic) as an intermediate before the DNA is fully repaired. The Ap site (Figure 3.1.) is recognized by Ap endonuclease, resulting into a single-stranded break. Furthermore, it is filled by DNA polymerase and ligated by DNA ligase. Eukaryotic DNA polymerase occurs

with an error rate of 1 error per 10^7 synthetised bases. It is also able to control itself and repair its mistakes [6-8].

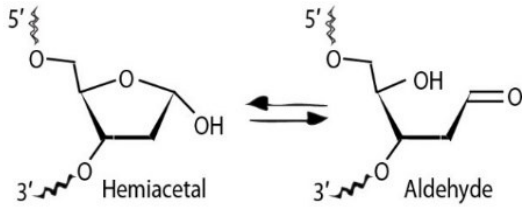

Figure 3.1. Structure of two equilibrating forms of abasic site (Ap site) – closed form of hemiacetal and open form of aldehyde.

Table 3.1. DNA lesions summary [9] NER stands for nucleotide excision repair, HR for homology recombination and BER for base excision repair.

Source of the lesion	Type of lesion	DNA repair pathway
IR light, radicals, alkylating agents	Altered base, Ap site, single strand break	BER
UV light, polyaromatic hydrocarbons, cisplatin	ICL, bulky DNA adduct	NER
IR light, cisplatin, alkylating agents	ICL, double strand break	HR recombination
Replication errors	Insertion, deletion, mismatched base	Mismatched repair

Ap sites can also be generated spontaneously in significant quantities, with approximately 10,000 Ap sites formed each day in a single human cell [10,11]. Several thousand of Ap sites are present in mammalian cells at steady state. Uracil DNA glycosylase is a well-known enzyme involved in base excision repair and somatic hypermutation during antibody diversification is yet another source of Ap sites [12-15]. Its primary function is to eliminate uracil residues, which can result from the spontaneous or enzymatic deamination of cytosine [16].

The Ap site, besides spontaneous formation of DNA interstrand crosslink (ICL) it can undergo either spontaneous or enzymatic cleavage. Specifically, Ap endonucleases cleave the phosphodiester bond at the Ap site, generating a nick with a 3′-deoxyribose-5′-phosphate and a single-strand break [17]. Alternatively, the lesion can spontaneously break with a half-life of 8 days. The reaction rate is influenced by amines and pH increase due to the nature of reaction, which involves base hydrolysis catalysis.

Adenine and guanine are liberated from DNA at similar rates. In comparison, the release of thymine and cytosine is only 5% of the rate of purines, and the most common pyrimidine lesion are thymine glycol and cytosine glycol, which can lead to 5-hydroxycytosine or undergo rapid deamination to uracil glycol, resulting in 5-hydroxyuracil formation. In dsDNA. the velocity of depurination is only four times higher than in ssDNA [10,18].

3.2 Fpg/Nei family

One of the most common oxidative lesions is 8-oxoguanine (8-oxoG), which can cause G/C to T/A transversion after DNA replication. 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG) results from 8-oxoG, and similar fate awaits oxidized adenine, leading to 4,6-diamino-5-formamidopyrimidine (FapyA) [19,20]. Therefore, Fapy lesions has been shown to block DNA replication [21]. Fapy lesions are potentially lethal or, at least, mutagenic, as these lesions can spontaneously depurinate, leaving Ap site that could lead to GC – TA transversions [22]. One of the key enzymes repairing all the lesions mentioned above is MutM DNA glycosylase, belonging to Fpg/Nei family. It is also called Formamidopyrimidine DNA glycosylase (Fpg/MutM) due to its ability to repair Fapy lesions [23]. In this whole study, the enzyme is addressed as MutM.

Besides Fpg, Endonuclease VIII (Nei) is an important member of Fpg/Nei family responsible for repairing oxidised DNA bases. Despite high sequence homology *E.*

coli has a copy of each in its genome [24]. In mammalian cells, three Nei-like proteins have been characterized, NEIL1, NEIL2 and NEIL3 (Nei Endonuclease VIII-like 1-3).

DNA glycosylases are divided into two groups: Helix-Hairpin-Helix (HhH) represented by 8-oxo-guanine DNA glycosylase (Ogg1), and Fpg/Nei represented by Fpg (MutM) and Nei. Fpg/Nei family possesses two domains connected by a linker region. The active site is located in the clench of the two domains, and binding amino acids are on the very N-terminus, making the N-terminus conserved [18,19].

Two main families of lesions originate directly from it: 8-oxopurines and formamidopyrimidines. Both these lesions are repaired by the MutM protein [25]. The expression of MutM is correlated with the mutagenicity of gamma radiolysis, reflecting its biological relevance. These enzymes are conserved among procaryotes, and their homologues are often found in eucaryotic cells.

3.3 Formamidopyrimidine DNA glycosylase (MutM)

Fpg is bifunctional enzyme with its glycosylase activity that removes the base and its β,δ-lyase activity that processes the generated Ap site by cleaving phosphodiester backbone 3′ resulting in a break (Figure 3.2.) [26]. The first step of the repair mechanism is nucleophilic attack, where imine group of Pro2 attacks to C1′of the ribose of the damaged base via a covalent bond. A Schiff base with ribosyl residue in DNA is formed and Glu3 of MutM acts as a proton donor and stabilizes the intermediate, making it relatively stable. MutM is regenerated by β,δ-elimination of the ribose at the 3′ and 5′ ends, liberating a by-product 4-hydroxypenta-2,4-diene [27]. The reaction catalyzed by MutM leads to a DNA gap with another lesion, a toxic 3′-phosphate. This 3′-phosphate blocks the subsequent DNA repair polymerisation step and is commonly removed by the *Xth* family of Ap-endonucleases in bacteria. Pro2 is a part of conserved N-terminus catalytic motif PELPEV of MutM proteins, although the third and sixth amino acids may vary [28].

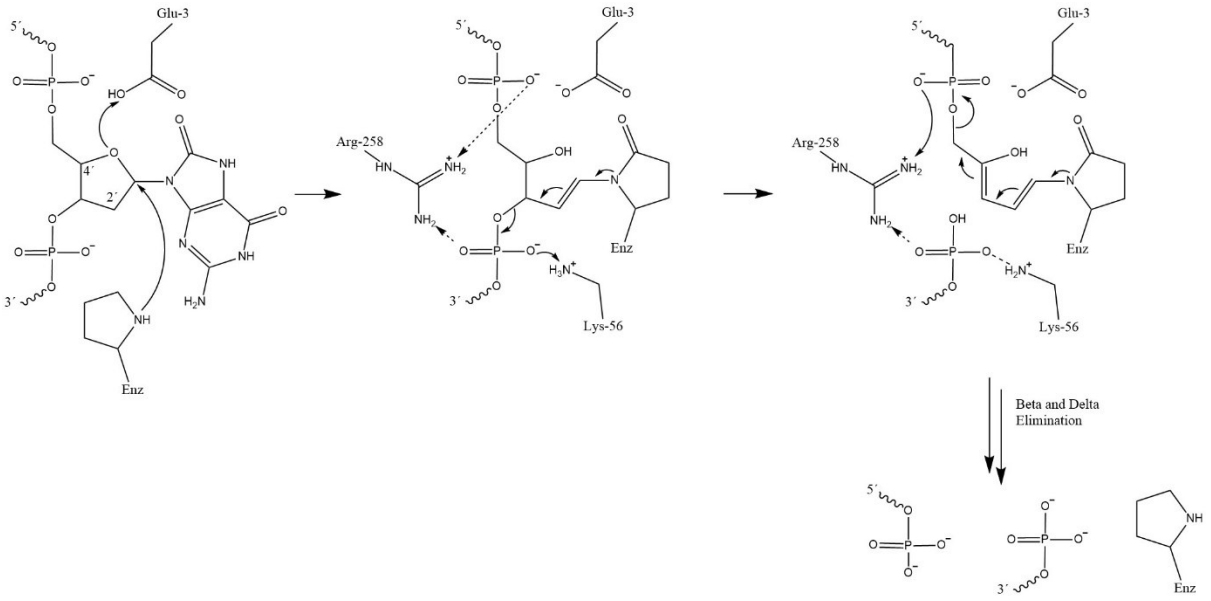

Figure 3.2. Schematics of the DNA repair mechanism of 8-oxo-guanine by MutM. Firstly, second amino acid proline attacks C1′of the DNA lesion ribose. Amino acids Glu-3, Lys-56 and Arg-258 stabilises the intermediates during damaged base unhooking.

To date, a vast number of crystal structures of Fpg/Nei glycosylases have been published (Figure 7.5.). The very first described the architecture of the holoenzyme from *Bacillus stearothermophilus* (PDBID: 1EE8) [29]. Subsequently, DNA-bound MutM for the same organism was reported (PDBID: 1L1T, 1L1Z, 1L2B, 1L2C, 1L2D) [28,30]. Most of the subsequent studies focused on substrate search, recognition, preference, the importance of the orphaned base to the lesion, and the reaction mechanism [27,31-34].

MutM and other members of the Fpg/Nei family exhibit a similar domain organization and secondary structure (Figure 3.3.), characterized by two main domains separated by a flexible region that acts as a hinge. Previous studies have provided insights into the overall architecture of holoenzymes and the DNA-bound structures of MutM enzymes. Specifically, the N-terminal domain is composed of eight β-sandwiched sheets, with two adjacent α-helices flanked on each side. On the other hand, the C-terminal domain consists of four α-helices and two antiparallel β-strands, forming a Zn-finger motif that is characteristic of this protein family. Notably, the last two α-helices within the CTD feature a conserved helix-two-turn-helix (H2TH) motif [35-

37], which contributes to the structural and functional diversity of these enzymes (Figure 3.4.).

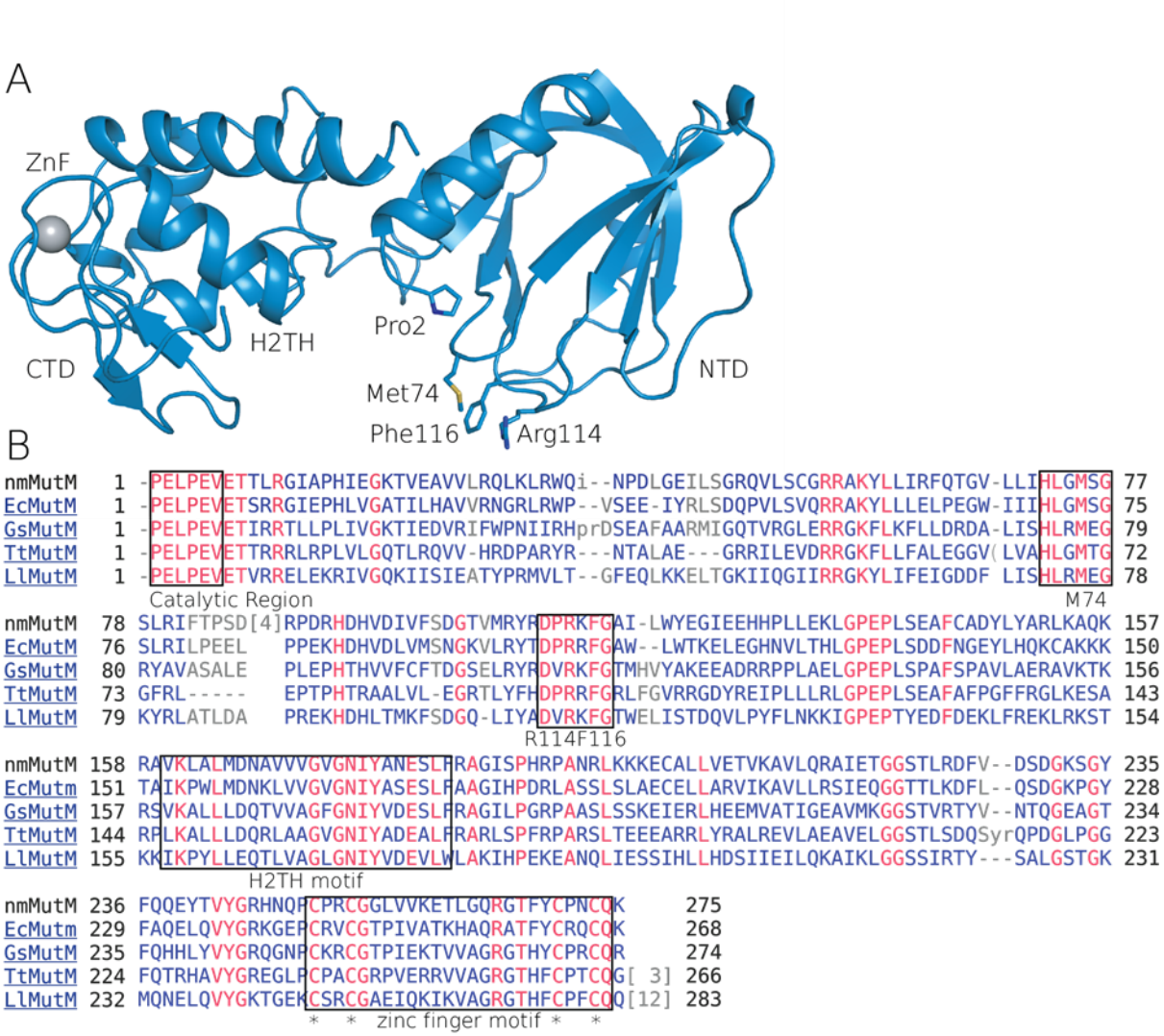

Figure 3.3. Structure of meningococcal MutM (PDBID:6TC6). (A) Crystal structure of MutM from this study from Neisseria meningitidis. Important structural motifs and amino acids involved in repair mechanism of this enzyme are labelled and named in the structure. (B) Sequence alignment of MutM proteins both from this study and previously published studies. Identical residues are labelled in red and the least conserved regions with the most deletions and loops are labelled in grey. The rest of the residues stay in blue.

The H2TH motif, along with other DNA-binding motifs such as helix-hairpin-helix (HhH) and helix-turn-helix, is commonly found in a wide range of DNA-binding proteins, including DNA glycosylases, transcription factors, and nucleases. In the case

of the Fpg/Nei family of proteins, the Zn-finger motif has been identified as crucial for DNA binding, although it is not directly in contact with the DNA. Interestingly, certain eukaryotic homologues of NEIL1 exhibit a distinct “Zincless” finger fold in this region of the C-terminus, suggesting variations in the DNA binding mechanisms between procaryotes and eukaryotes (Figure 3.4.) [26,38]. Enzymes from the Fpg/Nei family typically induce significant bend of approximately 45-60° in the substrate DNA upon binding [30,33,38]. This distortion in the DNA allows for efficient recognition and processing of damaged bases.

Understanding of the dynamic and conformational changes that occur in MutM during the DNA binding process is limited. Nevertheless, a comparative study of a MutM paralogue called EcoNei from *Excherichia coli* provided some insight. In this study, both the structures of free and DNA-bound EcoNei enzymes were examined (PDBID: 1Q3C, 1K3W). The findings revealed that EcoNei undergoes significant interdomain conformational changes upon DNA binding. Notably, both NTD and CTD undergo a substantial rotation towards the DNA molecule by an angle of approximately 50° [39,40].

A similar structural comparison was carried out with a more distant MutM homologue from *Arabidopsis* (PDBID:3TWL) that contains a “Zincless” finger domain instead of the canonical Zn-finger. Furthermore, this homologue, which lacks the ability to remove 8-oxoG, undergoes no significant structural change between DNA-bound and unbound form, in contrast to the *E. coli* EcoNei enzyme [41]. Additionally, the crystallographic studies of thermophilic MutM showed slight domain rotations along the flexible hinge region between monomers in the asymmetric unit (PDBID: 1EE8). Significantly, biophysical measurements conducted on MutM from *E. coli* have provided evidence supporting the dynamic nature of the enzyme, which undergoes a transition towards increased rigidity upon binding to DNA [42-44].

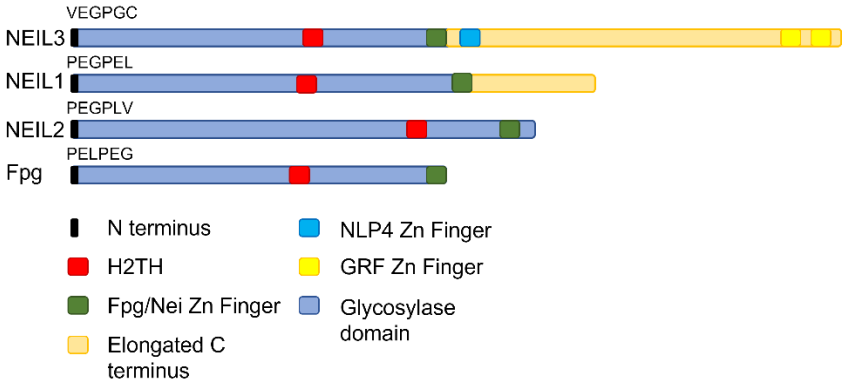

Figure 3.4. Schematic sequence representation of members of the Fpg/Nei family with marked important functional and structural motifs.

3.4 Endonuclease VIII-like 3 (NEIL3 glycosylase)

Earlier, NEIL3 was reported to exhibit limited Ap-lyase activity when repairing dsDNA substrates. However, subsequent investigations revealed that NEIL3 displays higher activity and specificity for ssDNA substrates [45]. Further research revealed that NEIL3 interacts with telomere sequences through its C-terminus, contributing to safeguarding telomere integrity during the S/G2 phase of cell proliferation [46].

Importantly, NEIL3 is a key participant in Ap-ICL repair. It has been shown that NEIL3 cleaves one of the N-glycosidic bonds within the Ap-ICL structure, resulting in the unhooking of this lesion. These experiments were carried out in *Xenopus laevis* egg extracts, and it represents the primary pathway for Ap-ICL repair [47-50]. However, the molecular details of the repair Ap-ICL and complete reaction mechanism of NEIL3 remains unknown.

The glycosylase domain of NEIL3 is catalytic and structurally similar to other members of the Fpg/Nei family. It has been shown that the presence of additional zinc fingers in the C-terminal domain could enable interaction with replication machinery, allowing the right DNA repair pathway to be initiated in case of DNA damage [47].

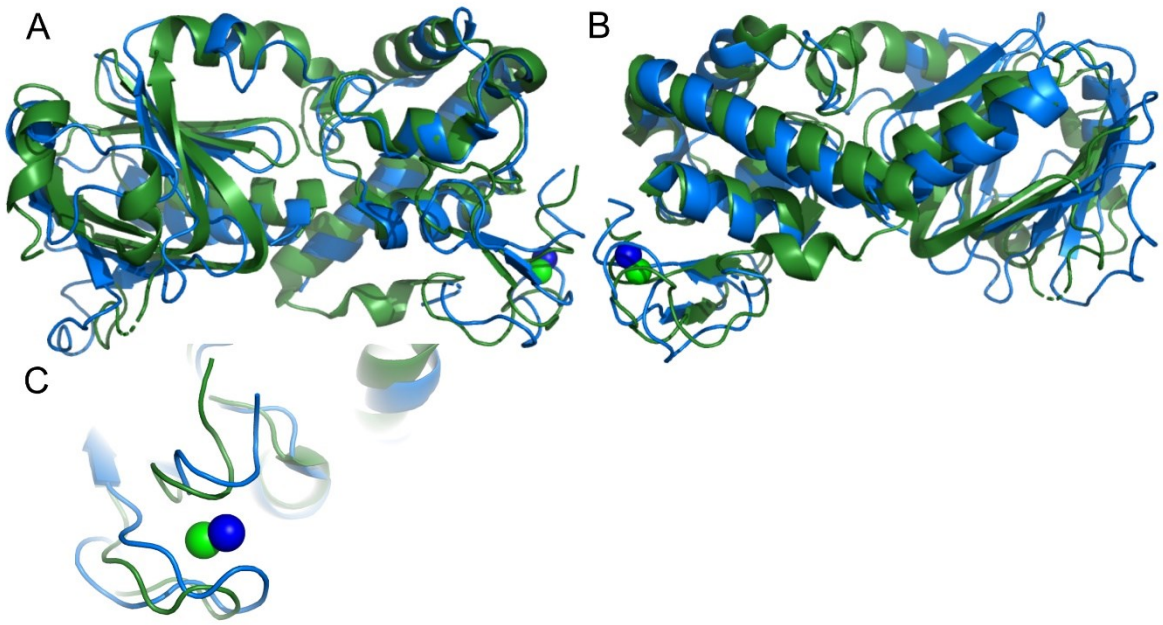

Figure 3.5. Structural alignment of MutM apoenzyme (blue) from this study (6TC6) and published structure (3W0F) of NEIL3 from Mus Musculus *(green) [51]. Layout of both MutM and NEIL3 NTD are similar as shown (A,B). (C) Focus on Fpg/Nei ZnF.*

It has been previously shown that short version containing only glycosylase domain (324 aa) of full length (606 aa) mNEIL3 maintain glycosylase activity. Moreover, it is significantly more stable and expression yields are higher [51]. Glycosylase domain is cut after the NTD typical for Fpg/Nei family which differ much in sequence but structurally is similar to Fpg (Figure 3.5.).

When NEIL3 recruited to a stalled replication fork, it likely targets bases positioned near the junction of a strand separated X-shaped or Y-shaped DNA configuration [48,52,53]. The ability of NEIL3 to catalytically remove psoralen and Ap-derived ICL can be explained. Comparing structures of these ICLs to previously characterized NEIL3 substrates – dihydrothymine (DTH), thymine glycol, FapyG and FapyA could provide insight to its activity (Figure 3.6.) [54,55]. The non-native glycosidic bond present in the Ap-dA ICL exhibits resemblance to the glycosidic bond found in FapyA and FapyG. Likewise, the saturated thymine rings in the psoralen ICL exhibit similarity to established substrates such as DTH and thymine glycol.

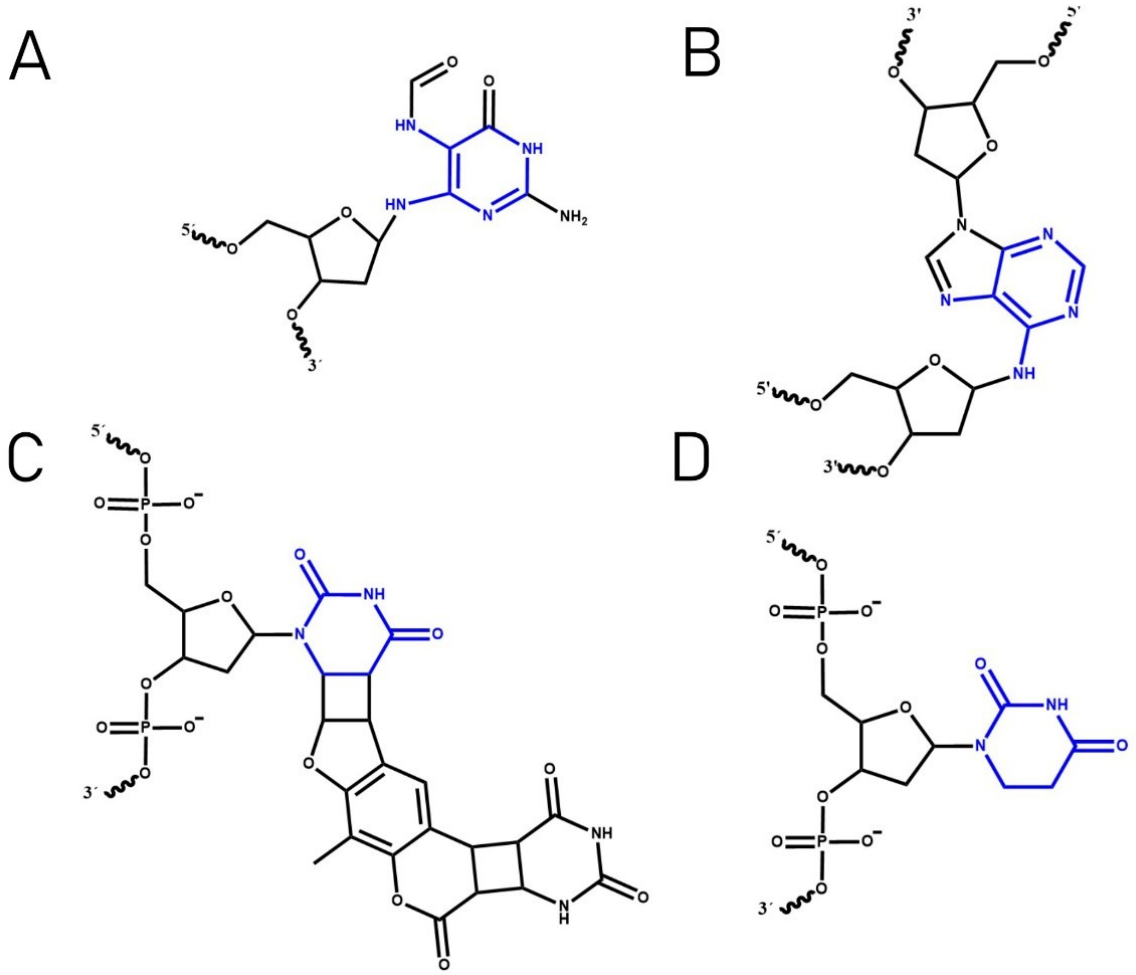

Figure 3.6. Psolaren-ICL (C) and Ap-ICL (B) are analogues to known NEIL3 substrates DTH (D) and Fapy-G (A). Identical parts of lesions are shown in blue.

3.5 DNA interstrand cross-link

An internal DNA interstrand cross-link (ICL) is a DNA lesion that covalently bonds both DNA strands, creating a barrier that prevents DNA replication [48]. Several forms of ICLs have been studied, differing both structurally and chemically [56-59]. The ICL is a toxic lesion originating from both endogenous and exogenous sources. It can also spontaneously form through a process where the exocyclic amino group on a base reacts with the aldehyde residue at the Ap site of adenine or guanine on the opposite DNA strand (Figure 3.7.). The aldehyde group of the ring-opened ribose at C1´forms an N-glycosidic bond with an amine group of a base on the opposite strand.

This reaction leads to the formation of carbinolamine, imine and aminoglycoside cross-links. The presence of the reducing agent $NaBH_3CN$ in the reaction mixture can significantly enhance the ICL yields and 3-aminopurine nucleobases [58,60]. Interstrand DNA cross-links derived from the reaction of a 2-aminopurine residue with an Ap site. ICLs block replication and transcription machineries by covalently linking the two DNA strands, necessitating removal by DNA repair pathway specialized at ICLs.

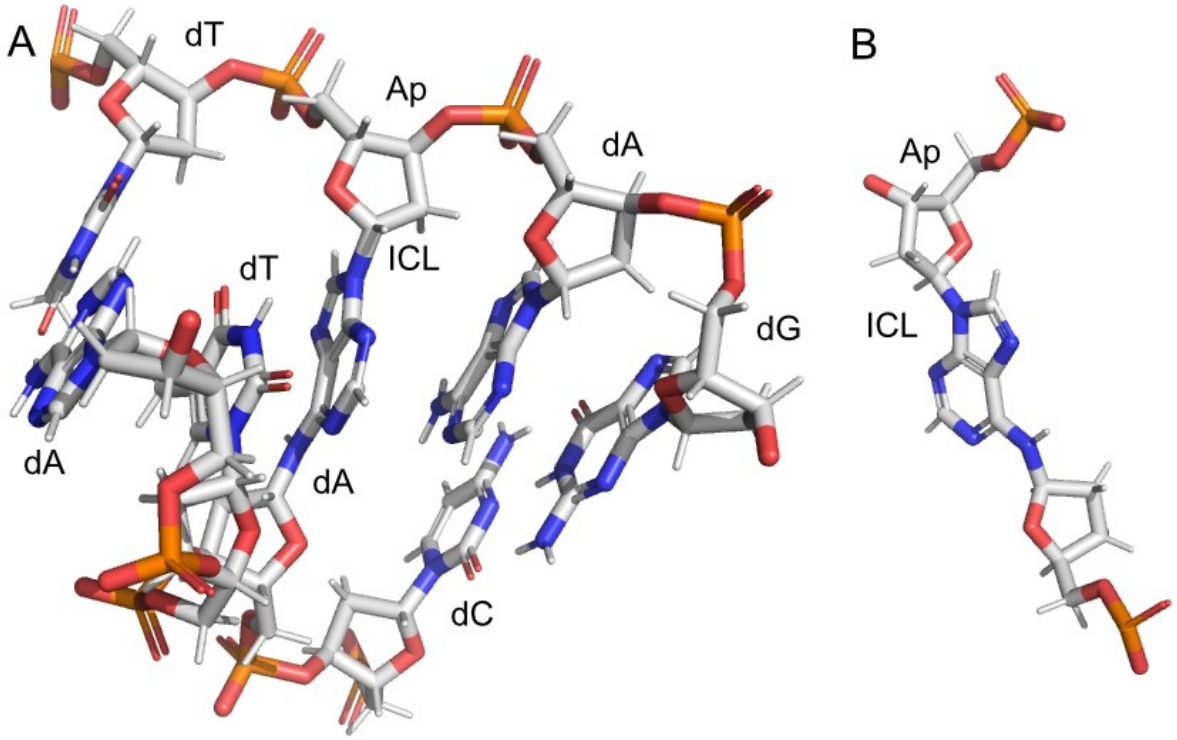

Figure 3.7. (A) Enlarged part from NMR structure of Ap-ICL with labelled nucleotides in the vicinity. N-glycosidic bond between Ap site and adenine from opposing strand is marked as ICL. (B) Top view of Ap-ICL from the same structure.

Three pathways are now known to be able to repair interstrand crosslink lesion: NEIL3 glycosylase pathway, acetaldehyde pathway and Fanconi anaemia (FA pathway) [47,49]. Initially, it was believed that FA pathway is responsible for repairing various types of ICLs. Nowadays, it is known to repair cisplatin-ICL and nitrogen mustard-ICL. If the NEIL3 pathway fails, FA proteins are responsible for double strand breaks, which are than repaired by homologous recombination (HR). Both pathways are linked to DNA replication machinery and complement each other

in repairing ICLs. The NEIL3 pathway takes charge in the early stages of replication and is coupled with replication fork, which stalls on the ICL. NEIL3 pathway is responsible for repairing Ap-ICL, psoralen-ICL, aldehyde-ICL, without generating double-strand breaks [61].

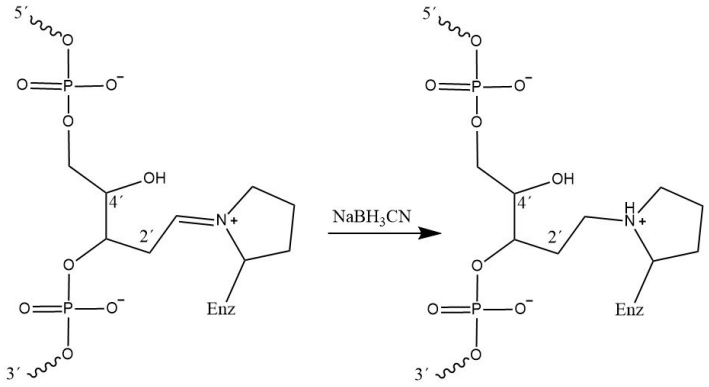

Figure 3.8. Schematic formation of stable crosslink (borohydride trapped complex) between Ap site and enzyme formed from Shiff base under reduction conditions created by addition of $NaBH_3CN$.

Initially, the Ap site undergoes a series of chemical reactions, starting with β-elimination followed by δ-elimination, resulting in the formation of a 3′-phosphate as the final product [62-65]. These 3′-lesions are highly toxic and can impede DNA ligation and replication processes. To address this, specific enzymes such as PNK or Ap exonucleases are responsible for repairing these lesions by recognizing and processing the 3′-phosphate [66-68]. Consequently, the resulting ssDNA break, which possesses a 3′-OH group, is repaired through the action of DNA polymerase and ligase enzymes [1,65,69]. Interestingly, the Ap site can also undergo nucleophilic attack, resulting in the formation of a DNA interstrand crosslink or a DNA-protein crosslink, both of which can have significant consequences for DNA structure and function [30].

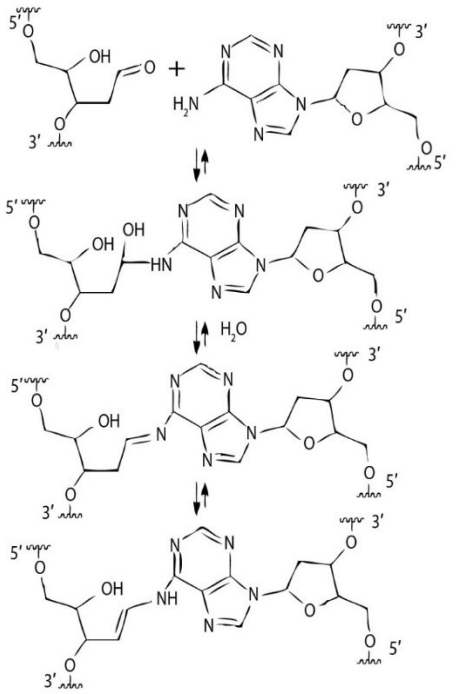

Figure 3.9. Schematic formation of Ap-ICL. Ap site was created in the lab using UDG, the rest of the reaction is spontaneous.

From a chemical perspective, an Ap site undergoes structural alterations, including mutarotation and spontaneous degradation [70]. The hydrolysis-induced spontaneous cleavage out of a base is caused by the attack of a water molecule on the C1 atom of the deoxyribose from the opposite side of the N-glycosidic bond which leads to the formation of alpha-hydroxyl at C1. The Ap site undergoes mutarotation until reaching the thermodynamic equilibrium. Notably, this equilibrium is established when only 1% of the ribose ring transforms into its aldehyde configuration [71,72]. Aldehyde form being reactive leads to formation of Shiff base when it interacts with amines [73-75]. Generating ICL from Ap site was widely investigated, mostly these experiments were carried under non-physiological low pH conditions around pH = 6.5. Estimated value of ICLs formed at pH = 6.8 in T7 phage DNA was 1 ICL per 140 Ap sites [76-79].

In vitro reconstruction of Ap-ICL at a specific site has been successfully achieved in relatively acidic conditions of pH = 5 and stabilized through reduction by NaBH3CN (Figure 3.8.). This Ap-ICL was formed between guanine and Ap site on the

complementary DNA strand within the 5′-dC-Ap sequence and characterized by mass spectroscopy [58]. Subsequently, Ap-ICL formed between Ap site linking with opposing adenine was characterized and confirmed to be stable under physiological conditions [80]. In this study we chose to create Ap-ICL using UDG, which cleaves out uracil and leaves Ap site and spontaneously Ap-ICL is formed (Figure 3.9.).

3.6 The NEIL3 pathway of ICL repair

The aforementioned NEIL3 pathway is responsible for Ap-ICL repair. ICLs pose a significant obstacle to DNA replication, potentially resulting in severe consequences. It is not surprising that the ICL repair process could be triggered by DNA replication forks [51]. During G1 phase of the cell cycle, cells undergo preparation for replication, characterized by the assembly of prereplication complexes consisting of minichromosome maintenance 2-7 (MGM2-7) double hexamers along vertebrate chromosomes. In the subsequent S phase, activation of cyclin-dependent kinase leads to the association of the cell division cycle 45 (CDC45) and go-ichi-ni-san (GINS) proteins with MCM2-7, forming the CDC45-MCM2-7-GINS (CMG) helicase complex. The CMG helicase translocates along the leading strand template, unwinding DNA at the forefront of the replisome [81].

All currently known mechanisms of replication coupled ICL repair (Figure 3.10.) are initiated when the TRAIP-dependent CMG helicase encounters the ICL resulting in a collision between the helicase and the lesion. After fork convergence at the lesion, CMG is activated by ubiquitylation and probably in the case of this ICL, it is placed to ssDNA which is NEIL3′s preferred substrate. If NEIL3 is not able to repair the ICL, for instance, in the case of cis-platin ICL, further ubiquitin chain growing leads to repair with FA pathway. NEIL3 cleaves one of the two N-glycosyl bonds which form cross-link, generating an Ap site and a monoadduct which is in the case of Ap-ICL normal adenosine in the other strand. The Ap site is bypassed by TLS, but generation of Ap itself promises the occurrence of a point mutation.

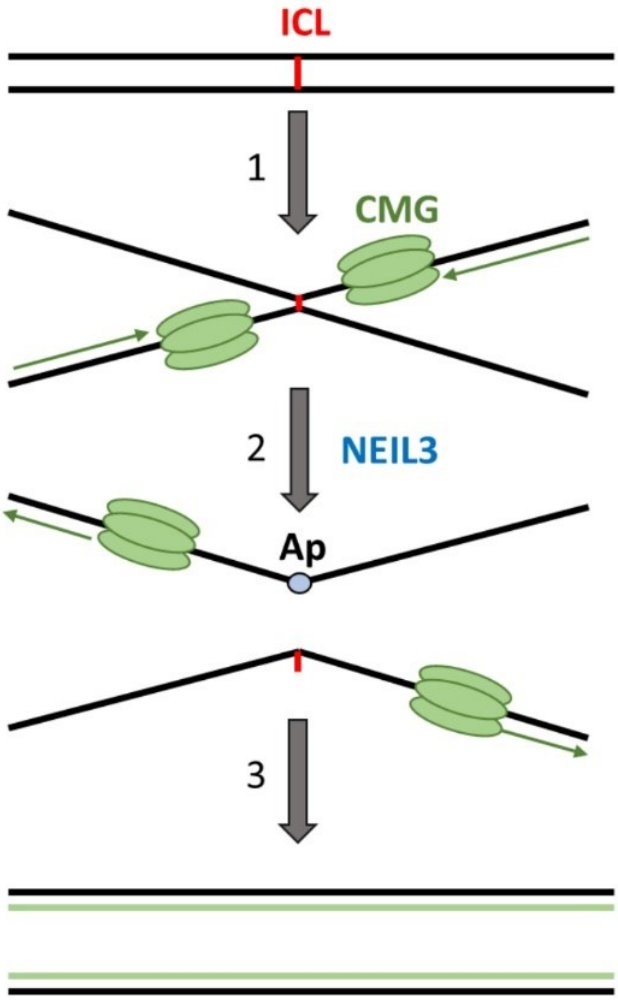

Figure 3.10. Schematic representation of NEIL3 pathway. After replication fork convergence at the ICL (red), CMG (green) helicase is ubiquitylated (1). NEIL3 is recruited (2) to unhook the lesion leaving one strand with monoadduct, in the case of Ap-ICL it is adenosine, and the other strand with Ap site (blue), which is bypassed by TLS (3) leaving nascent strand (green) with potential point mutation.

Neisseria meningitidis is a human pathogen responsible for severe infections, including bacterial meningitis, septicemia, and meningococcal pneumonia. Despite its relatively high fatality rate, this bacterium is primarily found in proximity to the human larynx [82,83]. Its small genome contains only a limited number of DNA repair genes, including one for Fpg/Nei glycosylase [84,85]. Through the study of this bacterium, insights can be gained into its DNA repair enzymes and pathways, providing a valuable model for understanding minimal DNA repair systems [86,87].

Investigating *Neisseria meningitidis* has contributed to our understanding of its pathogenicity and DNA repair mechanisms, which can have broader implications in the field of DNA repair research and potential therapeutic developments. Moreover, *Neisseria meningitidis* with its limited repertoire of DNA repair enzymes serves as an ideal model organism to study various DNA repair pathways and their interactions.

4 Aims of the study

- MutM structure: Investigate the *Neisseria meningitidis* MutM structure using protein crystallization to reveal its atomic structure and understand its role in DNA repair.
- MutM:DNA structural and functional characterization: Examine MutM from *Neisseria meningitidis* bound to DNA substrate, utilizing protein crystallization to reveal its atomic structure. Conduct binding experiments and determine dissociation constants to understand its interaction with potential substrates using fluorescence anisotropy and gel-based assays.
- Ap-ICL formation: Optimize spontaneous production of the abasic site DNA interstrand crosslink and explore how DNA sequence in the vicinity of the rising damage influences Ap-ICL formation and stability.

5 Publications

5.1 Publications included in dissertation thesis

Landova, B., and Silhan, J. (**2020**) Conformational changes of DNA repair glycosylase MutM triggered by DNA binding, *Febs Letters 594*, 3032-3044.

Huskova, A., Landova, B., Boura, E., and Silhan, J. (**2022**) The rate of formation and stability of abasic site interstrand crosslinks in the DNA duplex, *DNA Repair 113*.

5.2 Publications not included in dissertation thesis

Horova, V., Landova, B., Hodek, J., Chalupsky, K., Krafcikova, P., Chalupska, D., Duchoslav, V., Weber, J., Boura, E., and Klima, M. (**2021**) Localization of SARS-CoV-2 Capping Enzymes Revealed by an Antibody against the nsp10 Subunit, *Viruses-Basel 13*.

6 Materials and Methods

6.1 Protein expression and purification

MutM from *Neisseria meningitides* (strain MC58, UniProt P55044) was expressed in *E. coli* BL21 Star (DE3) and RosettaRM (DE3) pLys strains. Bacterial suspensions were grown in LB medium at 37°C at 220 rpm until O.D. (600 nm) = 0.6 was reached and then the temperature was lowered to 20°C. Protein production was induced by adding IPTG of final concentration 0.5 mM. The cells were incubated for further 12 – 16 h and then pelleted using centrifugation at 6000g for 6 min (JLA 9.1 centrifuge rotor). These pellets were frozen, thawed and resuspended in 1x PBS supplemented with 2 mM β-ME and 10% glycerol. The cells were homogenized using Dounce like homogenizator and sonicated 3 min at 80 % amplitude with protocol 1 s pulse and 3 s regeneration. Sonicated suspension was separated by centrifugation at 35 000 g and clarified lysates were loaded at anion – cation exchange chromatography tandem setup using 5 ml HiTrap Q HP and 5 ml Sp HP (GE Healthcare) columns. Loading buffer was 20 mM Tris buffer pH = 7.4, 125 mM NaCl, 2 mM β-ME, 10% glycerol. After the lysate was loaded, the anion column was removed from the tandem and cation column was washed. MutM protein was eluted with gradient of an elution buffer composed of 20 mM Tris buffer pH = 7.4, 1 M NaCl, 2 mM β-ME, 10% glycerol. The peak of MutM was observed at conductivity 25 $mS.cm^{-1}$. The final purification step was gel permeation chromatography where protein was loaded on Superdex 200, 16/60 (GE Healthcare) equilibrated in buffer: 20 mM Tris pH = 7.4, 500 mM NaCl, 2 mM β-ME, 10% glycerol. Protein was concentrated to 8 mg ml^{-1} and frozen for further use. The protein identity was confirmed using mass spectrometry.

His-tagged mNEIL3 from *Mus Musculus* was expressed in *E. coli* NiCo BL21 DE3 strain. Expression was the same as described above. Bacterial pellets were resuspended in lysis buffer containing 20% Tris pH 8.0, 300 mM NaCl, 20 mM Imidazole pH 8.0, 2 mM β-ME and 10% glycerol. Homogenization was again similar as in the case of MutM. First purification step for mNEIL3 was affinity Ni-NTA chromatography, where clarified lysates was incubated with Ni resin for 30 min at

4°C and rotating. After the incubation resin was washed with lysis buffer and then protein was eluated with elution buffer containing 20 mM Tris pH 8.0, 300 mM Imidazole pH 8.0, 300 mM NaCl, 2 mM β-ME and 10% glycerol. His-tag was cleaved out using 3C protease prepared in our lab during dialysis over night against buffer without imidazole. Last purification step was gel permeation chromatography carried in crystallisation buffer containing 20 mM Tris pH 8.0, 200 mM NaCl, 2 mM β-ME and 10% glycerol. Pure protein was concentrated to 10 mg ml^{-1} and frozen for further use.

6.2 MutM - DNA binding gel-based assay

HEX (hexachloro-fluorescein) labelled and unlabelled oligonucleotides containing THF (tetrahydrofuran) analogue that mimics Ap site were custom synthetized (Sigma-Aldrich, sequences are listed in Table 1. For binding assays all oligonucleotides, single-stranded (ssDNA) and double-stranded (dsDNA) substrates both containing THF or not were resuspended in reaction buffer composed of 20 mM Tris pH = 7.4, 120 mM NaCl, 0.5 mM TCEP, 5% glycerol. DsDNA substrates were prepared by mixing the complementary oligo strands in equimolar ratio and annealed by heating at 95°C for 2 min and allowed to slowly cool to 25°C at room temperature. 50 nM of each of the labelled DNA substrates (namely ss THF and ss no THF for ssDNA, ds THF and ds no THF for dsDNA) was titrated with pure MutM form 0 concentration to 2 uM and separated on 15% native PAGE and visualized on TYPHOON FLA 7000 fluorescence imager (GE Healthcare). Gels were qualitative control of DNA binding completing fluorescence anisotropy. No loading dye was used for native PAGE, only 10% glycerol.

Table 6.1. Sequence of oligonucleotides used in binding assays.

	Sequence 5′-3′
no THF	[HEX]GCAGCACCGACCACG
THF	[HEX]GCAGCAC[THF]GACCACG
comp	CGTGGTCGGTGCTGC

6.3 Fluorescence anisotropy DNA binding assay

To determine K_D we used the fluorescence anisotropy measurements in identical reaction buffer and with the same substrates as described in gel-based assays. Concentration of HEX labelled DNA was 50 nM in the reaction buffer. This reaction mixture was titrated with meningococcal MutM. Fluorescence spectrometer FluoroMax-4 HORIBA Scientific was set-up on wavelength 535 nm for absorption max and 556 nm fir emission max. Quantum yield (Q) represents the ratio of total intensities (I_{tot}) of free and bound fluorophore. Following formula was used for calculation of total intensity increases during titration:

$$I_{tot} = I_{VV} + 2 \times G \times I_{VH} \quad (1)$$

Where G is the G-factor depending on the optical system of the instrument. It is calculated using horizontal and vertical intensities. The fraction of bound DNA (f_B) was calculated using formula:

$$f_B = \frac{(A_{obs} - A_{min})}{(A_{max} - A_{obs}) \times Q + (A_{obs} - A_{min})} \quad (2)$$

Where A_{obs} is observed anisotropy for MutM concentration, A_{min} is the minimal observed anisotropy and A_{max} is the maximum observed anisotropy in the case of saturation [88]. We did four different measurements and therefore the error bars were calculated. Using formula:

$$f_B = (\tfrac{1}{2}[DNA]) \times ((K_D + [DNA] + [MutM] - \sqrt{(K_D + [DNA] + [MutM]^2) - 4[DNA][MutM]})) \quad (3)$$

The dissociation constant K_D was calculated. [DNA] and [MutM] are concentrations at given point of the titration. The nonlinear fit of the data was performed in GrafPad.

6.4 Crystallization, data collection and data analysis

The sitting drop vapor diffusion method was used to conduct the crystallization trials of Apoform of MutM enzyme, which were ser up on the Mosquito crystallization robot (TTPlabtech, Melbourn, UK). Commercialy composed crystallization screens mixed in our lab were used – ProPlex, JCSG Core I Suite, JSCG Core II Suite, Helix and Morpheus (Molecular Dimensions Limited, Sheffield, UK). Drops were set in ration 150:150 nl reservoir vs. protein solution and 250:150 nl. Crystals were grown in 18°C for minimum 48 hours and those chosen for diffraction experiment were grown in well solution containing 100 mM Bicine pH 8.5, 20% (w/v) PEG 6000, final pH 9.0. For cryoprotecting crystals before freezing in liquid nitrogen it was used supplemented original well solution with 20% glycerol. For testing diffraction on the frozen crystal, it was used the home source (Cu-Kα). Final dataset leading to the structure (PDBID: 6TC6) was collected on BESSY II Hemholtz-Zentrum Berlin, Germany.

For crystal structure of MutM bound to DNA firstly, the complement DNA oligos commercially synthetized: TGTCCA[THF]GTCTACC and AGGTAGACCTGGAC were annealed in buffer containing 20 mM Tris pH 8.0, 50 mM NaCl and 1 mM EDTA by heating at 98°C for 5 minutes and allowed to cool to the room temperature in PCR block. Enzyme was mixed with duplex DNA in molar ratio 1:1.5, desalted into crystallization buffer containing 10 mM Tris pH 8.0, 50 mM NaCl, 1mM β-ME using HiTrap Desalting Column (GE Healthcare). Afterwards the complex was purified on Superdex 200 (GE Healthcare) and concentrated before crystallization experiment. Complex presence and purity was confirmed using SDS PAGE and 15% native urea gel, where DNA was visualised with SYBR Gold Nucleic Acid Stain (Thermo Fisher Scientific, Waltham, MA, USA) and protein gel was coloured by Coomassie Brilliant Blue R-250. Fractions containing both protein and DNA were concentrated to final

concentration 200 μM. Crystals were grown at 18°C 1:1 ratio of complex to well solution set up in the same composition as described above. Crystals were observed in condition containing 100 mM Tris/HCl, 200 mM MgAc, 15% (w/v) PEG 6000. As cryoprotectant was again used 20% glycerol in the well solution and crystal was frozen in liquid nitrogen. Dataset was collected at the Diamond Light source, Didcot, UK.

Several datasets for bound enzyme form with DNA were collected and indexed in XDS, all with the same cell parameters. Two of them with the best parameters were combined in XSCALE and merged into mtz file [89]. The crystal structure of bound DNA form (6TC9) was determined using holoenzyme (6TC6) as the model. Unbound form structure was determined with molecular replacement in Phaser using model generated in Phyre2 based on *E.coli* MutM structure (1K82). In both cases CTD and NTD domains were separated because of their flexibility and solved individually. Afterwards, both domains were joint together by manual building in Coot and the refinements were performed in Phenix Refine [90,91]. All represented figures were modelled in PyMol.

6.5 Structure alignment

Structures were aligned in PyMol software using superposition tool. Holoenzyme (6TC6) was aligned to DNA bound form (6TC9) using conserved CTD domain (Ser-141 – Lys-275) composed of H2TH, HTH and ZnF, in total 704 atoms form CTD were used.

6.6 Preparation of DNA containing Ap site

All oligonucleotides used in this study were commercially synthetized by Sigma-Aldrich and all oligonucleotides are listed in Table 2. For Ap preparation all

oligonucleotides in need of Ap site were ordered with deoxy-uracil and all were labelled by HEX.

Table 6.2. Sequences of oligonucleotides used in ICL formation experiments.

Name	5′-3′sequence	Comp. strand	5′-3′sequence
4G	GTCGATGAACUTAGACAGCT	4Gc	AGCTGTCTAAGTTCATCGAC
4C	GTCGATGAAGUTAGACAGCT	4Cc	AGCTGTCTAACTTCATCGAC
4T	GTCGATGAAAUTAGACAGCT	4Tc	AGCTGTCTAATTTCATCGAC
4(5)A	GTCGATGAATUTAGACAGCT	4Ac	AGCTGTCTAAATTCATCGAC
5G	GTCGATGAACUTAGACAGCT	5Gc	AGCTGTCTAGATTCATCGAC
5C	GTCGATGAACUTAGACAGCT	5Cc	AGCTGTCTACATTCATCGAC
5T	GTCGATGAACUTAGACAGCT	5Tc	AGCTGTCTATATTCATCGAC
4I	GTCGATGAACUTAGACAGCT	4Ic	AGCTGTCTAAITTCATCGAC
6I	GTCGATGAACUTAGACAGCT	6Ic	AGCTGTCTIAGTTCATCGAC
TT	GTCAGATCAAAUTAAACCTG	TTc	CAGGTTTAGTTTGATCTGAC
TC	GTCAGATCAAGUTAAACCTG	TCc	CAGGTTTAGCTTGATCTGAC
GT	GTCAGATCCCAUTCCACCTG	GTc	CAGGTGGAGTGGGATCTGAC
GC	GTCAGATCCCGUTCCACCTG	GCc	CAGGTGGAGCGGGATCTGAC

6.7 DNA crosslink formation

Reaction buffer were optimized to its final composition of 20 mM HEPES pH 6.5, 140 mM NaCl, 0.5 mM TCEP, 5% glycerol. Experiments under physiological conditions were carried in 20 mM Tris pH 7.4, 140 mM NaCl and 5% glycerol for comparison. Labelled DNA oligonucleotides were mixed with nonlabelled complementary oligonucleotides in ratio 1:1 and annealed in Biometra Tprofessional Termocycler by heating at 95°C for 5 minutes and left to cool to room temperature. Final concentration in reaction buffer of dsDNA was 2.5 μM. Reactions were carried

out in 50 µl volume. Ap site was prepared by adding 0.5 U Uracyl-DNA-glycosylase (UDG) (New England Biolabs) per one reaction and kept for 5 min at room temperature. Subsequently, UDG was deactivated by heating the reaction in another round of annealing in thermocycler. To verify annealing process reactions were separated on 15% native gel. Afterwards, reactions were kept in 37°C and at given points 1 µl of crosslinking reactions were taken out and stopped by addition of 9 µl of formamide and either frozen or directly separated on 20% denaturating urea gel composed of 1x TBE buffer, 20% acrylamide/bisacrylamide (19:1) and 7M urea. Gels were visualised using Amersham Typhoon Biomolecular Imager (GE Healthcare) set up at 532 nm laser and using Cy3 570BP20 filter. Separated visualised gels were quantified using commercial software ImageQuant TL. The lines were manually chosen, and the rolling ball method was applied for background subtraction. All bands were calculated in separate lines including substrate, product and intermediates to calculate percentage of the crosslink formed. The percentage of the ICL formation was calculated from three independent reaction sets and gels. The methodology was described before in detail.

6.8 Isolation of ICL from polyacrylamide urea gel for further use

Ap-ICL was purified from 20% denaturating polyacrylamide urea gel. Separated bands were visualized using HeroLabUVT-20 S/M/L and the crosslinked band was cut out and transferred to a tube containing 200 µl of elution buffer composed of 20 mM Tris pH 7.4, 140 mM NaCl, 0.5 mM TCEP, 5% glycerol. The tube was left in rotating mixer at 8°C over night. The sample was centrifugated at 800 rpm for 1 min a supernatant was transferred into desalting columns (Cytiva MicroSpin Comlumns G-25) equilibrated with elution buffer to remove excess urea. For verification, pure sample was run at 20% denaturating gel to confirm its purity.

6.9 Degradation analysis of ICL

The stability experiments were carried out in 37°C incubator in elution buffer. 2 µl of the reaction was taken out at given time points (every 1 day at the same time for 14 days). Reaction was stopped by addition of 8 µl of formamide and stopped reaction was frozen at -80°C to prevent further degradation before gel analysis. All terminated reactions were run on 20% denaturation urea gel. Gels were visualised using Amersham Typhoon Biomolecular Imager (GE Healthcare) set up at 532 nm laser and using Cy3 570BP20 filter. Separated visualised gels were quantified using commercial software ImageQuant TL. The lines were manually chosen, and the rolling ball method was applied for background subtraction. The percentage of the ICL degradation products was calculated from three independent reaction sets based on the ratio of the remaining crosslinked DNA and the degradation product.

7 Results

7.1 Crystal structure of Meningococcal MutM

Initially, the crystal structure of holoenzyme MutM from *Neisseria Meningitidis* was solved from crystal diffracting to a resolution of 2.9 Å. This structure was solved by molecular replacement using the MutM from *E. coli* (PDBID: 1K82). A homology model was constructed using Phyre2 web server, with a sequence identity of 54% [92,93]. The initial homologous model, based on the protein-DNA complex, did not produce satisfactory electron density maps after performing molecular replacement using Phaser [91]. However, we achieved a successful solution by dividing the model into NTD and CTD subdomains. These subdomains were used as independent search models for molecular replacement, highlighting differences in the orientation of the NTD and CTD subdomains between the free and DNA-bound structures [30]. Subsequently, the model of the meningococcal MutM enzyme was refined to its final structure at a resolution of 2.9 Å. Notably, the region spanning Thr217 to Tyr242, typically well-ordered only when a base is flipped out, exhibited poor visibility in the electron density and was therefore excluded from the final model.

In general, the crystal structure of full-length MutM protein from Neisseria meningitidis exhibits a similar fold to its homologues in other organisms. The NTD consists of a large and one small α-helix positioned on opposite sides of a β-stranded sandwich. This sandwich structure is composed of two halves, each containing four antiparallel β-sheets (Figure 7.1.). The NTD is connected to the CTD through a coiled-coil region (Pro130-Phe144) with well-defined electron density. This region includes the flexible hinge region mentioned earlier. The CTD consists of four helices, with the last two helices featuring an H2TH motif, followed by a zinc finger domain characterized by four cysteine residues ([Cys]4) [29].

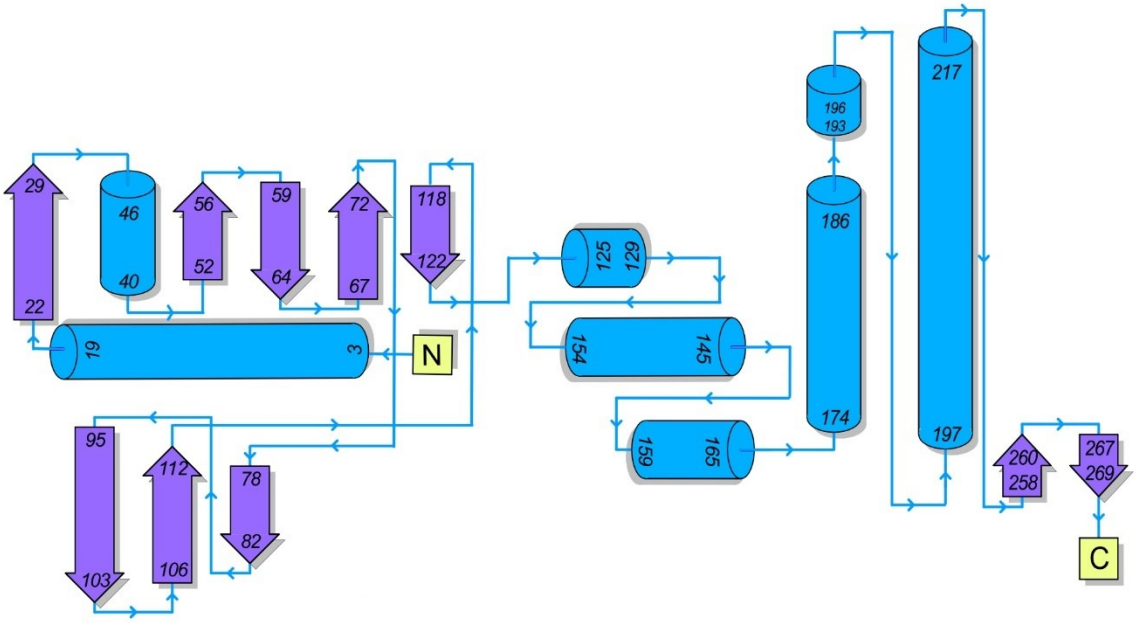

Figure 7.1. Topology plot of apo structure MutM (6TC6). Helices are visualised in blue and β-sheets in purple.

7.2 MutM binds synthetic oligonucleotides with Ap-like analogue THF

Before crystallization bound MutM with its substrate, we carried out binding experiments to identify the best binding partner with Ap site analogue THF (replaces ribose, i.e., 1′,2′-dideoxyribofuranose-5′-phosphate). For comparison, we also used DNA without any damage and a mimic of damaged DNA. Both double-stranded and single-stranded DNA were tested using fluorescence anisotropy to measure dissociation constants (Figure 7.2.). The oligonucleotides were commercially synthetized and their sequence was GCAGCAC(THF/C)GACCACG/ CGTGGTCGGTGCTGC, where in THF replaced cytosine to mimic the Ap site in one strand. The complement strand for dsDNA and dsTHF remained the same. We observed the weakest affinity, measured at 27 nM, in the case of ssDNA (without damage), which still falls within the low nanomolar range. K_D values for the rest of the substrates, including dsDNA, ssTHF and dsTHF were similar, ranging between 7 – 9 nM.

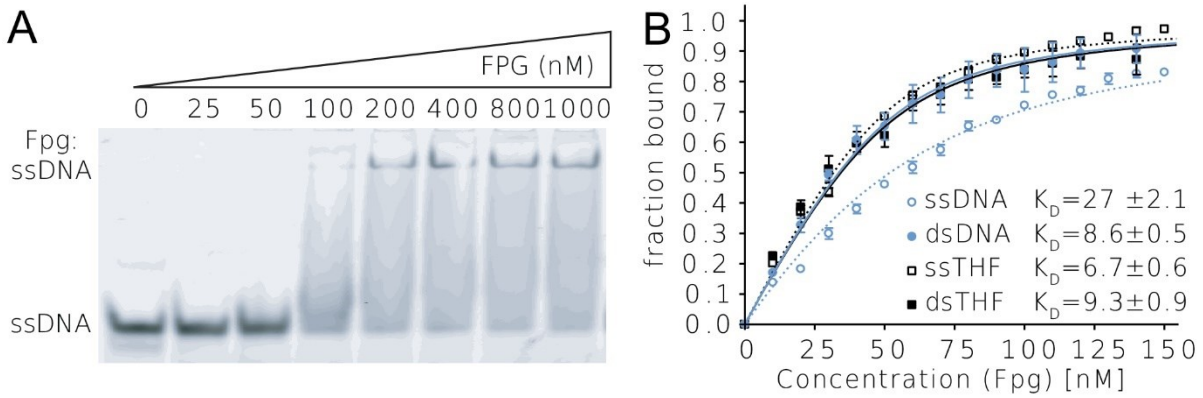

Figure 7.2. DNA binding properties of MutM. DNA containing THF which mimics Ap site in DNA substrates was used. (A) Gel-based DNA binding assay based on electrophoretic mobility shift used for determination of qualitative access of the THF-containing ssDNA labelled by fluorescent dye HEX to MutM. The oligonucleotide was incubated with the MutM protein, and reaction separated on 15% native PAGE. (B) Fluorescent anisotropy binding assay used to compare the K_D of substrates, either containing or lacking THF, which mimics the Ap site in both ssDNA and dsDNA. Blue curves represent undamaged oligonucleotides, while black curves represent oligonucleotides containing THF. K_D values were fitted using formula (3). Error bars represent standard errors resulting from four independent titrations.

7.3 Crystal structure of MutM bound to DNA

An oligonucleotide containing Ap site analogue THF and the sequence TGTCCA[THF]GTCTACC/AGGTAGACCTGGAC was used for diffraction experiment in its double stranded form. The crystal structure of MutM with DNA (PDBID: 6TC9) was solved by molecular replacement using MutM structure from our study (PDBID: 6TC6) as a search model. To achieve a successful molecular replacement, the model had to be further divided into NTD and CTD parts. Each domain was independently placed in Phaser software [91]. The resulting model of DNA-bound MutM was subsequently refined to a resolution of 2.15 Å using Coot software (Figure 7.3.) [90]. Data collection statistics for both crystal structures can be found in Table 7.1. Upon binding to DNA molecule, MutM induces a twist of

approximately 60° in the DNA, in comparison to the canonical B-DNA structure. Importantly, this conformational change in the DNA is not attributed to the presence of THF, as unbound DNA containing similar Ap sites maintains a normal B-DNA structure.

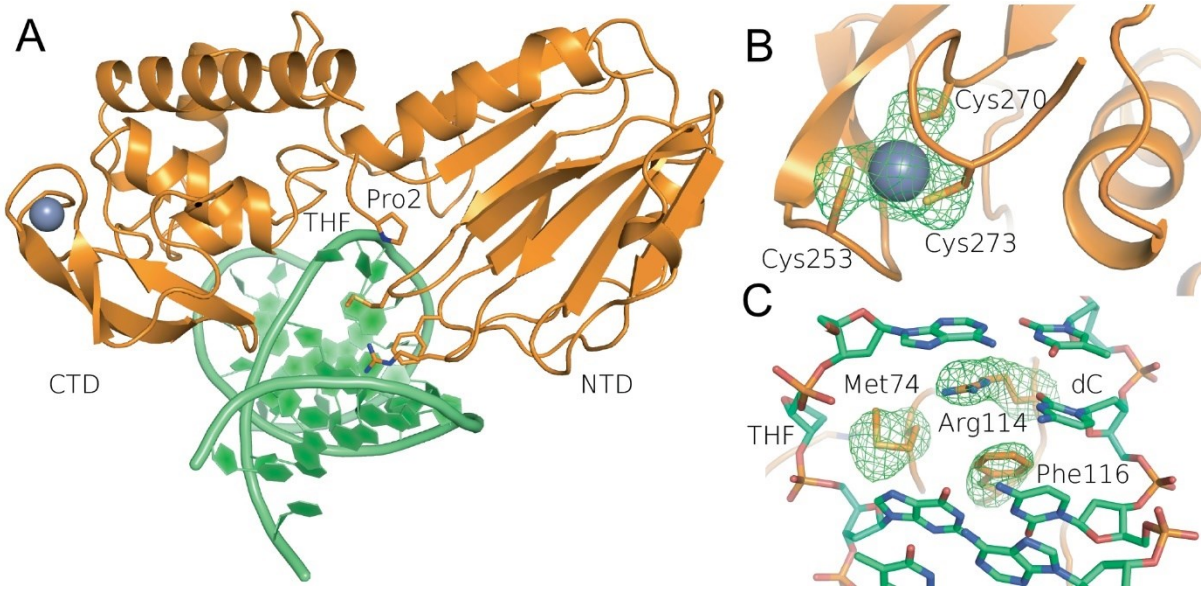

Figure 7.3. Crystal structure of DNA-bound meningococcal MutM (6TC9). (A) Structure of MutM enzyme (orange) bound to THF containing dsDNA substrate (green). (B) Focus on the Fpg/Nei family ZnF - Zn[Cys]$_4$, where Cys250 is hidden behind. The omit map of electron density (green) is shown at sigma level=3. (C) Focus on highly conserved residues Met74, Arg114 and Phe116 with electron density.

Specifically, the presence of THF in the DNA structure is exposed and rotated towards the active site, accompanied by the catalytic Pro2 residue as if the base were "flipped out." It is important to note that this MutM:DNA structure does not contain any damaged bases; instead it features an Ap site analogue. The binding process to the damaged base involves the nucleophilic attack of the Pro2 nitrogen side chain on the sugar at position C1′ of the damaged base. The resulting ring-opening intermediate is stabilized by Glu3 through protonation at the C5′-OH position. Additionally, Lys 54 protonates the 3′-phosphate, leading to β-elimination [39]. These observations align with previous MutM:Ap-DNA structures, where the presence of a damaged base was

not necessary for the occurrence of the “base-flipping” mechanism [94]. During DNA binding, a distinct “void space” emerges between the orphaned base and the THF residue.

Within the Fpg/Nei family of proteins, three highly conserved residues, namely Met74, Arg114 and Phe116, play significant roles in the region. These residues extend into the space between the orphaned base and the widened minor groove of the DNA substrate. Specifically, Phe116 is positioned in a stacked arrangement between the orphaned base and the neighbouring base pair. Arg114 extends into the major groove of the DNA, establishing hydrogen bonding interaction with the orphaned cytosine. Met74 undergoes an inward movement towards the damage site (THF), located opposing to the catalytic Pro2 residues. These observations highlight the intricate structural rearrangements that occur within the DNA-protein complex during the recognition and binding of damaged DNA.

Table 7.1. Data collection and refinement statistics

Crystal structure	***Fpg: DNA THF (6TC9)***	***Fpg (6TC6)***
Diffraction source	Diamond I-04-1	BessyII
Wavelength (Å)	0.91730	0.9184
Space group	C121	P 1 21 2
Cell a, b, c (Å)	130.2 80.1 83.4	66.4 38.0 110.3
Cell α, β, γ (°)	90, 90.035, 90	90, 89.96, 90
Resolution range (Å)	34.1 - 2.2 (2.28 - 2.2)	42.43 - 2.85 (2.95 - 2.85)
Total no. of reflections	80506 (8009)	375339 (36903)
No. of unique reflections	40438 (4140)	12736 (1162)
Multiplicity	2.0 (1.9)	29.5 (29.5)
Completeness (%)	0.93 (0.96)	0.90 (0.90)
Mean I/σ(I)	15 (2.61)	9.23 (1.05)
R-merge (%)	3.4 (50.1)	4.11 (31.41)
CC1/2	0.999 (0.693)	0.844 (0.207)
CC*	0.998 (0.905)	0.957 (0.586)
R-work (%)	25.26 (38.97)	23.02 (37.82)
R-free (%)	27.79 (39.22)	25.53 (36.96)
Number of non-H atoms	4934	3971
RMSD Bonds (Å)	0.004	0.013
RMSD Angles (°)	0.85	1.6
Ramachandran (%)	92/0.8	96.14/0.4

Statistics for the highest-resolution shell are shown in parentheses.

7.4 The conformational changes induced in MutM after binding DNA

The binding of DNA induces significant conformational changes in MutM. Upon comparing the structures, it is evident that both domains undergo substantial

reorganization. To examine these molecular movements in more detail, the proteins were superimposed using least-square methods, both as whole and for individual domains, using PyMol software. The overall root-mean-square deviation (RMSD) for the entire protein chain was relatively high at 2.41 Å, resulting in poor overlap of the protein chains. However, aligning the individual domains revealed a rotation of the C-terminal domain (CTD) and N-terminal domain (NTD) towards the DNA by an angle of 22.30°. Additionally, the superimposition of NTDs showed a relative displacement of the zinc atom from the ZnF domains by 11.4 Å (Figure 7.4.).

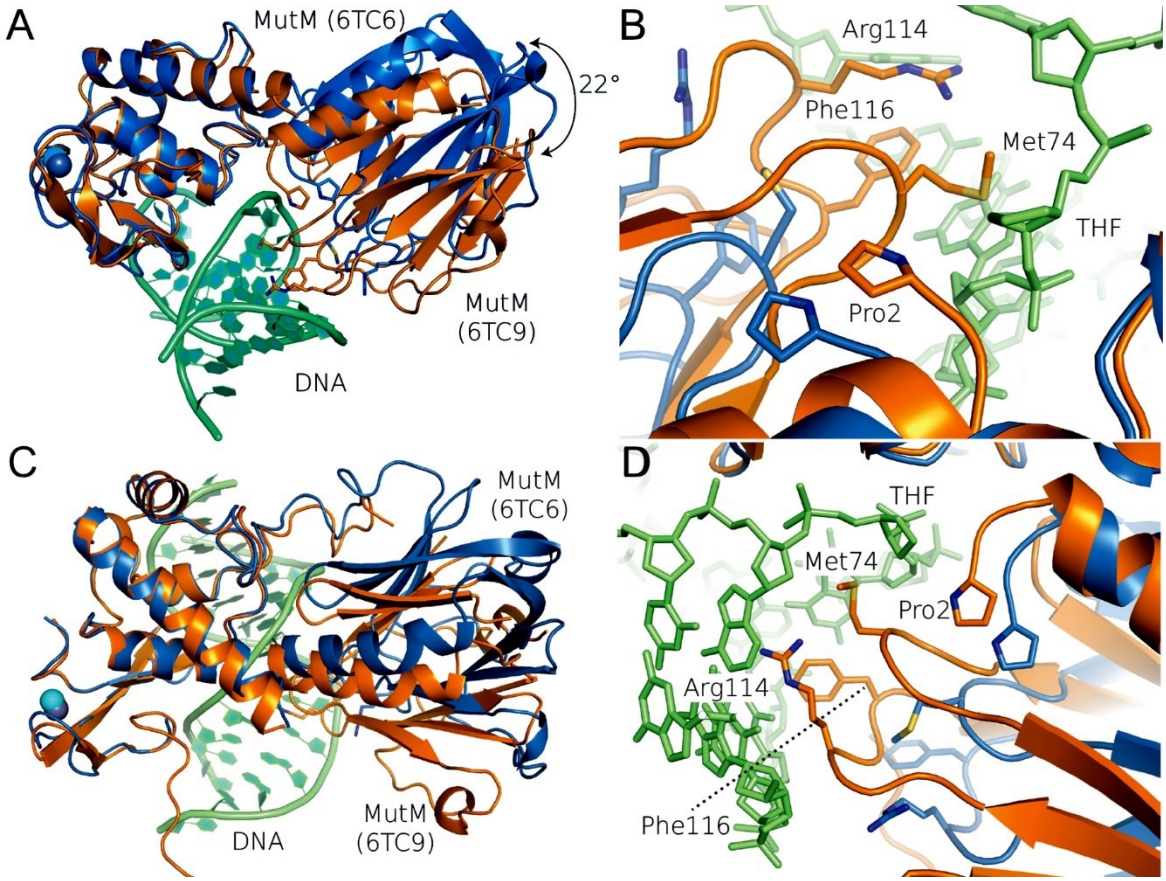

Figure 7.4. Structural alignments of CTDs of DNA bound MutM in orange and free MutM in blue (A-D) showing conformation changes upon DNA (green) binding. View is identical as in Figure 1.4 and 7.3 A illustrates the rotation movement of 22° in A panel with the top view in C panel. Views from different angles (B, D) show detail of residues Met 74, Arg114 and Phe 116 movement toward bound DNA. Met74 moves towards the THF to the opposite position to Pro2. All three residues seem to stabilize the protein in the substrate recognition state.

In summary, the overall RMSD was relatively high, indicating notable structural changes upon DNA binding. However, when examining the separate domains individually, the CTDs showed a low RMSD of only 0.649 Å, while the NTDs exhibited an even lower RMSD of 0.562 Å. This suggests that the domains largely maintain their overall architecture and experience primarily rotational and translational movements upon binding to DNA.

Structure alignment of previously published structures of related proteins to MutM or from different organisms from PDB database to structures from this study (Figure 7.5.). RMSD values of individual structures are shown in Supplementary table 1. Both our holoenzyme (PDBID:6TC6) and DNA-bound MutM (PDBID:6TC9) were aligned to *Bacillus stearothermophilus* holoenzyme (PDBID:1EE8), *E. coli* DNA-bound MutM (PDBID:1K82), *A. Thaliana* both DNA-bound and holoenzyme of MutM (PDBIDs: 3TWL, 3TWK) and DNA-bound and free Nei form *E. coli* (PDBIDs: 1K3W, 1Q3C) [27-34].

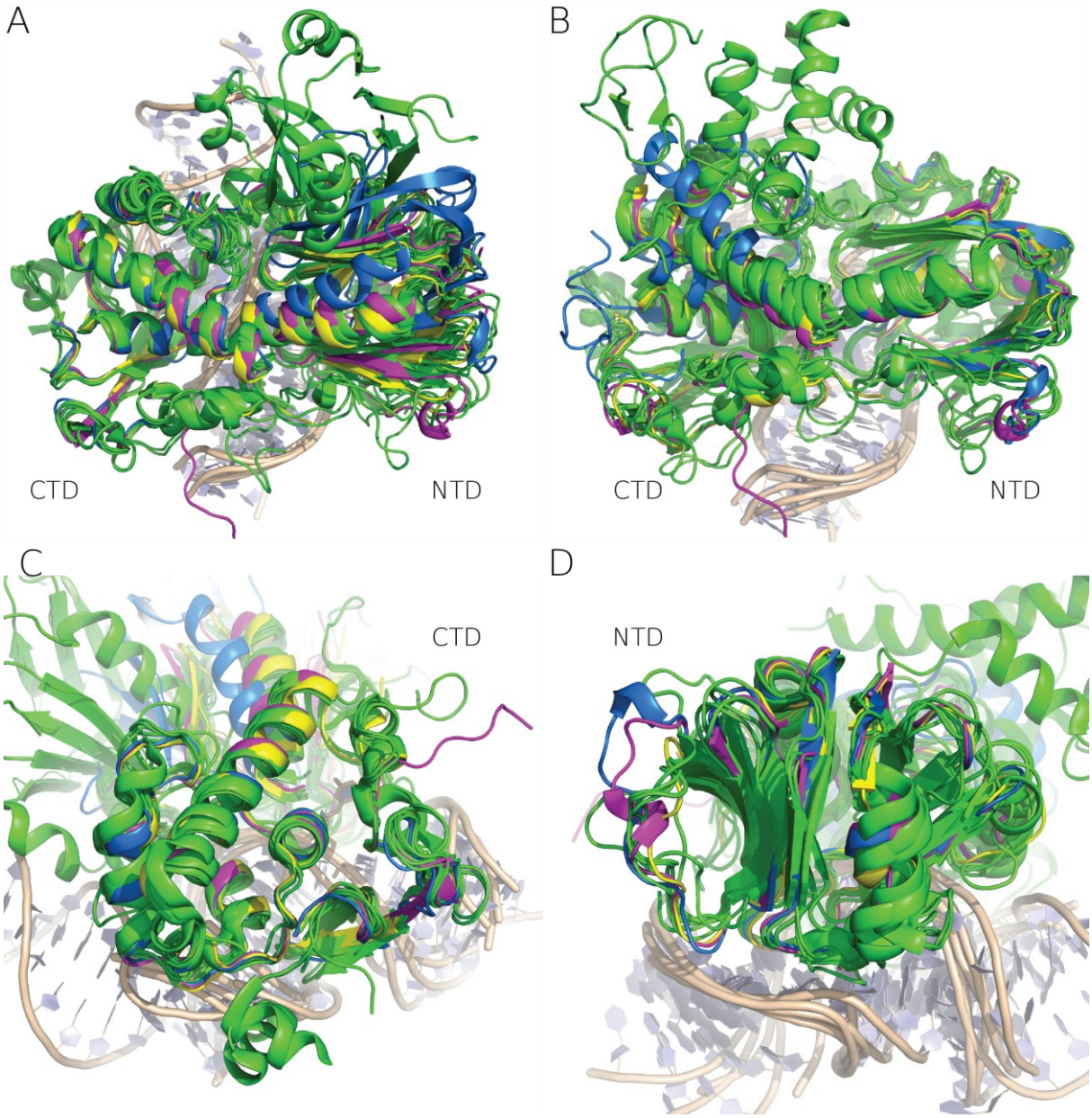

Figure 7.5. Structural superposition of previously solved structures of MutM from different views with meningococcal *MutM apoenzyme (PDBID: 6TC6) (blue) and MutM (6TC9) (magenta) with DNA (wheat) from this study. E. coli MutM (PDBID: 1K82) is shown in yellow and the rest of the structures from different organisms all in green (PDBIDs: 1EE8, 1XC8, 1L1T, 3GQ5, 4CIS, 3TWL, 3TWK). All structures were aligned by their CTDs (A, C) and structures aligned by their NTDs (B, D). Alignment values RDSM are shown in Supplementary table 1.*

7.5 DNA substrates design

The oligonucleotide duplex has the uracil residue, representing the Ap site, located at the center. To ease naming, all residues in the vicinity were labelled accordingly, and the position of the Ap site was designated as position 2. The crosslinks primarily formed with adenine at position 6, which was a fixed base in our experimental setup. Consequently, position 2 and 6 remained constant as they were involved in the formation of Ap-ICL, except for the control experiment that verified the preferential formation of these sites. Sequences with labeled positions are shown in Table 7.2. As control we performed experiments with guanine at position 6, which is another ICL possibility. In the vicinity of the Ap site, we introduce alternations only to the natural DNA nucleobases in order to explore all potential combinations. Position 5 served as the complement to the Ap site lesion. Given the natural complementarity between position 1 and 4, subsequent experiments focused on position 4 and 5. Additionally, a broader region surrounding the Ap site was examined with the specific aim of comparing the AT-rich and GC-rich regions.

Table 7.2. DNA oligonucleotides sequences

4G	5'[HEX]-GTCGATGAAC$_1$U$_2$T$_3$AGACAGCT 3'-CAGCTACTTG$_4$A$_5$A$_6$TCTGTCGA	4I	5'[HEX]-GTCGATGAAC$_1$U$_2$T$_3$AGACAGCT 3'-CAGCTACTTI$_4$A$_5$A$_6$TCTGTCGA
4C	5'[HEX]-GTCGATGAAG$_1$U$_2$T$_3$AGACAGCT 3'-CAGCTACTTC$_4$A$_5$A$_6$TCTGTCGA	6I	5'[HEX]-GTCGATGAAC$_1$U$_2$T$_3$AGACAGCT 3'-CAGCTACTTG$_4$A$_5$I$_6$TCTGTCGA
4T	5'[HEX]-GTCGATGAAA$_1$U$_2$T$_3$AGACAGCT 3'-CAGCTACTTT$_4$A$_5$A$_6$TCTGTCGA		
4(5)A	5'[HEX]-GTCGATGAAT$_1$U$_2$T$_3$AGACAGCT 3'-CAGCTACTTA$_4$A$_5$A$_6$TCTGTCGA	TT	5'[HEX]-GTCAGATCAAA$_1$U$_2$T$_3$AAACCTG 3'-CAGTCTAGTTT$_4$G$_5$A$_6$TTTGGAC
5G	5'[HEX]-GTCGATGAAT$_1$U$_2$T$_3$AGACAGCT 3'-CAGCTACTTA$_4$G$_5$A$_6$TCTGTCGA	TC	5'[HEX]-GTCAGATCAAG$_1$U$_2$T$_3$AAACCTG 3'-CAGTCTAGTTC$_4$G$_5$A$_6$TTTGGAC
5C	5'[HEX]-GTCGATGAAT$_1$U$_2$T$_3$AGACAGCT 3'-CAGCTACTTA$_4$C$_5$A$_6$TCTGTCGA	GT	5'[HEX]-GTCAGATCCCA$_1$U$_2$T$_3$CCACCTG 3'-CAGTCTAGGGT$_4$G$_5$A$_6$GGTGGAC
5T	5'[HEX]-GTCGATGAAT$_1$U$_2$T$_3$AGACAGCT 3'-CAGCTACTTA$_4$T$_5$A$_6$TCTGTCGA	GC	5'[HEX]-GTCAGATCCCG$_1$U$_2$T$_3$CCACCTG 3'-CAGTCTAGGGC$_4$G$_5$A$_6$GGTGGAC

Sequences of dsDNA oligonucleotides used as substrates in this study shown with their names. Ap site is generated by UDG from uracil (U) on the position 2 and Ap-ICL is formed with adenine on the position 6.

7.6 Ap site and Ap-ICL formation

The synthetic DNA oligonucleotides, commercially synthetized, contained deoxyuridine and were labelled at 5´with the fluorescent dye HEX. The presence of uracil allows generation of the Ap site. Initially, the complement oligonucleotides were annealed and subsequently incubated with Uracil-DNA glycosylase [95]. UDG was titrated into the reaction mixture to determine the minimal amount required. To assess the presence of the generated Ap site, the reaction mixture was separated on native PAGE gel electrophoresis after hydrolysis by NaOH. Typically, an incubation of 5 minutes with UDG was sufficient for complete conversion of uracil to the Ap site.

The crosslinking reaction was kept in cabinet incubator at 37°C. To obtain timepoints, aliquots of reaction were taken, and the reaction was stopped by adding formamide. All timepoints were analyzed by separation of 15% denaturating PAGE gels, visualized and quantified. The percentage of ICL formed was plotted on a graph.

Maximum yields of Ap-ICL were measured values and initial ones were calculated using a linear fit to the data.

7.7 The influence of the opposite base to the Ap site to the rate of formation of Ap-ICL

Ap sites are generated in DNA molecules spontaneously, regardless of the DNA sequence context. In our initial experiments, we focused on varying the single base located opposite the Ap site while keeping the sequence of the remaining DNA duplex identical. This position was referred to as the 5^{th} position. We measured the rate of Ap-ICL formation with all four nucleotides and determined the maximum yields for each of them. The obtained data are presented in both graph and bar chart formats with error bars generated based on three different measurements.

The observed rates of Ap-ICL formation were as follows: 1.1% per hour for 5A, 0.76% per hour for 5G, 0.91% per hour for 5C, and 0.99% per hour for 5T. We express the rates via converting them into the percentage of original Ap sites that were converted to Ap-ICL within one hour. The maximum yields for the respective oligonucleotides were 24.25% for 5A, 47.2% for 5G, 25.8% for 5C, and 35.9% for 5T, relative to the total DNA (Figure 7.6.).

As the amount of degradation product increases over time, the amount of Ap-ICL decreases. The graph shows significant differences in the rate of Ap-ICL among the tested bases. However, these changes were relatively small, and regardless of the nucleobase present opposite the Ap site, the formation of Ap-ICL occurred to a similar extent.

7.8 Varying of the base-pairs in the vicinity of Ap on its 5´end

Similarly, variations of the opposite bases, nucleotides occupying position 4 and 1 were tested (Figure 7.6.). For these positions, all four possible base pair combinations

were tested, resulting in the DNA oligonucleotides named 4G, 4T, 4C, and 4(5)A. Notably, sequences 4A and 5A were co-named 4(5)A since 4A remained constant when vary 5th position, and vice versa for sequence 5A. Significant differences in the formation of Ap-ICL were observed among these combinations. Surprisingly, despite these differences, Ap-ICLs were formed in all possible base combinations. The variations in formation rates were not exceptional, showing only an approximate 2-fold difference. The maximum yield varied slightly among the combinations and was reached at different time points. The initial rate of ICL formation was lower for the pyrimidine bases. Both thymine and cytosine-containing duplexes reached their maximum yield of ICL later, but the percentage were not significantly different from the other two bases at this position. As with previous experiments, the amount of ICL eventually started to decline due to irreversible decomposition into smaller degradation products.

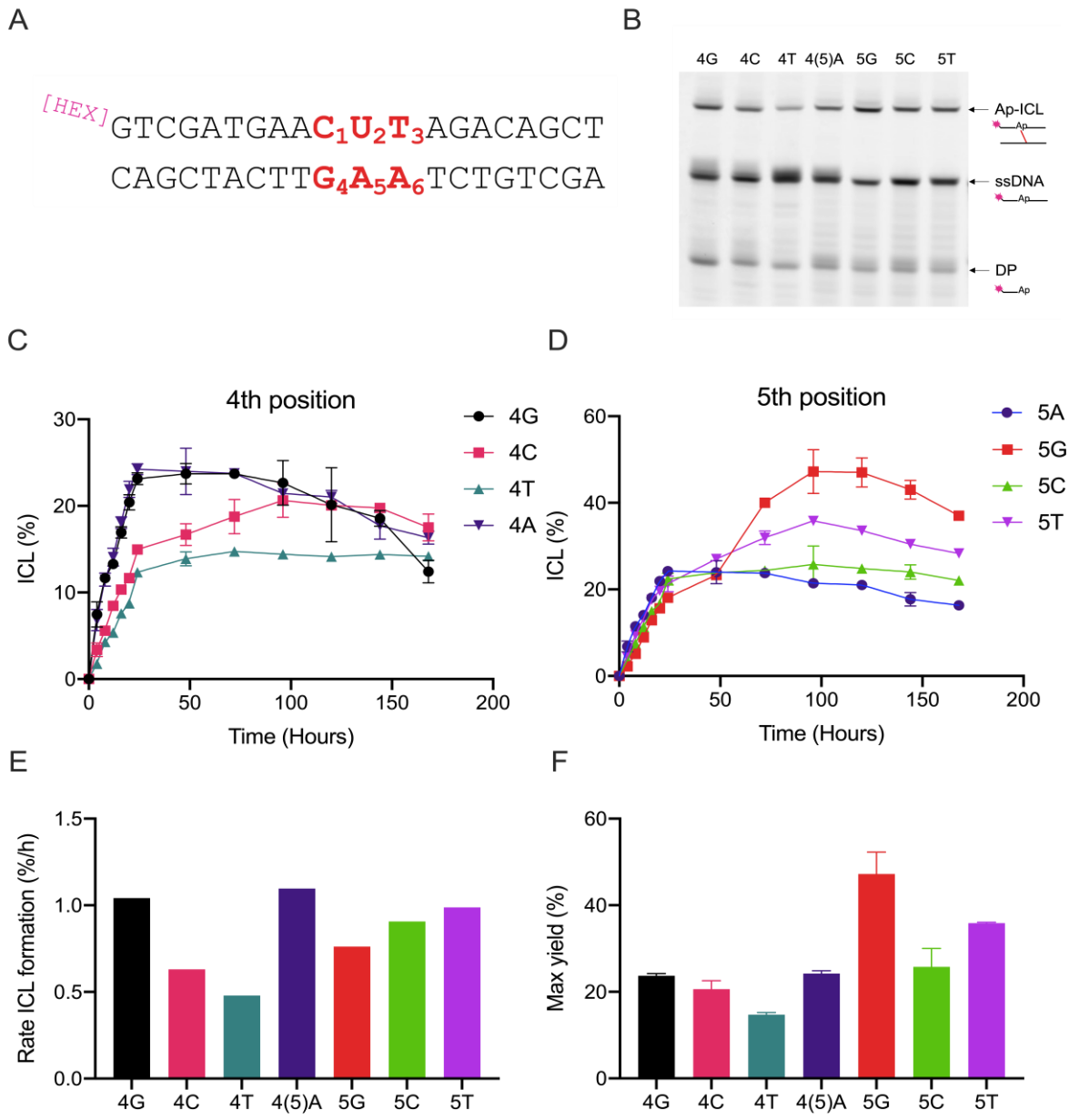

Figure 7.6. Reaction kinetics of the Ap-ICL formation when varying bases at 4th and 5th position respectively. (A) Schematic representation of DNA duplex containing Ap site with the nomenclature of differing positions. Uracil at position 2 is cleaved out by UDG and Ap-ICL is formed between positions 2 (where Ap is present) and adenine at position 6. The spontaneous reaction of Ap-ICL formation began after uracil cleavage. Reaction was kept in the dark at 37 °C and stopped by addition of formamide at defined time points (each 24 h) and resolved on 20% denaturating PAGE gel (B), where DP stands for product of degradation. After evaluation of Ap-ICL formed from DNA oligonucleotides with different bases in position 4 (C) and 5

(D) the proportion were plotted in the graph. The rates of Ap-ICL formation (E) and maximum yields of the reactions (F) were plotted as a bar chart. Error bars were calculated from 3 independent repeats.

To validate previously published work and demonstrate the specificity of Ap-ICL formation from Ap with adenine residues, two control experiments were conducted. In the case of sequence 4G, hypoxanthine was placed at positions 4 and 6. No Ap-G crosslink was formed when hypoxanthine was placed at position 6. On the other hand, when hypoxanthine was placed at position 4, an Ap-A crosslink was formed. These control experiments provide further evidence supporting the specificity of Ap-ICL formation with adenine residues.

7.9 AT-rich and GC-rich regions relation to amount of Ap-ICL formation

By investigating the sequence surrounding the Ap site, we have observed differences in the rate of Ap-ICL formation. However, to obtain a more comprehensive understanding, further investigation is required to explore additional sequence variations that may influence the overall rigidity of the region surrounding the Ap site.

To accomplish this, we modified the sequences in the wider vicinity of the Ap site by incorporating two AT or GC pairs on both sides. Our intent was to examine how AT-rich and GC-rich sequences impact the rate of Ap-ICL formation. These sequences were tested in non-identical duplicated, where the base at 4^{th} position was either cytosine or thymine. Consequently, the oligonucleotides were named accordingly (e.g., TT, TC, GT, GC). For instance, TC represents an At-rich oligonucleotide with a cytosine at the 4^{th} position. We specifically chose these bases to extend the AT-rich or GC-rich region around the Ap site.

The experimental setup was similar to the previous experiments, where the reaction progress was monitored using denaturating PAGE gel analysis. The formation of Ap-ICL was visually confirmed on the gel images. Subsequently, the gels were evaluated,

and the initial reaction rated were determined by calculating the slopes of the data points. This analysis allowed us to assess the impact of the modified sequences on the rate of Ap-ICL formation (Figure 7.7.).

Swapping the AT-rich and GC-rich regions resulted in noticeable disparities in the rate of Ap-ICL formation. The impact of the cytosine or thymine base at 4th position on the reaction rate was relatively minor, with the rates for oligos containing cytosine being slightly faster than those with thymine. This observation held true for both AT-rich and GC-rich DNA duplexes, affirming the reliability of our approach.

Notably, the AT-rich sequence with a cytosine at the 4th position (TC) exhibited an almost four-fold higher rate of formation compared to the GC-rich sequence with thymine at the 4th position (GT). This trend persisted across both pairs of sequences (TC>TT>>>GC>GT), and it also manifested in the maximum yield, which showed an approximate 2.5-fold difference.

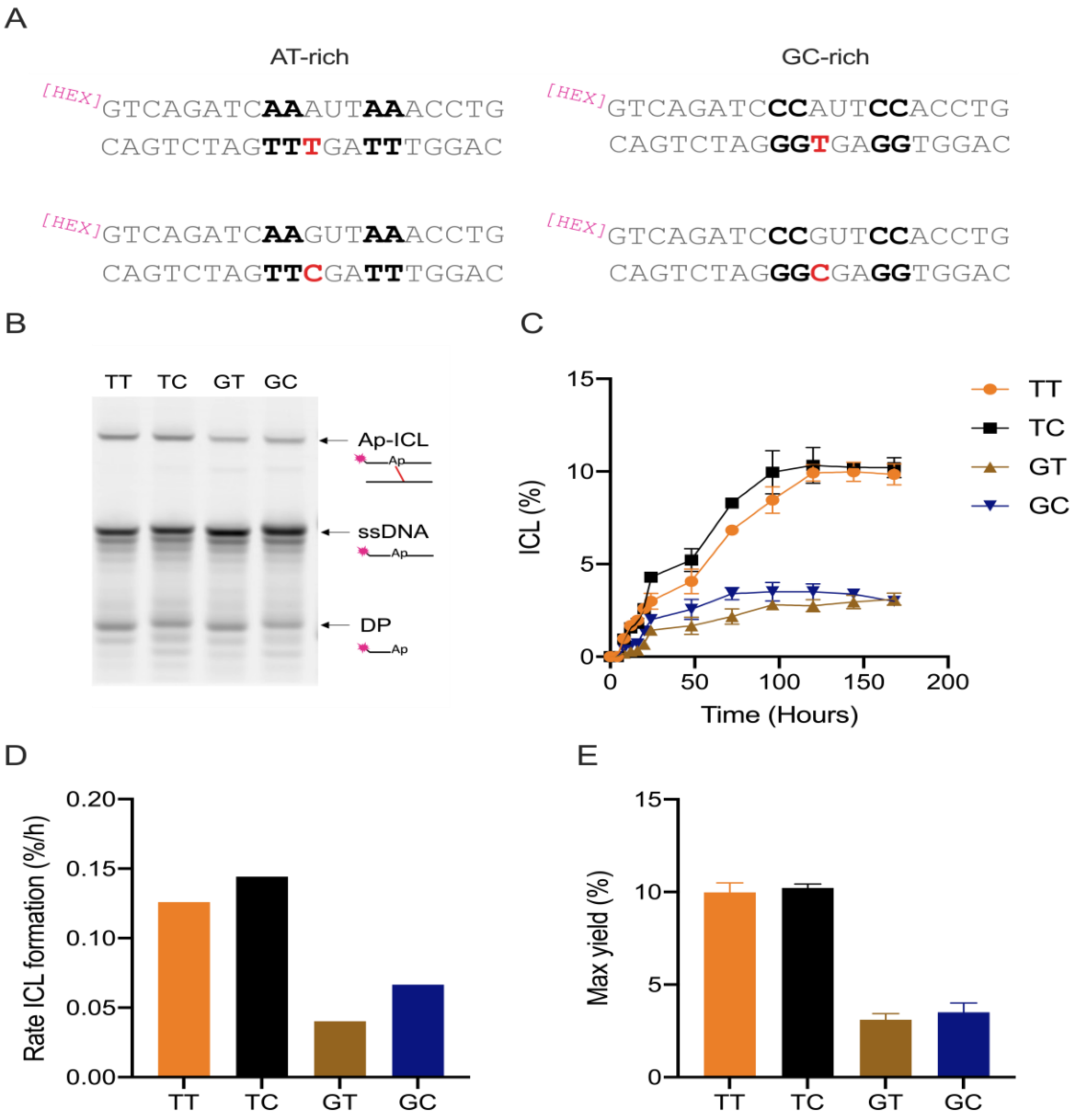

Figure 7.7. Formation of Ap-ICL in dsDNA containing AT-rich and GC-rich regions. Schematic representation of DNA duplexes containing AT-rich or GC-rich regions in the vicinity of the Ap site. AT-rich oligonucleotides with thymine or cytosine on the 4th position are named TT or TC. Likewise, GC-rich oligonucleotides with thymine and cytosine on the 4th position are named GT or GC. Stopped reaction were run on 20% denaturating PAGE gel (B), where DP stands for degradation product. (C) The reaction kinetics of Ap-ICL formation for AT and GC-rich oligonucleotides. (D) The rate of Ap-ICL formation and (E) the maximum yield plotted for all sequences. Error bars comes from three independent repeats.

These findings highlight the significant influence of sequence composition, specifically the AT-rich and GC-rich regions, on the rate of Ap-ICL formation. Additionally, the presence of a cytosine or thymine base at the 4th position contributed to subtle variations in the reaction rates. These results emphasize the importance of considering sequence context when studying Ap-ICL formation dynamics.

7.10 Stability of Ap-ICL

Finally, we conducted experiments to assess the stability of Ap-ICL under physiological conditions, without involvement of additional proteins or enzymes. The purified Ap-ICLs were obtained by isolating them from the gel. For this purpose, we halted the crosslinking reactions at a timepoint corresponding to the maximum yield of Ap-ICL. Following isolation, the pure Ap-ICLs were desalted to a physiological buffer to create a suitable environment for spontaneous degradation. Based on our initial experiment and existing literature, it has been indicated that the decomposition of Ap-ICL is a prolonged process [96].

To investigate the stability of Ap-ICL over time, subsamples were taken at 24-hour intervals. The collected samples were resolved on denaturating PAGE gels after a two-week period. The proportions of degrading Ap-ICL were plotted against time, and the initial rates of their decomposition were fitted using exponential decay and displayed on a bar chart (Figure 7.8.).

During the decomposition process, the ICL band gradually disintegrated into a band of similar size as the original substrate. Over time, additional products emerged, resulting from the spontaneous hydrolysis of the Ap site. Both AT-rich and GC-rich sequences were utilized in this experiment, and for the most part, the rates of decomposition were comparable among all reactions. However, there was a slightly slower rate of decomposition observed for AT-rich sequence with a thymine at the 4th position (TT). Within the first 10 days, less than one-fifth of the initial material had undergone decomposition.

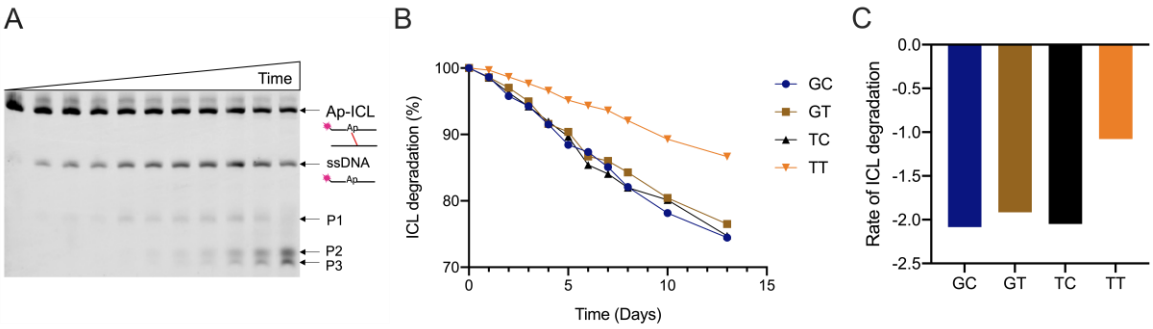

Figure 7.8. Stability and degradation of Ap-ICL in AT-rich and GC-rich oligonucleotides. Ap-ICL from AT-rich and GC-rich regions was gel purified and left for spontaneous degradation at physiological conditions. (A) Illustrative 20% denaturating PAGE gel showing Ap-ICL degradation in time. P1, P2 and P3 are degradation products. (B) Degradation of Ap-ICL plotted over time and fitted with exponential decay. (C) The rate of Ap-ICL degradation plotted and represented as a bar chart. The half-lives for TT, TC, GT and GC are 60.6; 29.3; 31.7 and 28.8 days.

8 Discussion

8.1 Structural characterization of MutM glycosylase

Firstly, two structures from pathogenic *Neisseria meningitidis* were solved, one in unbound form (6TC6) and one in DNA-bound form (6TC9). Upon DNA binding, there is a notable alternation in the relative orientations of the NTD and CTD. Specifically, the NTD undergoes rotation towards the DNA, resulting in the characteristic bending of the DNA molecule. This conformational change is likely facilitated by NTD´s strong affinity for DNA, allowing the energetically demanding alteration to occur. To our best knowledge, the study presented the first example of procaryotic MutM in both free and DNA-bound form from the same organism. Furthermore, the MutM protein utilized in this research is a full-length wild-type chain, adding to the significance of the findings. When considering the previously solved crystal structures, and biophysical measurements of the MutM domain movements, our data provide substantial support for the dynamic nature of the MutM enzyme. Specifically, it undergoes distinct conformational changes upon DNA binding, highlighting the dynamic adaptable characteristics of the enzyme in response. The overall movement and binding mechanism observed in MutM closely resembles the action of molecular clamps or pliers. These mechanical analogies help illustrate how MutM functions, efficiently recognizing and processing its DNA substrate. Despite the changes induced by binding, it is intriguing to note that the structures of the individual domains in MutM remain relatively unchanged upon DNA binding. Instead, these domains undergo subtle rearrangements, allowing efficient substrate binding and processing. A similar movement with wider degree of rotation has previously been described in a related protein, Nei from *E. coli* (1K3W) [40]. So far, such movements have not been observed in the case of eucaryotic MutM orthologs from *A. thaliana* (3TWL, 3TWM) [39]. Untill now, only one structure of free MutM from thermophilic bacterium has been solved (1EE8). Unexpectedly, the conformation of the two domains of this holoenzyme was similar to DNA-bound form. Minor

structural changes were observed between monomers of the asymmetric unit. The flexible hinge-bending region between NTD and CTD enables them to adjust their relative position [29]. This flexibility within the protein structure could have functional implications, as it allows MutM to accommodate various DNA substrates and repair DNA lesions. To compare the different results regarding conformational changes in our study, the lack of conformational changes in *E. coli* study (1K3W), and the minor changes in the study of *A. thaliana* (3TWL, 3TWM), we have a few hypotheses.

Firstly, the movement of the NTD and CTD could be dependent on the organism from which the studied protein originated. The source organism in the studies described above could lead to discrepancies in published studies. Another suggestion is that domain organization of MutM is dynamic in solution, allowing the NTD and CTD to change along the hinge region.

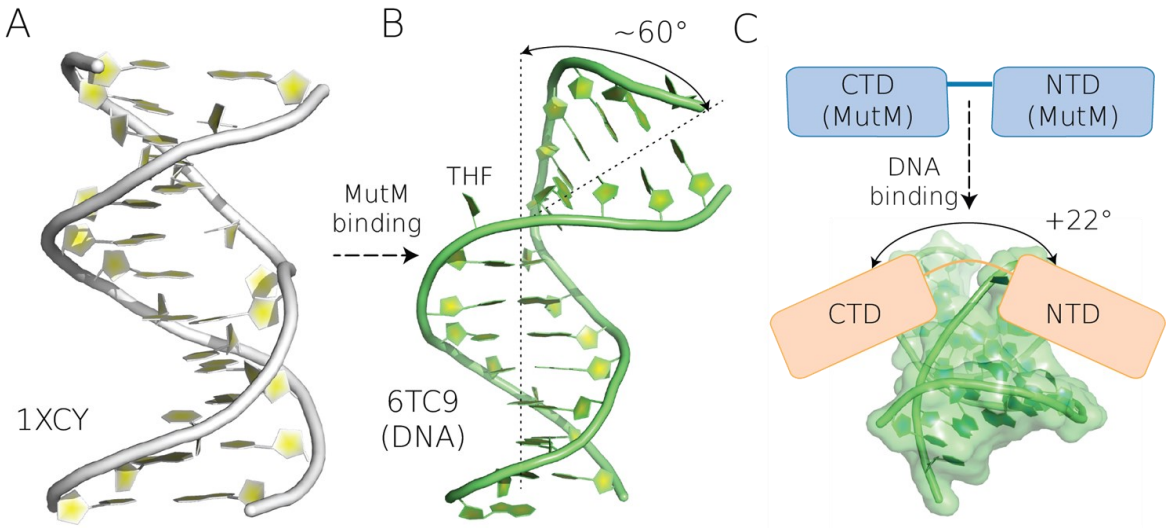

Figure 8.1. Schematic view of conformational changes in both DNA and MutM. (A) NMR structure of NDA (PDBID: 1XCY) with Ap site analogue α-hydroxy carbocyclic analogue of deoxyribose) does not show bends due to the presence of the mimic lesion. (B) DNA from DNA bound MutM complex form this study (PDBID:6TC9). MutM binding to DNA induces a DNA bend to an extent of 60°. At the same time the angle between the NTD and the CTD increases by 22°.

Originally, it was hypothesized that MutM, in contrast to Nei from *E. coli*, which shares only 21% similarity with *Neisseria* MutM, would not exhibit significant conformational changes upon binding to DNA. This hypothesis primarily relied on study showing crystal structure of apoform of MutM from thermophilic bacterium (PDBID:1EE8). The mechanism of this movement is similar to the much-accentuated domain movement observed in this study.

A comparison of the previously published MutM with DNA-bound MutM (PDBID:6TC9) from this study resulted in an RMSD = 1.661 Å for *A. Thaliana* (PDBID:3TWM), RMSD = 1.119 Å for thermophilic MutM (PDBID:1EE8), and RMSD = 0.908 Å for *E. coli* MutM (PDBID:1K82) (Figure 3.3.). These values indicate that the structures are very similar. In the contrary, when we compared them with the free MutM from this study (6TC6), those values increase dramatically (Table S.1.). As mentioned earlier, the increase in RMSD strongly depends on the rotation movement of NTDs and CTDs along the hinge region. We speculate that the differences in angles observed in this study compared to previous Fpg/Nei family structures result from crystal packing as an artifact, reflecting one of the possible conformations among all existing ones in solution. Furthermore, the difference between parted CTDs and NTDs among species are small, as these domains are structurally similar across species.

Both previously shown crystal structures and our structures show that both the open and closed conformations are supported, which is also confirmed by biophysical studies and NMR experiments for MutM from *E. coli* [43].

However, the “base flipping” mechanism remains unresolved, especially the energy source responsible for inducing DNA bending and flipping the damaged base towards the active site. The repairing process in biophysical studies requires no additional energy.

8.2 Formation and stability of Ap-ICL

Previous investigations of specific Ap-ICL formation have revealed two potential mechanisms for its occurrence [57,75]. These studies have presented evidence that within a duplex containing the sequence 5′-dC-Ap-dT, Ap-ICL can be formed through the crosslinking of Ap with either guanine or adenine in the complementary DNA strand. However, it was observed in subsequent study that an Ap-G crosslink is not inherently stable unless reduced with $NaBH_3CN$. This instability likely results from the steric unfavorable arrangement of the ICL, causing it to revert to the Ap site. Interestingly, the same sequence motif was used to determine NMR structure of Ap-ICL within a DNA duplex. In this study, only one type of ICL was observed, underscoring the significance of the 5′Ap-dT sequence as a "hotspot" for Ap-ICL formation in vivo. These findings underscore the intricate nature of Ap-ICL formation and highlight the importance of specific sequence motif and crosslink stability. Further research is necessary to gain a comprehensive understanding of the factors influencing Ap-ICL formation and its implications in DNA damage and repair processes.

A recent study shows another type of Ap-ICL, specifically involving the linking of the Ap site with the amino group of cytosine from the complementary DNA strand (Ap-C). This crosslink was observed to be preferentially formed in cases where cytosine was mispaired with adenine. Notably, this crosslink was found to occur under acidic conditions with a pH of 5, but it does not form under physiological conditions in non-mispaired DNA [97]. This new finding expands our understanding of Ap-ICL formation, highlighting the potential for different types of crosslinks depending on the specific mispairing of nucleotides and the surrounding conditions. Further exploration is warranted to elucidate the underlying mechanisms and physiological relevance of Ap-C crosslink formation, as it may provide insights into the repair processes and consequences of DNA damage associated with mispaired bases.

In our study, we focused on Ap-ICL formed within the 5′-Ap-dT sequence. We designed and have synthesized panel of DNA oligonucleotides with diverse DNA sequence surrounding the Ap site. Throughout the investigation, we focused on non-enzymatic Ap-ICL formation and aimed to elucidate the kinetics of this process by examining the influences of the surrounding DNA sequence. To initiate the generation of Ap-ICLs, we enzymatically converted uracil to the Ap site using UDG. Subsequently, we conducted reactions using individual oligonucleotides with varying sequences to measure the rates of Ap-ICL formation. The reactions were allowed to proceed for specific durations, after which they were terminated. The resulting reaction products were separated and analyzed using PAGE gels. By quantifying the fractions of crosslinks formed at each time point, we constructed a time-dependent profile, providing insights into the kinetics of Ap-ICL formation. This approach revealed that different nucleotides in the vicinity of Ap site influence the rate of Ap-ICL formation. While we observed some variations in the rates, one of the most notable findings from our study is the consistent formation of Ap-ICL across all tested DNA sequences. It is also remarkable the stability within all sequences, where half-life exceeds 30 days. This finding aligns with literature, which has also demonstrated the relative stability of different Ap-ICLs [96-98].

The data indicates that a single nucleotide near the Ap site only minimally affects the rate of Ap-ICL formation. The presence of AT-rich or GC-rich sequences around the Ap site significantly impacts ICL formation. Notably, AT-rich sequences exhibit considerably higher formation rates. We hypothesize that this phenomenon could be credited to the lower melting temperature of AT-rich regions, which facilitates structural fluctuations in DNA, known as “DNA breathing.” We speculate that weaker base-pairing in such regions may promote crosslinking reactions, resulting in faster and more frequent ICL formation. While our study focuses only on the *in vitro* studies, non-enzymatic formation of Ap-ICLs, it contributes to a broader understanding of the frequencies and conditions under which these crosslinks can be formed *in vivo*.

To enhance the importance of the study recently published paper showed reproducibility of our data and presented similar results in the Nucleic Acid Research journal [99].

10,000 Ap sites are generated through spontaneous processes such as depurination and deamination each day within a cell [10,11]. Furthermore, various DNA glycosylases contribute to the production of Ap sites during DNA damage repair processes. However, accurately quantifying this number can be a challenging task. Our rough estimate does not take into consideration the presence of DNA repair enzymes, which actively eliminate most of the Ap sites. Consequently, the observable Ap sites at any given moment may only represent a fraction ranging from 0.25% to 4% of the total Ap sites that could potentially lead to the Ap-ICL formation through the day [96,100-102].

Recent research involved estimating the relative number of Ap sites and comparing it with the knockdown of human Ap endonuclease. This study revealed that an additional 77% more Ap sites were detected in the knockdown scenario, when the removal the enzyme was 90% [103]. While it remains challenging to precisely determine the real number of Ap sites from the relative comparison, these findings strongly suggest that a substantial number of Ap sites persist within the genome.

Further methods for Ap site determination have been developed. One of them is chemical O-(pyridine-3-yl-methyl) hydroxylamine (PMOA) mass spectroscopy, which is a sensitive quantitative technique for directly detecting Ap sites within the DNA duplex [104]. This technique's sensitivity allows the identification of up to 2 Ap sites in 10^8 nucleotides, leading to an estimated count of approximately 2 Ap sites per 10^7 nucleotides in a steady-state scenario [105]. Considering the human genome size of 3.2×10^9 nucleotides, one human cell would have about 160 Ap sites. Most of the Ap sites are conjugated with an aldehyde reactive probe linked to biotin if sufficient time and reagents are supplied, as demonstrated in a biochemical study [106]. However, it is important to note that experiments involving aldehyde reactive probes could be biased, as these probes can interact with other aldehydes in DNA. While both Ap sites and sugars can undergo mutarotation, but usually only less than

1% of the sugars exist in the aldehyde form [70]. Therefore, the Ap site is determined by a ring-opening reaction leaving aldehyde form of the ribose, detectable by mass spectrometry with the PMOA reagent. PMOA studies use excessive quantity of PMOA, leading to first-order kinetics, where aldehyde form of Ap site is rate-limiting. Finally, the steady state value of Ap sites present in the genome is about 100-200 Ap sites. DNA repair mechanisms or DNA binding proteins could protect the genome by interfering with Ap-ICL formation [107,108]. NEIL3 glycosylase is responsible for repairing Ap-A ICL, while Ap endonuclease is responsible for repairing Ap-G ICL. These protective processes help explain the relatively low number of Ap sites detected by PMOA mass spectroscopy. Taking these predictions together with data from our study, we estimate that approximately 1-5 Ap-ICLs are formed or presented in a human cell.

9 Conclusion

One of the main goals of the thesis was to investigate the dependence of Ap-ICL formation on the nucleotide sequence in the vicinity of the rising damage, while also shedding light on the mechanism prior to the damage origin.

Our findings highlight the critical role of the neighboring sequence, particularly in arrangement mimicking AT or GC-rich regions, where we observed a notable preference for this lesion to manifest in the less tightly packed AT-rich regions. This preference indicates a spatial requirement for the formation of this type of damage. Moreover, we discovered that the base adjacent to the damaged site has a slide impact on the rate of Ap-ICL formation.

Our work has significantly contributed to understanding Ap-ICL formation, an area that has received relatively limited attention within the field of nucleic acid damage and repair. Our results, particularly regarding sequence dependency and the frequency of occurrence, offer valuable insights into the complexity of Ap-ICL formation.

Another aspect of our work is that we were the first to conduct a comprehensive comparison between the bound and unbound form of the MutM protein, both originating from the same organism by crystallization both holoenzyme and DNA-bound form. This analysis provide a valuable insight into structural changes that occur when MutM interacts with DNA and also how these changes are linked to its role in DNA repair.

10 Abbreviations

Ap	Apurinic/apyrimidinic, abasic
Ap-ICL	Abasic site interstrand crosslink
BER	base excision repair
CDC45	cell division cyclw 45
CMG	CDC45-MCM2-7-GINS
CTD	C-terminal domain
DNA	deoxyribonucleic acid
dsDNA	double-strand DNA
DTH	dihydrothimine
EcoNei	Endonuclease
FA	Fanconi anemia
Fapy	Formamidopyrimidine
FapyA	4,6-diamino-5-formamidopyrimidine
FapyG	2,6-diamino-4-hydroxy-5-formamidopyrimidine
Fpg	Formamidopyrimidine DNA glycosylase
GINS	go-ichi-ni-san
HEX	hexachlorofluorescein
HhH	Helix-Hairpin-Helix
HR	homologue recombination
H2TH	helix-two-turn-helix
ICL	interstrand crosslink
IR	infrared
MGM2-7	minichromosome maintenance 2-7
Nei	Endonuclease VIII
NEIL	Endonuclease VIII-like
NER	nucleotide excision repair
NMR	nuclear magnetic resonance

NTD	N-terminal domain
Ogg1	8-oxo-guanine DNA glycosylase
PMOA	O-(pyridine-3-yl-methyl) hydroxylamine
ssDNA	single-strand DNA
THF	tetrahydrofuran
UDG	Uracyl-DNA glycosylase
UV	ultraviolet
ZnF	Zinc finger
8-oxoG	8-oxoguanine

11 References

[1] D.E. Barnes, T. Lindahl, Repair and genetic consequences of endogenous DNA base damage in mammalian cells, Annual Review of Genetics 38 (2004) 445-476. 10.1146/annurev.genet.38.072902.092448.

[2] A. Tubbs, A. Nussenzweig, Endogenous DNA Damage as a Source of Genomic Instability in Cancer, Cell 168 (2017). 10.1016/j.cell.2017.01.002.

[3] T. Lindahl, INSTABILITY AND DECAY OF THE PRIMARY STRUCTURE OF DNA, Nature 362 (1993) 709-715. 10.1038/362709a0.

[4] J.A. Imlay, S. Linn, DNA DAMAGE AND OXYGEN RADICAL TOXICITY, Science 240 (1988) 1302-1309. 10.1126/science.3287616.

[5] S.S. Wallace, Base excision repair: A critical player in many games, DNA Repair 19 (2014) 14-26. 10.1016/j.dnarep.2014.03.030.

[6] S.D. McCulloch, T.A. Kunkel, The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases, Cell Research 18 (2008) 148-161. 10.1038/cr.2008.4.

[7] J.C. Fromme, G.L. Verdine, Base excision repair, in: W. Yang (Ed.) DNA Repair and Replication2004, pp. 1-41.

[8] A.L. Lu, X.H. Li, Y.S. Gu, P.M. Wright, D.Y. Chang, Repair of oxidative DNA damage - Mechanisms and functions, Cell Biochemistry and Biophysics 35 (2001) 141-170. 10.1385/cbb:35:2:141.

[9] S. Maynard, S.H. Schurman, C. Harboe, N.C. de Souza-Pinto, V.A. Bohr, Base excision repair of oxidative DNA damage and association with cancer and aging, Carcinogenesis 30 (2009) 2-10. 10.1093/carcin/bgn250.

[10] T. Lindahl, B. Nyberg, RATE OF DEPURINATION OF NATIVE DEOXYRIBONUCLEIC ACID, Biochemistry 11 (1972) 3610-&. 10.1021/bi00769a018.

[11] T. Lindahl, A. Andersson, RATE OF CHAIN BREAKAGE AT APURINIC SITES IN DOUBLE-STRANDED DEOXYRIBONUCLEIC ACID, Biochemistry 11 (1972) 3618-+. 10.1021/bi00769a019.

[12] C.D. Mol, A.S. Arvai, R.J. Sanderson, G. Slupphaug, B. Kavli, H.E. Krokan, D.W. Mosbaugh, J.A. Tainer, CRYSTAL-STRUCTURE OF HUMAN URACIL-DNA GLYCOSYLASE IN COMPLEX WITH A PROTEIN INHIBITOR - PROTEIN MIMICRY OF DNA, Cell 82 (1995) 701-708. 10.1016/0092-8674(95)90467-0.

[13] C. Rada, G.T. Williams, H. Nilsen, D.E. Barnes, T. Lindahl, M.S. Neuberger, Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice, Current Biology 12 (2002) 1748-1755 Pii s0960-9822(02)01215-0. 10.1016/s0960-9822(02)01215-0.

[14] A.L. Ross, J.E. Sale, The catalytic activity of REV1 is employed during immunoglobulin gene diversification in DT40, Molecular Immunology 43 (2006) 1587-1594. 10.1016/j.molimm.2005.09.017.

[15] E.J. Steele, Mechanism of somatic hypermutation: Critical analysis of strand biased mutation signatures at A:T and G:C base pairs, Molecular Immunology 46 (2009) 305-320. 10.1016/j.molimm.2008.10.021.

[16] D. Svilar, E.M. Goellner, K.H. Almeida, R.W. Sobol, Base Excision Repair and Lesion-Dependent Subpathways for Repair of Oxidative DNA Damage, Antioxidants & Redox Signaling 14 (2011) 2491-2507. 10.1089/ars.2010.3466.

[17] B. Demple, T. Herman, D.S. Chen, CLONING AND EXPRESSION OF APE, THE CDNA-ENCODING THE MAJOR HUMAN APURINIC ENDONUCLEASE - DEFINITION OF A FAMILY OF DNA-REPAIR ENZYMES, Proceedings of the National Academy of Sciences of the United States of America 88 (1991) 11450-11454. 10.1073/pnas.88.24.11450.

[18] P. Jaruga, M. Dizdaroglu, Repair of products of oxidative DNA base damage in human cells, Nucleic Acids Research 24 (1996) 1389-1394. 10.1093/nar/24.8.1389.

[19] M. Dizdaroglu, CHEMICAL CHARACTERIZATION OF OXIDATIVE DAMAGE TO DNA, Age 8 (1985) 83-83.

[20] T. Stevnsner, T. Thorslund, N.C. de Souza-Pinto, V.A. Bohr, Mitochondrial repair of 8-oxoguanine and changes with aging, Experimental Gerontology 37 (2002) 1189-1196 Pii s0531-5565(02)00142-0. 10.1016/s0531-5565(02)00142-0.

[21] T.R. Oconnor, S. Boiteux, J. Laval, RING-OPENED 7-METHYLGUANINE RESIDUES IN DNA ARE A BLOCK TO INVITRO DNA-SYNTHESIS, Nucleic Acids Research 16 (1988) 5879-5894. 10.1093/nar/16.13.5879.
[22] J. Laval, S. Boiteux, MISCODING LESIONS INDUCED IN DNA BY NITRITE AND THEIR REPAIR, Mutation Research 113 (1983) 274-274. 10.1016/0165-1161(83)90098-5.
[23] S. Boiteux, T.R. Oconnor, J. Laval, FORMAMIDOPYRIMIDINE-DNA GLYCOSYLASE OF ESCHERICHIA-COLI - CLONING AND SEQUENCING OF THE FPG STRUCTURAL GENE AND OVERPRODUCTION OF THE PROTEIN, Embo Journal 6 (1987) 3177-3183. 10.1002/j.1460-2075.1987.tb02629.x.
[24] D.Y. Jiang, Z. Hatahet, R.J. Melamede, Y.W. Kow, S.S. Wallace, Characterization of Escherichia coli endonuclease VIII, Journal of Biological Chemistry 272 (1997) 32230-32239. 10.1074/jbc.272.51.32230.
[25] M.L. Michaels, L. Pham, C. Cruz, J.H. Miller, MUTM, A PROTEIN THAT PREVENTS G.C- T.A TRANSVERSIONS, IS FORMAMIDOPYRIMIDINE-DNA GLYCOSYLASE, Nucleic Acids Research 19 (1991) 3629-3632. 10.1093/nar/19.13.3629.
[26] D.O. Zharkov, G. Shoham, A.P. Grollman, Structural characterization of the Fpg family of DNA glycosylases, DNA Repair 2 (2003) 839-862. 10.1016/s1568-7864(03)00084-3.
[27] J.C. Fromme, G.L. Verdine, DNA lesion recognition by the bacterial repair enzyme MutM, Journal of Biological Chemistry 278 (2003) 51543-51548. 10.1074/jbc.M307768200.
[28] J.C. Fromme, G.L. Verdine, Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM, Nature Structural Biology 9 (2002) 544-552. 10.1038/nsb809.
[29] M. Sugahara, T. Mikawa, T. Kumasaka, M. Yamamoto, R. Kato, K. Fukuyama, Y. Inoue, S. Kuramitsu, Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, Thermus thermophilus HB8, Embo Journal 19 (2000) 3857-3869. 10.1093/emboj/19.15.3857.

[30] R. Gilboa, D.O. Zharkov, G. Golan, A.S. Fernandes, S.E. Gerchman, E. Matz, J.H. Kycia, A.P. Grollman, G. Shoham, Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA, Journal of Biological Chemistry 277 (2002) 19811-19816. 10.1074/jbc.M202058200.

[31] F. Coste, M. Ober, T. Carell, S. Boiteux, C. Zelwer, B. Castaing, 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 (2004) 44074-44083. 10.1074/jbc.M405928200.

[32] Y. Qi, M.C. Spong, K. Nam, A. Banerjee, S. Jiralerspong, M. Karplus, G.L. Verdine, Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme, Nature 462 (2009) 762-U779. 10.1038/nature08561.

[33] Y. Qi, K. Nam, M.C. Spong, A. Banerjee, R.J. Sung, M. Zhang, M. Karplus, G.L. Verdine, Strandwise translocation of a DNA glycosylase on undamaged DNA, Proceedings of the National Academy of Sciences of the United States of America 109 (2012) 1086-1091. 10.1073/pnas.1111237108.

[34] K.P. de Jesus, L. Serre, C. Zelwer, B. Castaing, Structural insights into abasic site for Fpg specific binding and catalysis: comparative high-resolution crystallographic studies of Fpg bound to various models of abasic site analogues-containing DNA, Nucleic Acids Research 33 (2005) 5936-5944. 10.1093/nar/gki879.

[35] D.J. Hosfield, C.D. Mol, B.H. Shen, J.A. Tainer, Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: Coupling DNA and PCNA binding to FEN-1 activity, Cell 95 (1998) 135-146. 10.1016/s0092-8674(00)81789-4.

[36] M.M. Thayer, H. Ahern, D.X. Xing, R.P. Cunningham, J.A. Tainer, NOVEL DNA-BINDING MOTIFS IN THE DNA-REPAIR ENZYME ENDONUCLEASE-III CRYSTAL-STRUCTURE, Embo Journal 14 (1995) 4108-4120. 10.1002/j.1460-2075.1995.tb00083.x.

[37] X.G. Shao, N.V. Grishin, Common fold in helix-hairpin-helix proteins, Nucleic Acids Research 28 (2000) 2643-2650. 10.1093/nar/28.14.2643.

[38] J.C. Fromme, G.L. Verdine, Structure of a trapped endonuclease III-DNA covalent intermediate, Embo Journal 22 (2003) 3461-3471. 10.1093/emboj/cdg311.

[39] D.O. Zharkov, G. Golan, R. Gilboa, A.S. Fernandes, S.E. Gerchman, J.H. Kycia, R.A. Rieger, A.P. Grollman, G. Shoham, Structural analysis of an Escherichia coli endonuclease VIII covalent reaction intermediate, Embo Journal 21 (2002) 789-800. 10.1093/emboj/21.4.789.

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

[41] S. Duclos, P. Aller, P. Jaruga, M. Dizdaroglu, S.S. Wallace, S. Doublie, Structural and biochemical studies of a plant formamidopyrimidine-DNA glycosylase reveal why eukaryotic Fpg glycosylases do not excise 8-oxoguanine, DNA Repair 11 (2012) 714-725. 10.1016/j.dnarep.2012.06.004.

[42] O.S. Fedorova, G.A. Nevinsky, V.V. Koval, A.A. Ischenko, N.L. Vasilenko, K.T. Douglas, Stopped-flow kinetic studies of the interaction between Escherichia coli Fpg protein and DNA substrates, Biochemistry 41 (2002) 1520-1528. 10.1021/bi011524u.

[43] G.W. Buchko, K. McAteer, S.S. Wallace, M.A. Kennedy, Solution-state NMR investigation of DNA binding interactions in Escherichia coli formamidopyrimidine-DNA glycosylase (Fpg): a dynamic description of the DNA/protein interface, DNA Repair 4 (2005) 327-339. 10.1016/j.dnarep.2004.09.012.

[44] V.V. Koval, N.A. Kuznetsov, D.O. Zharkov, A.A. Ishchenko, K.T. Douglas, G.A. Nevinsky, O.S. Fedorova, Pre-steady-state kinetics shows differences in processing of various DNA lesions by Escherichia coli formamidopyrimidine-DNA glycosylase, Nucleic Acids Research 32 (2004) 926-935. 10.1093/nar/gkh237.

[45] M.M. Liu, K. Imamura, A.M. Averill, S.S. Wallace, S. Doublie, Structural Characterization of a Mouse Ortholog of Human NEIL3 with a Marked Preference for Single-Stranded DNA, Structure 21 (2013) 247-256. 10.1016/j.str.2012.12.008.

[46] J. Zhou, J. Chan, M. Lambele, T. Yusufzai, J. Stumpff, P.L. Opresko, M. Thali, S.S. Wallace, NEIL3 Repairs Telomere Damage during S Phase to Secure Chromosome Segregation at Mitosis, Cell Reports 20 (2017) 2044-2056. 10.1016/j.celrep.2017.08.020.

[47] D.R. Semlow, J.Q. Zhang, M. Budzowska, A.C. Drohat, J.C. Walter, Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase, Cell 167 (2016) 498-+. 10.1016/j.cell.2016.09.008.

[48] J.Q. Zhang, J.C. Walter, Mechanism and regulation of incisions during DNA interstrand cross-link repair, DNA Repair 19 (2014) 135-142. 10.1016/j.dnarep.2014.03.018.

[49] M. Raschle, P. Knipscheer, M. Enoiu, T. Angelov, J.C. Sun, J.D. Griffith, T.E. Ellenberger, O.D. Scharer, J.C. Walter, Mechanism of replication-coupled DNA interstrand crosslink repair, Cell 134 (2008) 969-980. 10.1016/j.cell.2008.08.030.

[50] R.A. Wu, D.R. Semlow, A.N. Kamimae-Lanning, O.V. Kochenova, G. Chistol, M.R. Hodskinson, R. Amunugama, J.L. Sparks, M. Wang, L. Deng, C.A. Mimoso, E. Low, K.J. Patel, J.C. Walter, TRAIP is a master regulator of DNA interstrand crosslink repair, Nature 567 (2019) 267-+. 10.1038/s41586-019-1002-0.

[51] M.M. Liu, V. Bandaru, J.P. Bond, P. Jaruga, X.B. Zhao, P.P. Christov, C.J. Burrows, C.J. Rizzo, M. Dizdaroglu, S.S. Wallace, The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo, Proceedings of the National Academy of Sciences of the United States of America 107 (2010) 4925-4930. 10.1073/pnas.0908307107.

[52] F.A. Ali, L. Renault, J. Gannon, H.L. Gahlon, A. Kotecha, J.C. Zhou, D. Rueda, A. Costa, Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate, Nature Communications 7 (2016) 10708. 10.1038/ncomms10708.

[53] Z.N. Yuan, L. Bai, J.C. Sun, R. Georgescu, J. Liu, M.E. O'Donnell, H.L. Li, Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation, Nature Structural & Molecular Biology 23 (2016) 217-224. 10.1038/nsmb.3170.

[54] M.I. Nejad, K. Housh, A.A. Rodriguez, T. Haldar, S. Kathe, S.S. Wallace, B.F. Eichman, K.S. Gates, Unhooking of an interstrand cross-link at DNA fork structures by the DNA glycosylase NEIL3, DNA Repair 86 (2020) 102752. 10.1016/j.dnarep.2019.102752.

[55] M.M. Liu, S. Doublie, S.S. Wallace, Neil3, the final frontier for the DNA glycosylases that recognize oxidative damage, Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis 743 (2013) 4-11. 10.1016/j.mrfmmm.2012.12.003.

[56] D.M. Noll, T.M. Mason, P.S. Miller, Formation and repair of interstrand cross-links in DNA, Chemical Reviews 106 (2006) 277-301. 10.1021/cr040478b.

[57] F. Coste, J.M. Malinge, L. Serre, W. Shepard, M. Roth, M. Leng, C. Zelwer, Crystal structure of a double-stranded DNA containing a cisplatin interstrand cross-link at 1.63 angstrom resolution: hydration at the platinated site, Nucleic Acids Research 27 (1999) 1837-1846. 10.1093/nar/27.8.1837.

[58] K.M. Johnson, N.E. Price, J. Wang, M.I. Fekry, S. Dutta, D.R. Seiner, Y.S. Wang, K.S. Gates, On the Formation and Properties of Interstrand DNA-DNA Cross-Links Forged by Reaction of an Abasic Site with the Opposing Guanine Residue of 5 '-CAp Sequences in Duplex DNA, Journal of the American Chemical Society 135 (2013) 1015-1025. 10.1021/ja308119q.

[59] J. Nakamura, M. Nakamura, DNA-protein crosslink formation by endogenous aldehydes and AP sites, DNA Repair 88 (2020) 102806. 10.1016/j.dnarep.2020.102806.

[60] X. Guo, M.I. Nejad, L.Q. Gu, K.S. Gates, Selective covalent capture of a DNA sequence corresponding to a cancer-driving C > G mutation in the KRAS gene by a chemically reactive probe: optimizing a cross-linking reaction with non-canonical duplex structures, Rsc Advances 9 (2019) 32804-32810. 10.1039/c9ra08009k.

[61] P.R. Martin, S. Couve, C. Zutterling, M.S. Albelazi, R. Groisman, B.T. Matkarimov, J.L. Parsons, R.H. Elder, M.K. Saparbaev, The Human DNA glycosylases NEIL1 and NEIL3 Excise Psoralen-Induced DNA-DNA Cross-Links in a Four-Stranded DNA Structure, Scientific Reports 7 (2017) 17438. 10.1038/s41598-017-17693-4.

[62] M. Bhagwat, J.A. Gerlt, 3'- and 5'-strand cleavage reactions catalyzed by the Fpg protein from Escherichia coli occur via successive beta- and delta-elimination mechanisms, respectively, Biochemistry 35 (1996) 659-665. 10.1021/bi9522662.

[63] H.M. Nash, S.D. Bruner, O.D. Scharer, T. Kawate, T.A. Addona, E. Sponner, W.S. Lane, G.L. Verdine, Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily, Current Biology 6 (1996) 968-980. 10.1016/s0960-9822(02)00641-3.

[64] B. Demple, L. Harrison, REPAIR OF OXIDATIVE DAMAGE TO DNA - ENZYMOLOGY AND BIOLOGY, Annual Review of Biochemistry 63 (1994) 915-948. 10.1146/annurev.biochem.63.1.915.

[65] V. Bailly, W.G. Verly, AP ENDONUCLEASES AND AP LYASES, Nucleic Acids Research 17 (1989) 3617-3618. 10.1093/nar/17.9.3617.

[66] C.C. Richardson, A. Kornberg, DEOXYRIBONUCLEIC ACID PHOSPHATASE-EXONUCLEASE FROM ESCHERICHIA COLI, Journal of Biological Chemistry 239 (1964) 242-&.

[67] L. Wiederhold, J.B. Leppard, P. Kedar, F. Karimi-Busheri, A. Rasouli-Nia, M. Weinfeld, A.E. Tomkinson, T. Izumi, R. Prasad, S.H. Wilson, S. Mitra, T.K. Hazra, AP endonuclease-independent DNA base excision repair in human cells, Molecular Cell 15 (2004) 209-220. 10.1016/j.molcel.2004.06.003.

[68] J. Silhan, Q. Zhao, E. Boura, H. Thomson, A. Forster, C.M. Tang, P.S. Freemont, G.S. Baldwin, Structural basis for recognition and repair of the 3 '-phosphate by NExo, a base excision DNA repair nuclease from Neisseria meningitidis, Nucleic Acids Research 46 (2018) 11980-11989. 10.1093/nar/gky934.

[69] A. Alvarez-Quilon, J.L. Wojtaszek, M.C. Mathieu, T. Patel, C.D. Appel, N. Hustedt, S.E. Rossi, B.D. Wallace, D. Setiaputra, S. Adam, Y. Ohashi, H. Melo, T.F. Cho, C. Gervais, I.M. Munoz, E. Grazzini, J.T.F. Young, J. Rouse, M. Zinda, R.S. Williams, D. Durocher, Endogenous DNA 3 ' Blocks Are Vulnerabilities for BRCA1 and BRCA2 Deficiency and Are Reversed by the APE2 Nuclease, Molecular Cell 78 (2020) 1152-+. 10.1016/j.molcel.2020.05.021.

[70] K.S. Ryu, C. Kim, I. Kim, S. Yoo, B.S. Choi, C. Park, NMR application probes a novel and ubiquitous family of enzymes that alter monosaccharide configuration, Journal of Biological Chemistry 279 (2004) 25544-25548. 10.1074/jbc.M402016200.

[71] C. de los Santos, M. El-khateeb, P. Rege, K. Tian, F. Johnson, Impact of the Cl ' configuration of abasic sites on DNA duplex structure, Biochemistry 43 (2004) 15349-15357. 10.1021/bi048400c.

[72] J.A. Wilde, P.H. Bolton, A. Mazumder, M. Manoharan, J.A. Gerlt, CHARACTERIZATION OF THE EQUILIBRATING FORMS OF THE ALDEHYDIC ABASIC SITE IN DUPLEX DNA BY O-17 NMR, Journal of the American Chemical Society 111 (1989) 1894-1896. 10.1021/ja00187a062.

[73] S. Dimitrov, Y. Koleva, T.W. Schultz, J.D. Walker, O. Mekenyan, Interspecies quantitative structure-activity relationship model for aldehydes: Aquatic toxicity, Environmental Toxicology and Chemistry 23 (2004) 463-470. 10.1897/02-579.

[74] H.S. Yu, T. Oyama, T. Isse, K. Kitagawa, T.T.P. Pham, M. Tanaka, T. Kawamoto, Formation of acetaldehyde-derived DNA adducts due to alcohol exposure, Chemico-Biological Interactions 188 (2010) 367-375. 10.1016/j.cbi.2010.08.005.

[75] I.G. Minko, I.D. Kozekov, T.A. Harris, C.J. Rizzo, R.S. Lloyd, M.P. Stone, Chemistry and Biology of DNA Containing 1,N-2-Deoxyguanosine Adducts of the alpha,beta-Unsaturated Aldehydes Acrolein, Crotonaldehyde, and 4-Hydroxynonenal, Chemical Research in Toxicology 22 (2009) 759-778. 10.1021/tx9000489.

[76] J. Burnotte, W.G. Verly, CROSSLINKING OF METHYLATED DNA BY MODERATE HEATING AT NEUTRAL PH, Biochimica Et Biophysica Acta 262 (1972) 449-&. 10.1016/0005-2787(72)90488-1.

[77] E. Freese, M. Cashel, CROSSLINKING OF DEOXYRIBONUCLEIC ACID BY EXPOSURE TO LOW PH, Biochimica Et Biophysica Acta 91 (1964) 67-&. 10.1016/0926-6550(64)90171-9.

[78] C. Goffin, W.G. Verly, INTERSTRAND DNA CROSSLINKS DUE TO AP (APURINIC APYRIMIDINIC) SITES, Febs Letters 161 (1983) 140-144. 10.1016/0014-5793(83)80747-9.

[79] A.S. Prakash, N.W. Gibson, SEQUENCE-SELECTIVE DEPURINATION, DNA INTERSTRAND CROSS-LINKING AND DNA STRAND BREAK FORMATION ASSOCIATED WITH ALKYLATED DNA, Carcinogenesis 13 (1992) 425-431. 10.1093/carcin/13.3.425.

[80] N.E. Price, K.M. Johnson, J. Wang, M.I. Fekry, Y.S. Wang, K.S. Gates, lnterstrand DNA-DNA Cross-Link Formation Between Adenine Residues and Abasic Sites in Duplex DNA, Journal of the American Chemical Society 136 (2014) 3483-3490. 10.1021/ja410969x.

[81] J.M. Dewar, J.C. Walter, Mechanisms of DNA replication termination, Nature Reviews Molecular Cell Biology 18 (2017) 507-516. 10.1038/nrm.2017.42.

[82] C. Feldman, R. Anderson, Meningococcal pneumonia: a review, Pneumonia 11 (2019) 3. 10.1186/s41479-019-0062-0.

[83] T. Davidsen, T. Tonjum, Meningococcal genome dynamics, Nature Reviews Microbiology 4 (2006) 11-22. 10.1038/nrmicro1324.

[84] T. Davidsen, E.K. Amundsen, E.A. Rodland, T. Tonjum, DNA repair profiles of disease-associated isolates of Neisseria meningitidis, Fems Immunology and Medical Microbiology 49 (2007) 243-251. 10.1111/j.1574-695X.2006.00195.x.

[85] T. Davidsen, H.K. Tuven, M. Bjoras, E.A. Rodland, T. Tonjum, Genetic interactions of DNA repair pathways in the pathogen Neisseria meningitidis, Journal of Bacteriology 189 (2007) 5728-5737. 10.1128/jb.00161-07.

[86] J. Silhan, K. Nagorska, Q.Y. Zhao, K. Jensen, P.S. Freemont, C.M. Tang, G.S. Baldwin, Specialization of an Exonuclease III family enzyme in the repair of 3' DNA lesions during base excision repair in the human pathogen Neisseria meningitidis, Nucleic Acids Research 40 (2012) 2065-2075. 10.1093/nar/gkr905.

[87] K. Nagorska, J. Silhan, Y.W. Li, V. Pelicic, P.S. Freemont, G.S. Baldwin, C.M. Tang, A network of enzymes involved in repair of oxidative DNA damage in Neisseria meningitidis, Molecular Microbiology 83 (2012) 1064-1079. 10.1111/j.1365-2958.2012.07989.x.

[88] C.C. Harper, J.M. Berg, S.J. Gould, PEX5 binds the PTS1 independently of Hsp70 and the peroxin PEX12, Journal of Biological Chemistry 278 (2003) 7897-7901. 10.1074/jbc.M206651200.

[89] W. Kabsch, XDS, Acta Crystallographica Section D-Biological Crystallography 66 (2010) 125-132. 10.1107/s0907444909047337.

[90] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallographica Section D-Biological Crystallography 66 (2010) 486-501. 10.1107/s0907444910007493.

[91] D. Liebschner, P.V. Afonine, M.L. Baker, G. Bunkoczi, V.B. Chen, T.I. Croll, B. Hintze, L.W. Hung, S. Jain, A.J. McCoy, N.W. Moriarty, R.D. Oeffner, B.K. Poon, M.G. Prisant, R.J. Read, J.S. Richardson, D.C. Richardson, M.D. Sammito, O.V. Sobolev, D.H. Stockwell, T.C. Terwilliger, A.G. Urzhumtsev, L.L. Videau, C.J. Williams, P.D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix, Acta Crystallographica Section D-Structural Biology 75 (2019) 861-877. 10.1107/s2059798319011471.

[92] A. Banerjee, W.L. Santos, G.L. Verdine, Structure of a DNA glycosylase searching for lesions, Science 311 (2006) 1153-1157. 10.1126/science.1120288.

[93] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J.E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nature Protocols 10 (2015) 845-858. 10.1038/nprot.2015.053.

[94] L. Serre, K.P. de Jesus, S. Boiteux, C. Zelwer, B. Castaing, Crystal structure of the Lactococcus lactis formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA, Embo Journal 21 (2002) 2854-2865. 10.1093/emboj/cdf304.

[95] S.R.W. Bellamy, G.S. Baldwin, A kinetic analysis of substrate recognition by uracil-DNA glycosylase from herpes simplex virus type 1, Nucleic Acids Research 29 (2001) 3857-3863. 10.1093/nar/29.18.3857.

[96] S.J. Admiraal, P.J. O'Brien, Base Excision Repair Enzymes Protect Abasic Sites in Duplex DNA from Interstrand Cross-Links, Biochemistry 54 (2015) 1849-1857. 10.1021/bi501491z.

[97] M.J. Catalano, S. Liu, N. Andersen, Z.Y. Yang, K.M. Johnson, N.E. Price, Y.S. Wang, K.S. Gates, Chemical Structure and Properties of Interstrand Cross-Links Formed by Reaction of Guanine Residues with Abasic Sites in Duplex DNA, Journal of the American Chemical Society 137 (2015) 3933-3945. 10.1021/jacs.5b00669.

[98] J.G. Varela, L.E. Pierce, X. Guo, N.E. Price, K.M. Johnson, Z.Y. Yang, Y.S. Wang, K.S. Gates, Interstrand Cross-Link Formation Involving Reaction of a

Mispaired Cytosine Residue with an Abasic Site in Duplex DNA, Chemical Research in Toxicology 34 (2021) 1124-1132. 10.1021/acs.chemrestox.1c00004.
[99] S.B.M. Amin, T. Islam, N.E. Price, A. Wallace, X. Guo, A. Gomina, M. Heidari, K.M. Johnson, C.D. Lewis, Z.Y. Yang, K.S. Gates, Effects of Local Sequence, Reaction Conditions, and Various Additives on the Formation and Stability of Interstrand Cross-Links Derived from the Reaction of an Abasic Site with an Adenine Residue in Duplex DNA, Acs Omega 7 (2022) 36888-36901. 10.1021/acsomega.2c05736.
[100] P.D. Chastain, J. Nakamura, S.B. Rao, H.T. Chu, J.G. Ibrahim, J.A. Swenberg, D.G. Kaufman, Abasic sites preferentially form at regions undergoing DNA replication, Faseb Journal 24 (2010) 3674-3680. 10.1096/fj.09-145276.
[101] S. Boiteux, F. Coste, B. Castaing, Repair of 8-oxo-7,8-dihydroguanine in prokaryotic and eukaryotic cells: Properties and biological roles of the Fpg and OGG1 DNA N-glycosylases, Free Radical Biology and Medicine 107 (2017) 179-201. 10.1016/j.freeradbiomed.2016.11.042.
[102] K.N. Mohni, S.R. Wessel, R.X. Zhao, A.C. Wojciechowski, J.W. Luzwick, H. Layden, B.F. Eichman, P.S. Thompson, K.P.M. Mehta, D. Cortez, HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA, Cell 176 (2019) 144-+. 10.1016/j.cell.2018.10.055.
[103] N. Wang, H.Y. Bao, L. Chen, Y.H. Liu, Y. Li, B.X. Wu, H.D. Huang, Molecular basis of abasic site sensing in single-stranded DNA by the SRAP domain of E. coli yedK, Nucleic Acids Research 47 (2019) 10388-10399. 10.1093/nar/gkz744.
[104] K.P. Roberts, J.A. Sobrino, J. Payton, L.B. Mason, R.J. Turesky, Determination of apurinic/apyrimidinic lesions in DNA with high-performance liquid chromatography and tandem mass spectrometry, Chemical Research in Toxicology 19 (2006) 300-309. 10.1021/tx0502589.
[105] J.H. Guo, H.Q. Chen, P. Upadhyaya, Y. Zhao, R.J. Turesky, S.S. Hecht, Mass Spectrometric Quantitation of Apurinic/Apyrimidinic Sites in Tissue DNA of Rats Exposed to Tobacco-Specific Nitrosamines and in Lung and Leukocyte DNA of Cigarette Smokers and Nonsmokers, Chemical Research in Toxicology 33 (2020) 2475-2486. 10.1021/acs.chemrestox.0c00265.

[106] S.E. Bennett, J. Kitner, Characterization of the aldehyde reactive probe reaction with AP-sites in DNA: Influence of AP-lyase on adduct stability, Nucleosides Nucleotides & Nucleic Acids 25 (2006) 823-842. 10.1080/15257770600726133.
[107] B.F. Eichman, P.S. Thompson, K.M. Amidon, K.N. Mohni, D. Cortez, Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA crosslink, Acta Crystallographica a-Foundation and Advances 75 (2019) A207-A207. 10.1107/s0108767319097988.
[108] L.T. Da, J. Yu, Base-flipping dynamics from an intrahelical to an extrahelical state exerted by thymine DNA glycosylase during DNA repair process, Nucleic Acids Research 46 (2018) 5410-5425. 10.1093/nar/gky386.

12 Supplementary

Supplementary table 1. Structural alignments of individual MutM

PDB	MutM	M-DNA	NTD6	CTD6	NTD9	CTD9	I.(%)	Organism	Interesting feature
1K82	2.598	0.908	0.654	0.841	0.526	0.544	54.4	*E.coli*	DNA bound
1EE8	3.133	1.119	1.061	0.968	1.134	0.749	42.6	*Thermus thermophilus*	Free MutM
3TWL	2.779	1.527	1.389	1.194	1.461	1.14	27.7	*A. thaliana*	DNA bound
3TWK	2.713	1.634	1.484	1.361	1.585	1.297	27.1	*A. thaliana*	Free MutM
1K3W	3.290	2.275	2.106	1.12	1.952	1.02	25.7	*E.coli*	DNA NEI
1Q3C	6.889	9.142	2.451	1.225	2.474	1.102	25.4	*E.coli*	Free NEI
6TC9	2.408		0.562		0.562		100	*N. meningitidis*	this study
6TC6		2.408		0.649		0.649	100	*N. meningitidis*	this study

RMSD values in Å were computed for the structural alignment of individual PDB structures. The alignments were performed between the free MutM structure obtained in this study (6TC6) and DNA-bound MutM (6TC9). In these calculations, both the NTD and CTD regions of the protein were considered. “6” and “9” serve as the identifier for free and DNA-bound structures. The RMDS value determination is described in materials and methods. I (%) indicated sequence identity when compared to the meningococcal MutM. The last two rows present a comparison of 6TC6 with 6TC9.

FEBS Letters

FEBS PRESS science publishing by scientists

Conformational changes of DNA repair glycosylase MutM triggered by DNA binding

Barbora Landová and Jan Šilhán

Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, Czech Republic

Correspondence
J. Šilhán, Institute of Organic Chemistry and Biochemistry, Flemingovo náměstí 542/2, 160 00 Praha 6, Czech Republic
Tel: (+420) 220 183 477
E-mail: silhan@uochb.cas.cz

(Received 1 February 2020, revised 28 May 2020, accepted 23 June 2020)

doi:10.1002/1873-3468.13876

Edited by Christian Griesinger

Bacterial MutM is a DNA repair glycosylase removing DNA damage generated from oxidative stress and, therefore, preventing mutations and genomic instability. MutM belongs to the Fpg/Nei family of prokaryotic enzymes sharing structural and functional similarities with their eukaryotic counterparts, for example, NEIL1–NEIL3. Here, we present two crystal structures of MutM from pathogenic *Neisseria meningitidis*: a MutM holoenzyme and MutM bound to DNA. The free enzyme exists in an open conformation, while upon binding to DNA, both the enzyme and DNA undergo substantial structural changes and domain rearrangement. Our data show that not only NEI glycosylases but also the MutMs undergo dramatic conformational changes. Moreover, crystallographic data support the previously published observations that MutM enzymes are rather flexible and dynamic molecules.

Keywords: base excision DNA repair; DNA repair; Fpg/Nei; MutM; *Neisseria meningitidis*

DNA inevitably undergoes spontaneous chemical reactions leading to DNA damage and loss of genome integrity. Further on, highly reactive endogenous and endogenous compounds constantly attack susceptible groups on the DNA molecule. Oxidative stress results in a vast number of DNA lesions. These can generate mutations that may result in genetic instability, aging, and cancer. To counteract these potentially harmful effects, organisms have evolved various kinds of DNA repair mechanisms. The first line of defense is base excision DNA repair where the damaged base is removed by a DNA glycosylase forming an abasic site (AP-site, from apurinic/apyrimidinic). This lesion is subsequently excised by AP endonuclease resulting in a single-stranded DNA break. It is further filled by DNA repair polymerase and ligated by DNA ligase [1–4].

There are various sources of oxidative DNA damage in the cell involving respiration, inflammation, metal catalysis, and irradiation. Although the sources for single cellular and multicellular organisms may differ, the main agents responsible for the majority of oxidized DNA lesions are thought to be superoxide anion radical and mainly hydroxyl radical [5–7]. Both purine and pyrimidine nucleotides can be damaged leading to the formation of more than 100 different DNA lesions. Among the most common pyrimidine lesions are cytosine and thymine radicals, which are relatively unstable and undergo chemical changes leading to thymine glycol and cytosine glycol. The latter generates 5-hydroxycytosine, or it undergoes rapid deamination to uracil glycol and subsequent formation of 5-hydroxyuracil [8]. One of the most common and thoroughly studied purine lesions is 8-oxoguanine (8-oxoG). Subsequent chemical lesion leads to 2,6-diamino-4-hydroxy-5-formamidopyridine (FapyG), and a similar fate awaits oxidized adenine leading to 4,6-diamino-5-formamidopyrimidine (FapyA) [3,8,9]. One of the key enzymes

Abbreviations
8-oxoG, 8-oxoguanine; CTD, C-terminal domain; dsDNA, double-stranded DNA; H2TH, helix-two-turn-helix; NTD, N-terminal domain; ssDNA, single-stranded DNA; THF, tetrahydrofuran.

removing all of these aforementioned lesions is a MutM DNA glycosylase belonging to the Nei/Fpg class. MutM is also called Fpg for its ability to repair Fapy lesions (formamidopyridine DNA glycosylase) [10]. These enzymes are conserved among prokaryotes, and their homologues are found often in multiple paralogues in all studied eukaryotic species. The non-prokaryotic homologs have been described as specialized enzymes with unique substrate specificities and therefore fulfilling relatively distinct DNA repair roles [11].

MutM is a bifunctional glycosylase with its glycosylase activity removing the base and β,δ-lyase activities processing the AP site. The molecular mechanism of the reaction starts with a nucleophilic attack of the secondary amine group of Pro2 on C1′ of the ribose of the damaged site so that the damaged base is cleaved off. Pro2 is part of a conserved catalytic motif at the N terminus of MutM proteins with consensus PEL-PEV, where the third and sixth amino acid residues may vary (Fig. 1B) [12]. The aforementioned reaction leads to a covalent reaction intermediate connecting MutM with ribosyl residue in DNA by via a Schiff base. This relatively stable intermediate is subsequently excised by the associated lyase activity. Further on, the enzyme is regenerated by the β,δ-elimination ribose at the 3′ and 5′ ends liberating the 4-hydroxypenta-2,4-diene as a by-product [13]. The MutM-catalyzed reaction generates a DNA gap with another lesion, a toxic 3′-phosphate which blocks the subsequent DNA repair polymerization step. In bacteria, the 3′-phosphate is commonly removed by *Xth* family of AP-endonucleases [14,15]. Overall, bacterial MutMs have a broad range of substrate specificity, recognizing not only both damaged pyrimidines and purines but also AP-sites.

In the last two decades, a vast number of crystal structures of Fpg/Nei glycosylases have been reported (see Table S3 for some examples). The first described the overall architecture of the free enzyme (PDB 1EE8, [13], and in the next, the structure of DNA-bound MutM from *Bacillus stearothermophilus* was revealed (PDBID: 1L1T, 1L1Z, 1L2B, 1L2C, 1L2D, [12,16]. The other studies have been focused on the substrate search, its recognition, preference, and the importance of orphaned base to the lesion and the details of the reaction mechanisms [12,13,17–21]. MutM and other members of Fpg/Nei family share similar domain organization and secondary structure with the two main domains divided by a flexible region acting as a hinge [13]. Overall, these studies have described the architecture of holoenzymes and the DNA-bound structures of MutM enzymes. In greater detail, the N-terminal domain (NTD) is composed of eight β-sandwiched sheets flanked with two adjacent α-helices from either side. The C-terminal domain (CTD) contains four α-helices and two antiparallel β-strands forming a Zn-finger characteristic for this family of proteins. The latter two of the four α-helices contain conserved helix-two-turn-helix (H2TH) motif [13]. H2TH along with helix-hairpin-helix and helix turn helix can be found across many different DNA binding proteins including DNA glycosylases, transcription factors, and nucleases [22–24]. The Zn-finger has been reported to be essential for DNA binding of Fpg/Nei family of proteins although it does not seem to contact the DNA directly and some eukaryotic homologues of NEIL1 contain a 'Zincless' fold in this part of the C terminus of the protein [26,27]. Typically, the enzymes from Fpg/Nei family induce a strong aprox. 45–60° bend to its substrate DNA (e.g., [16,21,27].

Only little is known about the dynamics in the solution and conformational changes induced to MutM during the process of DNA binding. There is an example of a comparative study of a MutM paralogue EcoNei from *Escherichia coli* with both structures of free and DNA-bound enzymes (PDBID: 1Q3C, 1K3W). It has been shown that EcoNei undergoes dramatic interdomain conformational changes upon DNA binding, where both NTD and CTD rotate toward the DNA molecule by an angle of ~ 50° [28,29]. A similar structural comparison was carried out for the eukaryotic MutM homologue from *Arabidopsis* (PDBID: 3TWL). This is a more distant MutM homologue and contains a 'Zincless' domain instead of the canonical Zn-finger, and it is unable to remove 8-oxoG. Most importantly, there were no significant structural changes observed between free and DNA-bound structures, in contrast to the *E. coli* Nei enzyme [30]. Unlike the case of Nei, the crystallographic studies of MutM do not suggest dramatic domain rearrangement, dynamic behavior or conformational changes triggered upon DNA binding. On the other hand, minute but noticeable domain rotations along the flexible hinge region were observed between the monomers in the asymmetric unit for thermophilic MutM (PDBID: 1EE8) [13]. Importantly, biophysical measurements on MutM from *E. coli* suggested the dynamic character of the enzyme with an increase in rigidity during DNA binding [31–33].

Neisseria meningitidis is a human pathogenic bacterium responsible for bacterial meningitis, septicemia and meningococcal pneumonia and is characterized by a relatively high fatality rate, but its natural habitat is

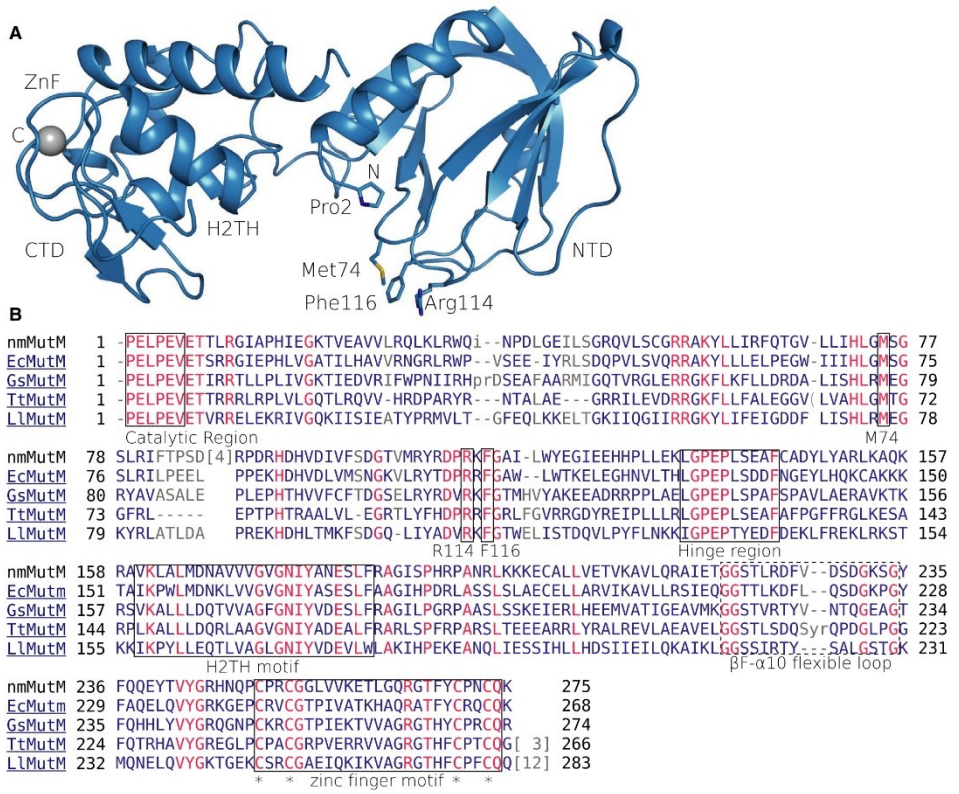

Fig. 1. Structure of meningococcal MutM (PDBID: 6TC6). (A) Crystal structure of MutM enzyme from *Neisseria meningitidis*. NTD and CTD are the N- and C-terminal domains of the protein. Within the NTD catalytic region with Pro2 and the nucleophilic amine group, Met74, Arg114, and Phe116 are residues important for interaction with the minor groove of DNA. CTD consists of four helices where the latter two are part of the conserved H2TH domain. At the C terminus of CTD dwells zinc finger domain (ZnF), (B) sequence alignment of MutM proteins from different prokaryotic species for which the crystal structure has been determined (PDBID: 1K82, 1LTT, 1EE8, 1KFV) [12,13,16,50]. Identical residues are labeled in red while other residues are blue and the least conserved regions with deletions and loops labeled in grey. Important regions are highlighted by boxes and annotated. The dashed line outlines a flexible region (βF-α10 loop) not observed in most crystal structures responsible recognition of the substrate base [18].

restricted to the vicinity of the human larynx [34,35]. Due to its pathogenic character and restricted habitat, its small genome contains only a handful of DNA repair genes with one Fpg/Nei glycosylase [36,37]. Besides the studying of this important human pathogen, we utilized the limited repertoire of DNA repair enzymes in this bacterium to reveal the network of enzymatic pathways and molecular mechanisms of its DNA repair enzymes [38,39]. Therefore, *Neisseria* offers us an excellent model of a minimal DNA repair system.

In this work, we seek to understand the extent and underlying principles of conformational changes triggered by DNA binding in bacterial MutMs. We have expressed, purified, and crystallized full-length wild-type MutM and solved structures of the free and the DNA-bound forms of this enzyme. Here, we reveal differences in global and local conformational changes triggered by DNA binding and point out differences and similarities with homologous Fpg/Nei proteins from other organisms.

Experimental procedures

Protein expression and purification

Full-length *MutM* untagged gene from *Neisseria meningitides* (strain MC58, UniProt P55044) was expressed and purified as described previously [38]. In brief, protein was expressed in *E. coli* BL21 Star (DE3) or RosettaTM(DE3) pLys strains. Bacteria were grown in LB medium at 37 °C at 220 r.p.m. until they reached an O.D. (600 nm) = 0.6, and then, the temperature was lowered to 20 °C and the protein production was induced by addition of 0.5 mM IPTG. The suspension was incubated for further 12–16 h. The cells were pelleted by centrifugation at 6000 ***g*** for 6 min (JLA 9.1 centrifuge). Pellets were frozen at −20 °C, thawed, resuspended in 1× PBS supplemented with 2 mM bME and 10% glycerol, and processed by sonication. Sonicated suspension was clarified by centrifugation at 35 000 ***g***, and clarified lysates were loaded on the tandem anion/cation exchange chromatography with 5 mL HiTrap Q HP and 5ml HiTrap Sp HP (GE Healthcare). The running buffer was 20 mM Tris/HCl pH = 7.4, 125 mM NaCl, 2 mM βME, 10% glycerol. After the load of the lysate, the Q HP column was removed from the tandem and the Sp HP column was thoroughly washed. MutM protein was eluted with the gradient of an elution buffer: 20 mM Tris/HCl pH = 7.4, 1 M NaCl, 2 mM βME, 10% glycerol. The peak of eluted MutM protein was observed at conductivity 25 $mS{\cdot}cm^{-1}$. Finally, MutM was loaded on Superdex 200 pg, XK 16/60 (GE Healthcare) equilibrated in 20 mM Tris/HCl pH = 7.4, 500 mM NaCl, 2 mM βME, 10% glycerol. Protein was concentrated to 8 $mg{\cdot}mL^{-1}$ and flash-frozen for further use. The identity of the protein was validated using mass spectrometry (data not shown).

DNA binding gel-based assay; fluorescent anisotropy DNA binding assay

5′-HEX (hexachloro-fluorescein) labeled and unlabeled DNA oligonucleotides containing tetrahydrofuran (THF) analogue of AP site were custom synthesized, and the sequences are listed in Table S2 (Sigma-Aldrich). For all binding assays, single-stranded (ssDNA) and double-stranded (dsDNA) substrates containing either THF or not were resuspended in reaction buffer: 20 mM Tris pH 7.4, 120 mM NaCl, 0.5 mM TCEP, 5% glycerol. TE buffer: 10 mM Tris/Hcl pH = 8.0, 1 mM EDTA. Oligonucleotide sequences are listed in Table S2. For the preparation of dsDNA substrates, the complementary oligos were mixed in equimolar ration and annealed by heating to 95 °C for 2 min and allowed to cool gradually to 25 °C. Fifty nanomolar of each of the DNA substrates [namely ssTHF, ss no THF (ssDNA), ds THF, dsDNA (ds noTHF)] was titrated with pure MutM protein from concentration 0 to 2000 nM and subsequently separated on 15% native PAGE and visualized with TYPHOON FLA 7000 fluorescence imager (GE Healthcare). Gels were only qualitative control of DNA binding. No loading dye was used for native PAGE.

The fluorescence anisotropy measurements were used to determine K_D constants in identical reaction buffer with the same DNA substrates. Similarly, as above constant concentration of HEX labeled DNA was 50 nM was diluted in reaction buffer. DNA was titrated with meningococcal MutM. Set-up of wavelengths on fluorescence spectrometer FluoroMax-4, HORIBA Scientific was 535 nm absorption max and 556 nm emission max. Quantum yield (Q) represents the ratio of total intensities (I_{tot}) of free and bound phluorophore. Total intensity increases during titration and was calculated using following formula (1)

$$I_{tot} = I_{vv} + 2{*}G{*}I_{vh} \tag{1}$$

where G is the G-factor which depends on the optical system of the instrument. It is calculated as the ratio between horizontal and vertical intensity (values are listed in Table S4). The fraction of DNA bind (fB) was determined using formula (2)

$$fB = \frac{(A_{obs} - A_{min})}{(A_{max} - A_{obs}){*}Q + (A_{obs} - A_{min})} \tag{2}$$

where A_{obs} is the observed anisotropy for given MutM concentration, A_{min} is the minimum observed anisotropy and A_{max} is the anisotropy at saturation [40,41]. Standard errors were calculated from four independent measurements. Dissociation constants were calculated as described previously [42,43], using formula (3)

$$fB = (1/2[\mathrm{DNA}])((K_D + [\mathrm{DNA}] + [\mathrm{MutM}]) - \sqrt{\left(K_D + [\mathrm{DNA}] + [\mathrm{MutM}]^2\right) - 4[\mathrm{DNA}][\mathrm{MutM}]}) \tag{3}$$

where K_D is dissociation constant and the values in brackets, for example, [DNA], are concentrations at a given point of the titration. The nonlinear fit of the data was performed using GRAPHPAD PRISM (GraphPad Software).

Crystallization, data collection and analysis

The MutM crystallization trials were carried out using the sitting drop vapor diffusion method and set up on Mosquito® crystallization robot (TTPlabtech, Melbourn, UK). Three commercial crystallization screens JCSG Core I Suite, ProPlex, and Morpheus were used (Molecular Dimensions Limited, Sheffield, UK). The crystals appeared in drops containing 150 nL of protein solution and 250 nL of reservoir and grew within 48 h at 18 °C in the condition composed of 100 mM Bicine pH 8.5, 20% (w/v) PEG 6000, final pH 9.0. Crystals were cryoprotected using original well solution supplemented with 20% glycerol and frozen in liquid nitrogen. The frozen crystals were tested for diffraction on the home source (Cu–Kα). Dataset resulting in the final structure (PDBID: 6TC9) was collected at synchrotron beamline MX 14.1 at BESSYII Helmholtz–Zentrum Berlin, Germany.

To obtain the crystal structure of MutM with DNA (PDBID: 6TC9). DNA oligos, one containing synthetic a AP-site analogue (THF) (TGTCCA[THF] GTCTACC) and complementary oligo (AGGTA GACCTGGAC) (IDT, Leuven, Belgium), were resuspended in 20 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA, mixed, and annealed by heating to 98 °C for 5 min and allowed to cool to room temperature in a metal block. MutM was mixed with DNA duplex in 1 : 1.5 molar ration and desalted into crystallization buffer 10 mM Tris 8.0, 50 mM NaCl using HiTrap Desalting column (GE Healthcare). This complex was subsequently loaded on Superdex 200 XK16/60 prep grade (GE Healthcare) in 10 mM Tris/HCl pH = 8.0 and 50 mM NaCl, 1 mM EDTA, 1mM βME. Peak fractions were analyzed on 15% SDS/PAGE gel stained with Coomassie® Brilliant Blue R-250 (Sigma-Aldrich; Fig. S1). The presence of DNA in complex with MutM was confirmed on 15% Native PAGE (TBE gel) stained with SYBR® Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific, Waltham, MA, USA). Subsequently, this native gel was additionally stained with Coomassie® Brilliant blue R-250. The peak fraction containing both the protein and the DNA was pooled and concentrated (200 μM). The crystals were grown utilizing the sitting drop vapor diffusion method at 18 °C. Three hundred nanoliter of complex was mixed with 300 nL reservoir and set up by Mosquito crystallization robot. The reservoir condition for final crystals was in 100 mM Tris/HCl, 200 mM magnesium acetate, 15–19% PEG 6000. The crystals were briefly soaked in reservoir solution supplemented with 20% glycerol and immediately frozen in liquid nitrogen. Data were collected at the Diamond Light Source, Didcot, UK. Several datasets were collected from these crystals with identical cell parameters. Datasets were indexed with XDS, and two of them with the best parameters were combined using XSCALE and finally merged, scaled, and converted to mtz files using XDSCONV within the XDS package [44].

The crystal structure of MutM holoenzyme (PDBID: 6TC6) was determined by molecular replacement in Phaser and using the model generated in Phyre2 generated and based on MutM structure from *E. coli* (PDBID: 1K82). Subsequently, the crystal structure of DNA-bound meningococcal MutM (PDBID: 6TC9) was solved by molecular replacement in Phaser using the structure of MutM holoenzyme (PDBID: 6TC6) as the model. In both cases, the flexible domains (CTD and NTD) were separated and individual solutions for each domain were found in Phaser. The domains were joined together by manual rebuilding of the flexible linker in COOT, and the refinement was performed in Phenix Refine [45,46]. The figures representing the protein structures were generated by PYMOL 1.7.X (https://pymol.org).

Structural alignments

All structures used for structural alignment were superimposed using align command in PYMOL software. Unbound MutM (PDBID: 6TC6) was aligned to DNA-bound MutM (PDBID: 6TC9) using only CTD (Ser141 - Lys275) containing clearly defined residues of H2TH (helix-two-turns-helix) and helix turn helix and zinc finger located at C terminus. The outliers were rejected from final superimposition yielding the final RMSD values; for example, 704 atoms were used from CTD.

Results

Crystallization and the determination of the structure of meningococcal MutM

First, we have solved the crystal structure of meningococcal MutM enzyme from *N. meningitidis*. We obtained crystals that diffracted to 2.9 Å, and we solved the structure by molecular replacement using a homology model generated by the Phyre2 web server

where the structure of MutM from the *E. coli* enzyme (54% sequence identity) was used (PDBID: 1K82) [47,48]. This homologous model was based on the protein bound to DNA, and after molecular replacement with Phaser, we obtained poor electron density maps [46]. However, we could obtain a good solution by splitting the model into N- and C-terminal subdomains that were used as two independent search models for molecular replacement, indicating that the NTD and CTD subdomains are oriented differently in the free and the DNA-bound structures (Fig. S3) [16]. The model was refined to the final structure of enzyme of meningococcal MutM at a resolution of 2.9 Å (Fig. 1). The region Thr217–Tyr242 that is usually only ordered if a flipped-out base is present was poorly visible in the electron density and was omitted from the model.

Overall, the crystal structure of the full-length MutM from *N. meningitidis* shares the fold with its homologues from other organisms. The NTD is composed of one large and one small α-helix flanking the opposite sides of the β-stranded sandwich composed of two halves where each has four antiparallel β-sheets. The CTD is linked to NTD with a stretch of the coiled-coil region (Pro130–Phe144) with well-defined electron density and encompasses the aforementioned flexible hinge region. CTD contains four helices where the latter two contain a H2TH (helix-two-turns-helix) motif, followed by a zinc finger domain ([Cys]4) (Fig. 1) [13].

Meningococcal MutM binds DNA with the synthetic AP-site analogue THF with nanomolar affinity

To obtain a crystal structure of DNA-bound MutM, we carried out DNA binding experiments using a 15 base pair double-stranded DNA containing either the base C or THF (THF instead of the ribose, i.e., 1′,2′-dideoxyribofuranose-5′-phosphate) at position (GC AGCAC(THF/C)GACCACG/CGTGGTCGGTGCT GC). The THF is a frequently used analogue of an abasic site. We compared the binding of single-stranded and double-stranded oligonucleotides. Our DNA binding experiments revealed strong binding in the low nanomolar range even in the case of 'undamaged' ssDNA (27 nM, Fig. 2A,B), which bound with the weakest affinity. All the other constructs show K_D values between 7 and 9 nM.

MutM bound to DNA with abasic site analogue THF

The crystal structure of 14 base pair DNA with a THF (THF instead of the ribose, i.e., 1′,2′-dideoxyribofuranose-5′-phosphate) bound to MutM was solved by molecular replacement using the structure of enzyme MutM from this study (PDBID: 6TC6) as a search model. For successful molecular replacement, the model had to be split yet again to NTD and CTD. Each domain was placed separately using Phaser [46]. Final model of DNA-bound MutM was refined to 2.15 Å in Coot (Fig. 2 [45]). Data collection statistics for both crystal structures are listed in Table S1. The MutM binding to the DNA molecules induces a DNA twist of ~ 60° (Fig. 4) compared to canonical B-DNA. This conformational change in the DNA is not due to the presence of the THF since unbound DNA containing similar abasic sites exhibits a normal B-DNA structure (Fig. 4). More specifically, the THF of the DNA is exposed and rotated toward the active site with the catalytic Pro2 as it would be in case of the base being 'flipped out'. Since the substrate is AP site analogue, this MutM:DNA structure does not harbor any damaged base. Binding to the damaged base is initiated by the nucleophilic attack of the side chain Pro2 nitrogen at the C1′ of sugar of the damaged base. The ring opening intermediate is stabilized by protonation by Glu3 at the C5′-OH. Lys54 protonates the 3′-phosphate which leads to β-elimination [29]. This is consistent with previous structures of MutM:AP-DNA structures where a damaged base was not required for the 'base-flipping' mechanism to take place [12,49]. Concurrent with the DNA binding, a 'void space' in this part of the DNA molecule is created between the orphaned base and the THF residue. Three residues Met74, Arg114, and Phe116 are well conserved within the Fpg/Nei family of proteins (Fig. 2C). They protrude into the space between the orphaned base and the widened minor groove DNA substrate. Phe116 is stacked between the orphaned base and the adjacent base pair. Arg114 protrudes into the major groove of the DNA by hydrogen bonding with the orphaned cytosine. Finally, Met74 moves inwards to the damage site (THF site) opposite of catalytic Pro2.

The conformational changes in MutM induced by DNA binding

At first glance at both structures, it becomes obvious that their domains undergo substantial reorganization or conformational changes upon DNA binding (Fig. 3). To describe these molecular movements in greater details, the least-square superimpositions of whole proteins and individual domains were carried out (PYMOL software). The r.m.s.d of the entire protein chain was relatively high (RMSD = 2.41 Å), and

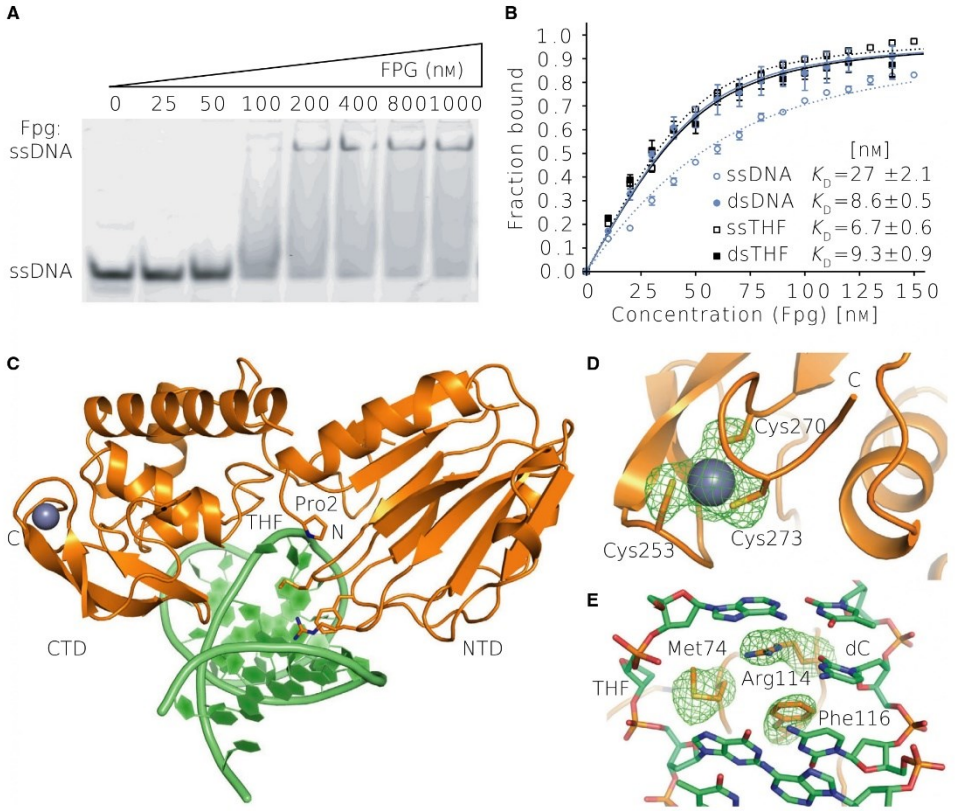

Fig. 2. DNA binding properties and structure of DNA-bound meningococcal MutM (PDBID: 6TC9). DNA containing THF were used to mimic an abasic (i.e., 'damaged') site in the DNA substrates.(A) Gel-based DNA binding assay (EMSA electrophoretic mobility shift assay) to qualitatively access DNA binding of MutM protein, HEX (hexachloro-fluorescein)-labeled THF-containing ssDNA 50 nM oligonucleotide was incubated with MutM and loaded on 15% native PAGE, (B) fluorescent anisotropy binding assay was used to compare the affinities to ssDNA and ds DNA substrates with and without the abasic site mimick THF. Blue represents 'undamaged' and black THF containing DNA. Dissociation constants (KD) were fitted to the data using formula (3). Error bars are shown as standard errors of four independent titrations, (C) crystal structure of meningococcal MutM (orange) bound to dsDNA substrate (green) containing THF (AP site analogue), (D) detail of omit map of Zn[Cys]4 where Cys250 is hidden behind, the omit map (fo-fc) of electron density (green) is shown at sigma level = 3.0, (E) view toward the opening in the minor groove of DNA where the DNA backbone containing the abasic site mimick THF is flipped out and the place of the missing base is filled by the chains of Met74, Arg114, and Phe116. Arg114 hydrogen bonds with the orphaned cytosine (dC). The omit map (fo-fc) of electron density (green) is shown at sigma level = 2.0.

alignment leads to poor overlap of the protein chains. While the alignment of individual domains revealed that the CTD and NTD domains rotate toward DNA by angle of 22.30° (Fig. 3, Fig. S2), the superimposition of NTDs reveals a relative displacement of the zinc atom from the finger domains is 11.4 Å (Fig. S2A).

In summary, a relatively high overall RMSD, while the alignment of the separate domains separately RMDS of the CTDs was only 0.649 Å and the RMSD of NTDs was only 0.562 Å and, therefore, the domains retain their overall architecture and mainly undergo rotation and translation upon the DNA binding (Fig. S2 and Table S3).

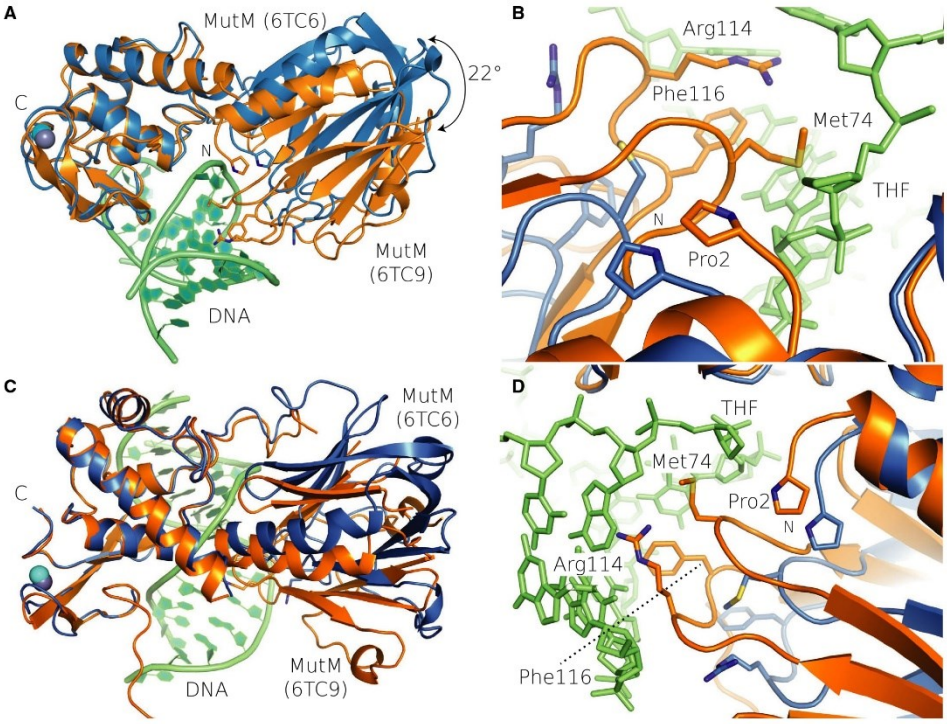

Fig. 3. Conformation changes in meningococcal MutM upon DNA binding. Structural alignments of CTDs of free and DNA-bound MutMs (shown in blue and orange resp.) (A–D), identical view as in Figs 1A and 2C depicts the rotational movement of ~ 22° in panel A with the top view on panel C, (B, D) views from different angles showing detail of residues Met74, Arg114, and Phe116 and their 'movement' toward bound DNA. For clarity, DNA is shown in green (See Fig. S1, for structures aligned to their NTDs). Finally, Met74 moves inwards to theTHF site, opposite of catalytic Pro2 (Fig. 2C). All three residues Met74, Arg114, and Phe116 seem to be stabilizing and thus anchoring the protein in this substrate recognition state.

Comparison with other systems and proteins

Currently, more than 30 different MutM structures from at least 5 different organisms have been solved (https://rcsb.org/). This includes distant and aforementioned homologue from *Arabidopsis thaliana* (PDBID: 3TWL, 3TWK, 3TWM). The RMSD of DNA-bound MutM from *E. coli* and DNA-bound MutM from this study (PDBID: 6TC9) is relatively low (RMSD = 0.908 Å) as opposed to that of free meningococcal MutM (PDBID: 6TC6) (RMSD = 2.598 Å). Unsurprisingly, the structural alignments of individual domains from meningococcal MutM structures with MutM structures from other organisms reveal a great degree of similarity in both NTDs and CTDs (Table S3, Fig. S3). These alignments of these individual domains separately have much greater structural similarities than when MutM structures are aligned with their entire protein chains (Table S3). Besides of Nei enzyme from *E. coli* (PDBID: 1K3W, 1Q3C), the most outstanding differences in all MutM structures belongs to meningococcal MutM from this study (PDBID: 6TC6). These significant changes arise mainly from the aforementioned rotational movement of NTD and CTD around flexible hinge region, rather than due to their sequence difference (Data S3, Table S3). Therefore, the overall conformation of DNA-bound MutM (PDBID: 6TC9) resembles overlaps with the majority of MutM structures. Also, the conformation and the bend of substrate DNA follows the same trend (Fig. S3).

Discussion

In this work, we have solved two structures of MutM from pathogenic bacterium *N. meningitidis* in free (6CT6) and DNA-bound forms (6TC9). Upon the binding to DNA, the relative orientations of the NTD and CTD significantly change when the NTD rotates toward DNA and concomitantly induces the typical bend of the DNA, presumably by exploiting the extremely high affinity to the DNA to allow for the energetically costly conformation change (Figs 3 and 4). To our best knowledge, this is the first example of prokaryotic MutM structures with and without DNA from the same organism. Moreover, the MutM protein used in this study is a full-length wild-type chain. Together with previously solved crystal structures and biophysical measurements of MutM domain movements, our data support a dynamic character MutM enzyme where it changes into more distinctive conformation upon the DNA binding.

The overall movement and binding mechanism of the substrate resembles the action of molecular clamps or pliers. Since the structure of individual domains remains practically unchanged upon the DNA binding as they are only rearranged into new positions (Figs 3 and 4). A similar movement with an even greater degree of rotation has been previously described in a related enzyme in the case of *E. coli* Nei-1K3W [28]. No such movement has been observed in eukaryotic MutM orthologue from *A. thaliana* between the free and DNA-bound forms (PDBID: 3TWL and 3TWM) [30]. As of yet, the only one crystal structure of prokaryotic MutM enzyme without DNA (from a thermophilic bacterium, PDBID 1EE8) has been solved. Surprisingly, it also followed the domain conformation of MutM bound to DNA. It is noteworthy that minor interdomain structural changes have been observed between the two monomers of the asymmetric unit, suggesting domains might undergo movement along flexible hinge-bending region connecting NTD and CTD [13]. We speculate that the conformational changes observed in our study compared to the lack of these changes in the *Arabidopsis* homologue (PDBID: 3TWL, 3TWM) or the only minor hinge-bending observed in the *Thermophilus*–MutM (PDBID: 1EE8) could arise due to the following hypotheses.

The first is that the observed angles between NTD and CTD of MutMs depend on the source organism. In other words, that differences between species drive the discrepancies in published structures. While the other suggests that the domain organization of unliganded MutM is dynamic in solution and changes along the hinge region. Therefore, each MutM structure might be locked within a discrete energetic minimum and therefore may represent crystallographic snapshots of distinct conformations (Fig. S3).

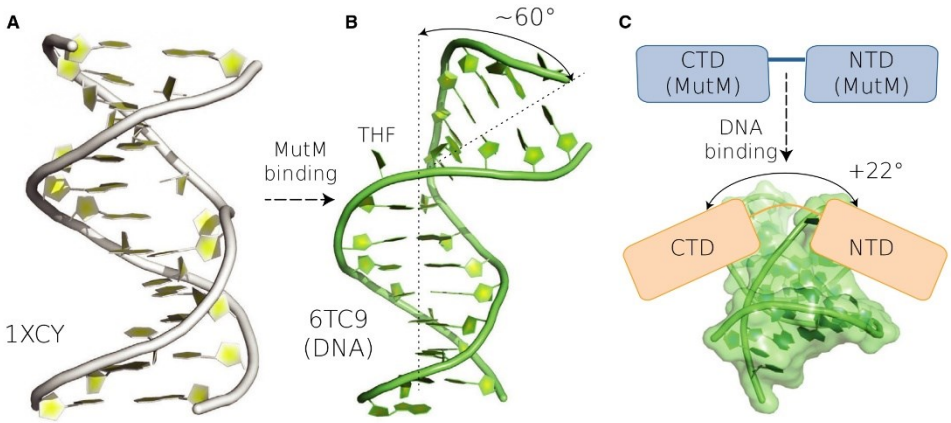

Fig. 4. Schematics of mutual conformational changes in DNA and MutM. (A) NMR structure of DNA with α-carba AP site analogue (α-hydroxy carbocyclic analogue of deoxyribose) (PDBID: 1XCY) does not display dramatic bends due to the presence of the lesion (50), (B) DNA from the MutM:DNA complex from this study (PDBID: 6TC9), where MutM binding induces a DNA bend, typical for Fpg/Nei family (in extend of ~ 60°), in the same time the angle between the CTD and NTD domain increases by 22° in the same direction but perpendicular to the DNA bend.

Initially, it has been suggested that unlike Nei from *E. coli* (with 21% sequence identity to the Neisseria MutM), MutM (Fpg) enzymes do not undergo substantial conformational changes upon DNA binding [12]. This prediction was fundamentally based on the crystal structure of the aforementioned thermophilic MutM holoenzyme (PDBID: 1EE8) [13]. Nonetheless, the mechanism of this movement is in principle similar to the much-accentuated interdomain movement observed in this study. When we compare published structures of MutM with MutM bound to DNA from this study (PDBID: 6TC9), the resulting RMSD for thermophilic MutM (PDBID: 1EE8) is 1.119 Å, while for *E. coli*-1K82 is 0.908 Å and for *A. thaliana* (PDBID: 3TWM) RMSD = 1.661 Å (See Table S3 for more complete comparison). This demonstrates that these structures are very similar. In contrast, when we do this comparison with the free MutM enzyme (PDBID: 6TC6), these differences increase dramatically (Table S3). As aforementioned, the most of these differences (the increase of RMSD) arise from substantial rotation movement of NTDs and CTDs along the hinge region of free MutM (PDBID: 6TC6). We speculate that different configurations of CTD and NTD from meningococcal MutM enzyme (PDBID: 6TC6) as well as slight different angles observed in other structures of Fpg/Nei family could be a result of crystal packing. We speculate that this is a crystallization artifact only reflecting one of the possible conformations in solution.

Additionally, the structural differences among various species are even smaller when comparing NTD and CTD individually (Table S3). Taking this into account and also the possible bias arriving from the crystallographic influence of the domain orientation, the first hypothesis is deemed unlikely as the domains are structurally quite similar among various species (Fig. S3).

Besides these data that crystal structures support both the open and the closed conformation, additional biophysical studies support this hypothesis. A high degree of backbone movements and dynamic character of the protein backbone were observed during the NMR experiments of the MutM enzyme from *E. coli* [31]. The additional studies supported this observation by measurements of the mobility by tryptophan fluorescence in MutM (*E. coli*), where the rigidity was observed to increase on DNA-bound complex [32,33,50]. These studies with our work bring compelling evidence of dynamic domain movements around 'hinge' region in MutM.

One further question still remains unresolved, and it is the mechanism of 'base flipping'. The driving force and the energy responsible for the induction of these changes including the DNA bending and twisting of the damaged site toward the active site pocket is still not fully understood. Remarkably, the process of repair does not require additional energy. Intriguingly, either the damaged or the undamaged DNA molecules are positioned almost in a completely identical manner when bound to MutM from *Geobacillus stearothermophilus* (PDBID: 3GQ5) [20]. DNA containing only the AP site or THF is positioned in conformation almost identical to that of 'flipped out base', and DNA backbone is distorted in a very similar extent regardless of the presence of damaged nucleotide. The question is where does the driving force for 'flipping out-like mechanism' come from. It is well established that the damage itself is not a sufficient determinant for proper DNA bending as the free DNA molecule containing damaged base (e.g., 8-oxoG, AP site analogue) does not display any dramatic conformational changes or bends (Fig. 4) [51–53]. In a study of free DNA, it was shown that DNA bending, deformation and slight unwinding may be facilitating the flipping mechanism and proper accommodation of the DNA substrate (8-oxoG) in the active site [54]. We speculate that the additional required energy may arise from the interactions of DNA and MutM (Fig. 4) as K_D of DNA is in nanomolar range (Fig. 2B). DNA binding to MutM induces harsh changes bending, unwinding, and widening the minor groove of the DNA molecule.

Acknowledgements

The work was supported by the Czech Science Foundation (17-21649Y), and we thank Academy of Sciences Czech Republic for support in form of J.E. Purkyne Fellowship awarded to JS. The project was also supported by the Academy of Sciences Czech Republic (RVO: 61388963). I would like thank Geoff S. Baldwin for kindly donating expression clones and sharing some unpublished data with us.

Author contributions

JS and BL designed the experiments, BL conducted biophysical experiments, crystallized and solved structure of MutM. JS crystallized and solved the structure of MutM:DNA. JS and BL wrote the manuscript.

References

1 Barnes DE and Lindahl T (2004) Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu Rev Genet* **38**, 445–476.

2 Fromme JC and Verdine GL (2004) Base excision repair. *Adv Protein Chem* **69**, 1–41.
3 Lu A-L, Li X, Gu Y, Wright PM and Chang D-Y (2001) Repair of oxidative DNA damage. *Cell Biochem Biophys* **35**, 141–170.
4 Wallace SS (2014) Base excision repair: a critical player in many games. *DNA Repair (Amst)* **19**, 14–26.
5 Imlay J and Linn S. (1988). Damage and oxygen radical. *Science (80.)* **240**, 1302–1309.
6 Keyer K, Gort AS and Imlay JA (1995) Superoxide and the production of oxidative DNA damage. *J Bacteriol* **177**, 6782–6790.
7 Téoule R (1987) Radiation-induced DNA damage and its repair. *Int J Radiat Biol* **51**, 573–589.
8 Jaruga P and Dizdaroglu M (1996) Repair of products of oxidative DNA base damage in human cells. *Nucleic Acids Res* **24**, 1389–1394.
9 Dizdaroglu M (1985) Formation of an 8-hydroxyguanine moiety in deoxyribonucleic acid on γ-irradiation in aqueous solution†. *Biochemistry* **24**, 4476–4481.
10 Boiteux S, O'Connor TR and Laval J (1987) Formamidopyrimidine-DNA glycosylase of *Escherichia coli*: cloning and sequencing of the fpg structural gene and overproduction of the protein. *EMBO J* **6**, 3177–3183.
11 Prakash A, Doublié S and Wallace SS (2012) The Fpg/Nei family of DNA glycosylases: substrates, structures, and search for damage. *Prog Mol Biol Transl Sci* **110**, 71–91.
12 Fromme JC and Verdine GL (2002) Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nat Struct Biol* **9**, 544–552.
13 Sugahara M, Mikawa T, Kumasaka T, Yamamoto M, Kato R, Fukuyama K, Inoue Y and Kuramitsu S (2000) Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, *Thermus thermophilus* HB8. *EMBO J* **19**, 3857–3869.
14 Richardson CC and Kornberg A (1964) A deoxyribonucleic acid phosphatase-exonuclease from *Escherichia coli* : I. Purification of the enzyme and characterization of the phosphatase activity. *J Biol Chem* **239**, 242–250.
15 Silhan J, Zhao Q, Boura E, Thomson H, Förster A, Tang CM, Freemont PS and Baldwin GS (2018) Structural basis for recognition and repair of the 3′-phosphate by NExo, a base excision DNA repair nuclease from *Neisseria meningitidis*. *Nucleic Acids Res* **46**, 11980–11989.
16 Gilboa R, Zharkov DO, Golan G, Fernandes AS, Gerchman SE, Matz E, Kycia JH, Grollman AP and Shoham G (2002) Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *J Biol Chem* **277**, 19811–19816.
17 Coste F, Ober M, Carell T, Boiteux S, Zelwer C and Castaing B (2004) Structural basis for the recognition of the FapydG lesion (2,6-diamino-4-hydroxy-5-formamidopyrimidine) by formamidopyrimidine-DNA glycosylase. *J Biol Chem* **279**, 44074–44083.
18 Fromme JC and Verdine GL (2003a) DNA lesion recognition by the bacterial repair enzyme MutM. *J Biol Chem* **278**, 51543–51548.
19 Pereira de Jésus K, Serre L, Zelwer C and Castaing B (2005) Structural insights into abasic site for Fpg specific binding and catalysis: comparative high-resolution crystallographic studies of Fpg bound to various models of abasic site analogues-containing DNA. *Nucleic Acids Res* **33**, 5936–5944.
20 Qi Y, Spong MC, Nam K, Banerjee A, Jiralerspong S, Karplus M and Verdine GL (2009) Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature* **462**, 762–766.
21 Qi Y, Nam K, Spong MC, Banerjee A, Sung RJ, Zhang M, Karplus M and Verdine GL (2012) Strandwise translocation of a DNA glycosylase on undamaged DNA. *Proc Natl Acad Sci USA* **109**, 1086–1091.
22 Hosfield DJ, Mol CD, Shen B and Tainer JA (1998) Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell* **95**, 135–146.
23 Shao X (2000) Common fold in helix-hairpin-helix proteins. *Nucleic Acids Res* **28**, 2643–2650.
24 Thayer MM, Ahern H, Xing D, Cunningham RP and Tainer JA (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *EMBO J* **14**, 4108–4120.
25 Doublié S, Bandaru V, Bond JP and Wallace SS (2004) The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. *Proc Natl Acad Sci USA* **101**, 10284–10289.
26 Zharkov DO, Shoham G and Grollman AP (2003) Structural characterization of the Fpg family of DNA glycosylases. *DNA Repair (Amst)* **2**, 839–862.
27 Fromme JC and Verdine GL (2003b) Structure of a trapped endonuclease III-DNA covalent intermediate. *EMBO J* **22**, 3461–3471.
28 Golan G, Zharkov DO, Feinberg H, Fernandes AS, Zaika EI, Kycia JH, Grollman AP and Shoham G (2005) Structure of the uncomplexed DNA repair enzyme endonuclease VIII indicates significant interdomain flexibility. *Nucleic Acids Res.* **33**, 5006–5016.

29 Zharkov DO (2002) Structural analysis of an *Escherichia coli* endonuclease VIII covalent reaction intermediate. *EMBO J* **21**, 789–800.
30 Duclos S, Aller P, Jaruga P, Dizdaroglu M, Wallace SS and Doublié S (2012) Structural and biochemical studies of a plant formamidopyrimidine-DNA glycosylase reveal why eukaryotic Fpg glycosylases do not excise 8-oxoguanine. *DNA Repair (Amst)* **11**, 714–725.
31 Buchko GW, McAteer K, Wallace SS and Kennedy MA (2005) Solution-state NMR investigation of DNA binding interactions in *Escherichia coli* formamidopyrimidine-DNA glycosylase (Fpg): a dynamic description of the DNA/protein interface. *DNA Repair (Amst)* **4**, 327–339.
32 Fedorova OS, Nevinsky GA, Koval VV, Ishchenko AA, Vasilenko NL and Douglas KT (2002) Stopped-flow kinetic studies of the interaction between *Escherichia coli* Fpg protein and DNA substrates. *Biochemistry* **41**, 1520–1528.
33 Koval VV, Kuznetsov NA, Zharkov DO, Ishchenko AA, Douglas KT, Nevinsky GA and Fedorova OS (2004) Pre-steady-state kinetics shows differences in processing of various DNA lesions by *Escherichia coli* formamidopyrimidine-DNA glycosylase. *Nucleic Acids Res* **32**, 926–935.
34 Davidsen T and Tønjum T (2006) Meningococcal genome dynamics. *Nat Rev Microbiol* **4**, 11–22.
35 Feldman C and Anderson R (2019) Meningococcal pneumonia: a review. *Pneumonia* **11**, 1–13.
36 Davidsen T, Tuven HK, Bjørås M, Rødland E andTønjum T (2007a) Genetic interactions of DNA repair pathways in the pathogen *Neisseria meningitidis*. *J Bacteriol* **189**, 5728–5737.
37 Davidsen T, Amundsen EK, Rødland EA and Tønjum T (2007b) DNA repair profiles of disease-associated isolates of *Neisseria meningitidis*. *FEMS Immunol Med Microbiol* **49**, 243–251.
38 Nagorska K, Silhan J, Li Y, Pelicic V, Freemont PS, Baldwin GS and Tang CM (2012a) A network of enzymes involved in repair of oxidative DNA damage in *Neisseria meningitidis*. *Mol Microbiol* **83**, 1064–1079.
39 Silhan J, Nagorska K, Zhao Q, Jensen K, Freemont PS, Tang CM and Baldwin GS (2012) Specialization of an Exonuclease III family enzyme in the repair of 3′ DNA lesions during base excision repair in the human pathogen *Neisseria meningitidis*. *Nucleic Acids Res* **40**, 2065–2075.
40 Harper CC, Berg JM and Gould SJ (2003) PEX5 binds the PTS1 independently of Hsp70 and the peroxin PEX12. *J Biol Chem* **278**, 7897–7901.
41 Obsilova V, Herman P, Vecer J, Sulc M, Teisinger J and Obsil T (2004) 14-3-3ζ C-terminal stretch changes its conformation upon ligand binding and phosphorylation at Thr232. *J Biol Chem* **279**, 4531–4540.
42 Boura E, Silhan J, Herman P, Vecer J, Sulc M, Teisinger J, Obsilova V and Obsil T (2007) Both the N-terminal loop and wing W2 of the forkhead domain of transcription factor Foxo4 are important for DNA binding. *J Biol Chem* **282**, 8265–8275.
43 Silhan J, Obsilova V, Vecer J, Herman P, Sulc M, Teisinger J and Obsil T (2004) 14-3-3 protein C-terminal stretch occupies ligand binding groove and is displaced by phosphopeptide binding. *J Biol Chem* **279**, 49113–49119.
44 Kabsch W (2010). XDS. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 125–132.
45 Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 486–501.
46 Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr Sect D Struct Biol* **75**, 861–877.
47 Banerjee A, Santos WL and Verdine GL (2006) Structure of a DNA glycosylase searching for lesions. *Science (80-.)* **311**, 1153–1157.
48 Kelley LA, Mezulis S, Yates CM, Wass MN and Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845–858.
49 Serre L, Pereira de Jésus K, Boiteux S, Zelwer C and Castaing B (2002) Crystal structure of the *Lactococcus lactis* formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA. *EMBO J* **21**, 2854–2865.
50 Koval VV, Kuznetsov NA, Ishchenko AA, Saparbaev MK and Fedorova OS (2010) Real-time studies of conformational dynamics of the repair enzyme *E. coli* formamidopyrimidine-DNA glycosylase and its DNA complexes during catalytic cycle. *Mutat Res Fundam Mol Mech Mutagen* **685**, 3–10.
51 Lipscomb LA, Peek ME, Morningstar ML, Verghis SM, Miller EM, Rich A, Essigmann JM and Williams LD (1995) X-ray structure of a DNA decamer containing 7,8-dihydro-8-oxoguanine. *Proc Natl Acad Sci USA* **92**, 719–723.
52 De Los Santos C, El-khateeb M, Rege P, Tian K and Johnson F (2004) Impact of the C1′ configuration of abasic sites on DNA duplex structure. *Biochemistry* **43**, 15349–15357.
53 Singh SK, Szulik MW, Ganguly M, Khutsishvili I, Stone MP, Marky LA and Gold B (2011) Characterization of DNA with an 8-oxoguanine modification. *Nucleic Acids Res* **39**, 6789–6801.

54 La Rosa G and Zacharias M (2016) Global deformation facilitates flipping of damaged 8-oxo-guanine and guanine in DNA. *Nucleic Acids Res* **44**, 9591–9599.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Gel filtration of MutM:DNA (THF) for crystallography.

Fig. S2. Structural superimposition of NTD domains free MutM and MutM:DNA.

Fig. S3. Structural superimposition of solved structures from different organisms with MutM and MutM:DNA from this study.

Table S1. Data collection and refinement statistics.

Table S2. DNA oligos for binding assays and crystallography.

Table S3. Structural alignments of MutMs from five different organisms.

Table S4. Quantum yields and *G*-factors from fluorescent anisotropy DNA binding assay.

DNA Repair 113 (2022) 103300

Contents lists available at ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

The rate of formation and stability of abasic site interstrand crosslinks in the DNA duplex

Andrea Huskova [1], Barbora Landova [1], Evzen Boura, Jan Silhan [*]

Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo namesti 2, 166 10 Prague 6, Czech Republic

ARTICLE INFO

Keywords:
DNA damage
Interstrand crosslink
Abasic site
Non-enzyme kinetics
DNA repair

ABSTRACT

DNA interstrand crosslinks (ICLs) strands pose an impenetrable barrier for DNA replication. Different ICLs are known to recruit distinct DNA repair pathways. NEIL3 glycosylase has been known to remove an abasic (Ap) site derived DNA crosslink (Ap-ICL). An Ap-ICL forms spontaneously from the Ap site with an adjacent adenine in the opposite strand. Lack of genetic models and a poor understanding of the fate of these lesions leads to many questions about the occurrence and the toxicity of Ap-ICL in cells.

Here, we investigate the circumstances of Ap-ICL formation. With an array of different oligos, we have investigated the rates of formation, the yields, and the stability of Ap-ICL. Our findings point out how different bases in the vicinity of the Ap site change crosslink formation *in vitro*. We reveal that AT-rich rather than GC-rich regions in the surrounding Ap site lead to higher rates of Ap-ICL formation. Overall, our data reveal that Ap-ICL can be formed in virtually any DNA sequence context surrounding a hot spot of a 5′-Ap-dT pair, albeit with significantly different rates and yields. Based on Ap-ICL formation *in vitro*, we attempt to predict the number of Ap-ICLs in the cell.

1. Introduction

Interstrand DNA crosslinks (ICLs) are a type of DNA damage in which opposite DNA strands are covalently linked. ICLs form an impenetrable barrier that prevents DNA replication [1]. Various forms of ICLs that differ both structurally and chemically have been identified [2–5]. To combat the deleterious and toxic consequences of these lesions, distinct DNA repair mechanisms have evolved. At present, two distinct DNA repair processes have been identified. The first repair process is the NEIL3 pathway involving NEIL3 glycosylase, and the second is the Fanconi Anaemia DNA repair pathway (FA pathway) [6,7]. These two pathways are intrinsically linked to DNA replication and complement each other in the repair of ICLs. The NEIL3 pathway has been linked to an early stage of replication coupled repair when replication forks stall on the crosslink and is known for the removal of abasic site ICLs (Ap-ICLs) [4,8]. Additionally, NEIL3 is also involved in the recognition and excision of psoralen-induced ICLs. The same ability has also another glycosylase, NEIL1. Importantly, the excision activity of these glycosylases doesn't generate double-strand breaks [9]. The later recruited FA pathway recognises and repairs crosslinks arising from small aldehydes such as acetaldehyde derived crosslinks. It has been identified and it is known for its ability to repair DNA crosslinks generated from exogenous crosslinking compounds, used as chemotherapeutics or in warfare (*e.g.* cis-platin and nitrogen mustard) [10–12].

An abasic site (apurinic/apyrimidinic), or an Ap site, can be generated spontaneously in large quantities, as it was estimated that 10,000 Ap sites form per cell every day [13,14]. This astronomical number is increased even further, mainly by some DNA glycosylases that remove damaged bases from the genome [15]. One of the most well-known is Uracil DNA glycosylase, an important enzyme not only in base excision repair but also in somatic hypermutation during the diversification of antibodies [16–19]. It removes uracil, generated either by spontaneously or enzymatic deamination of cytosine. The Ap site undergoes spontaneous or enzymatic cleavage. The Ap endonucleases cleave the phosphodiester bond of the Ap site while producing a nick with 3′-deoxyribose-5′-phosphate and a single-strand DNA (ssDNA) break [20,21]. Two groups of Ap lyases cleave the Ap site either *via* β-elimination generating an unsaturated 3′-aldehyde or through the β-δ

Abbreviations: Ap site, apurinic/apyrimidinic site, abasic site in DNA; ICL, Interstrand DNA crosslink; NEIL3, nei-like DNA glycoslase 3.

* Corresponding author.
E-mail address: silhan@uochb.cas.cz (J. Silhan).

[1] These authors contributed equally to this work

https://doi.org/10.1016/j.dnarep.2022.103300
Received 7 January 2022; Received in revised form 8 February 2022; Accepted 8 February 2022
Available online 11 February 2022
1568-7864/© 2022 Published by Elsevier B.V.

elimination to yield 3′-phosphate [22,23]. The Ap site also undergoes spontaneous breakage with a half-life of 8 days, and the rate can be enhanced by basic hydrolysis catalysed by amines or an increase of pH. First, the Ap site undergoes β-elimination followed by δ-elimination with 3′-phosphate as the final endproduct [24,25]. These toxic 3′-lesions block ligation and replication and are generally repaired by specific 3′-phosphates, such as PNK or Ap exonucleases [26–29]. Finally, the generated ssDNA break with 3′-OH is repaired by a DNA polymerase and ligase [25,30,31]. A nucleophilic attack of the Ap site may form a DNA interstrand crosslink or a DNA-protein crosslink [4,32].

From a chemical point of view, an Ap site undergoes structural changes, such as mutarotation and spontaneous decomposition [33]. Spontaneous base loss is caused by hydrolysis where the C1 of deoxyribose is attacked by a water molecule from the opposite side of the N-glycosidic bond [30]. This reaction leads to the formation of an Ap site with an alpha-OH group at C1. Further on the Ap site, it undergoes mutarotation until the thermodynamic equilibrium is reached [33,34]. Interestingly, this process reaches equilibrium when only 1% of the ribose ring opens to its aldehyde form [33,35]. Aldehydes are reactive groups that readily interact with amines, forming a relatively stable covalent imine commonly known as a Schiff base [36–38].

Very early on, it was suggested that the Ap site can generate ICL in double-stranded DNA [39,40]. This was later confirmed and further investigated in several publications [41–43]. Most of these experiments were carried out at relatively low pH, somewhat harsh and non-physiological conditions. Nonetheless, one rough estimate at pH = 6.8 in T7 phage DNA was that 1 crosslink is formed per 140 Ap sites [40]. Eventually, the nature of the chemical agent responsible for crosslink formation on the Ap site was found to be the aldehyde group, and the optimum for Ap-ICL formation is in the low pH region [44].

Successfully reconstituted Ap-ICL at a specific site *in vitro* was accomplished for the Ap site crosslinking with guanine in an opposite DNA strand within the 5′-dC-Ap sequence [4]. These crosslinks were formed under relatively acidic conditions (pH = 5) and stabilized by the reduction with sodium cyanoborohydride ($NaBH_3CN$) and characterized by mass spectroscopy. Shortly thereafter, another Ap-ICL was discovered where the Ap site linking with opposing adenine downstream of the Ap site, and therefore a different sequence context (5′-ApT) was confirmed. This Ap-ICL was shown to be stable, and it formed under physiological conditions [46]. A subsequent NMR study confirmed that one of the Ap-ICLs generated between ApT:A is attached with the N6-amino group of the dA from the opposite strand [47]. In this study, Ap-ICL was extracted from the duplex by enzymatic digestion and analysed by NMR. The presence of different sugar moiety isomers was found in equilibrium. The latest NMR study reveals that the structure of Ap-ICL has the Ap site's ribose ring closed in its furanose form [45]. In this work, the DNA contains a 5′-dC-Ap-dT sequence (Fig. 1).

These studies have sparked questions regarding the repair of these Ap-ICLs in the cell. The first repair that was discovered is linked to replication and is orchestrated by the DNA glycosylase NEIL3 [7]. Subsequently, it was demonstrated that this enzyme requires a replication fork with a specific orientation of the Ap part of the lesion to be catalytically efficient. Furthermore, NEIL3 is recruited to the ICL by TRAIP1 dependent ubiquitylation of DNA helicase which initiates the ICL repair [8]. Finally, it has been described that particular orientation of the Ap-ICL site within the DNA fork is necessary for sufficient removal

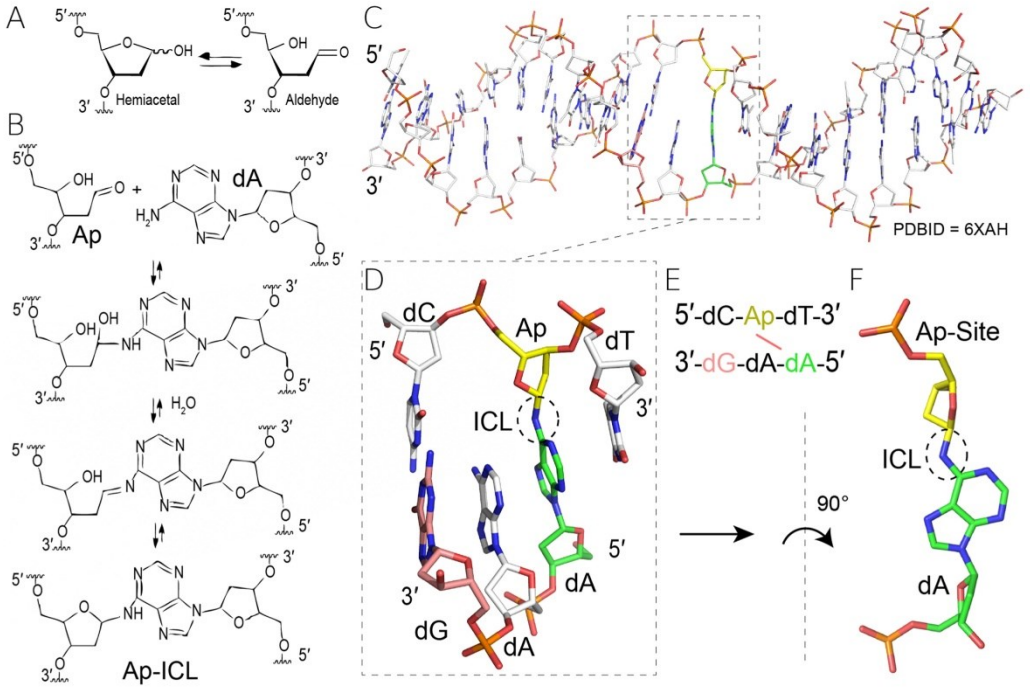

Fig. 1. Structure, sequence and formation of the Ap-ICL.A) The Ap site in an open and closed form, B) schematics of formation of Ap-ICL. C) The NMR model of the structure of Ap-ICL. D). An enlarged side view of the Ap-ICL within the duplex DNA with orphaned bases (Ap site in yellow, dG which is known to also form Ap-ICL in pink, and crosslinked adenine in green), E) the DNA sequence schematics in the vicinity of Ap-ICL in this NMR structure, and F) top view of the Ap-ICL crosslink (C-F) [45].(For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

of the Ap-ICL [48]. The knowledge regarding the feasibility and prevalence of the Ap-ICL, however, remains elusive. Moreover, the circumstances such as the local sequence leading to formations and its kinetics are also only fragmental.

Here, we aim to reveal a sequence of contexts and conditions leading to the formation of the Ap-ICL *in vitro* and to predict the relevance, frequency and feasibility of Ap-ICL formation *in vivo*. With this goal in mind, we designed a panel of DNA oligonucleotides to study the rate of formation of Ap-ICL *in vitro* with regard to the sequence context. Our data reveal that individual changes in nucleobases surrounding Ap-ICLs influence the rate of formation of the Ap-site, but only to a certain extent. Significant changes in formation rates have been observed for Ap-ICLs generated within the duplex with AT-rich regions surrounding the crosslink lesion. We then followed the stability of the purified Ap-ICLs, and from our biochemical experiments regarding the formation of Ap-ICLs, we have estimated a theoretical number of Ap-ICLs formed in a living cell.

2. Materials and methods

2.1. Oligonucleotides for preparation of DNA containing native Ap site

All synthetic DNA oligonucleotides were acquired from (Sigma-Aldrich). Oligonucleotides for the preparation of native Ap sites were ordered with deoxy-uracil and were labelled with HEX (Table 1).

2.2. DNA crosslink formation

The reaction buffer for crosslinking reactions was optimized. The final reaction buffer was composed of 20 mM HEPES, pH = 6.5, 140 mM NaCl, 0.5 mM TCEP, and 5% glycerol. Comparative crosslinking experiments were carried out at Tris-HCl pH = 7.4, 140 mM NaCl and 5% glycerol to outline the similarity of results under physiological conditions (Supplementary Fig. S4). We mixed labelled DNA oligonucleotides with non-labelled complementary oligonucleotides 1:1 in an optimized buffer to the final DNA concentration of 2.5 μM. The DNA reaction mixture (50 μl) was annealed using Biometra Tprofessional Thermocycler by heating to 95 °C and then cooling to room temperature using the slowest ramp. For the Ap site creation, 0.5 U uracil-DNA glycosylase (UDG) (New England Biolabs) per one sequence reaction was added and kept 5 min at room temperature. Afterwards, UDG was inactivated with another round of an annealing cycle. The annealed DNA reaction mixture containing the Ap site was incubated for the duration of the experiment at 37 °C. At given time points, 1 ul of the crosslinking reaction was taken and mixed with 9 ul of formamide to stop the reaction. The stopped reaction was separated on a 20% denaturing polyacrylamide gel (1 x TBE, 7 M Urea 20% polyacrylamide: bisacrylamide 19:1) and visualized using a 532 nm laser and using a Cy3 570BP20 filter on an Amersham Typhoon Biomolecular Imager (GE Healthcare). The progress of the crosslinking reaction was quantified using a commercial ImageQuant TL. All bands in the individual lanes (substrate, product and intermediates) were densitometrically quantified to determine the percentage of the crosslink formed, the methodology was described previously in detail [49]. Control reaction with thymine in place of uracil to prevent Ap-ICL reaction was carried out with and without UDG, this control aimed to demonstrate the specificity of crosslinking at Ap site (Supplementary Fig. S2).

2.3. Isolation of crosslinked DNA from polyacrylamide gel for further reactions

Ap-ICL crosslink was purified from the PAGE gel using the modified band excision method from the gel [50]. The sample of crosslinked DNA was run on a 20% denaturing polyacrylamide gel. Bands were visualized using HeroLab UVT-20 S/M/L, and a band of crosslinked DNA was cut out, chopped and transferred to a tube with 200 μl of elution buffer (20 mM Tris, pH 7.4, 140 mM NaCl, 0.5 mM TCEP, 5% glycerol). The tube was left on a rotating wheel at 8 °C overnight. The sample was then centrifuged at 800 *g* for 1 min. The remaining buffer was recovered and purified using Cytiva MicroSpin™ G-25 Columns against the elution buffer to remove excess urea. The final sample of purified Ap-ICL DNA was run on 20% denaturing polyacrylamide gel to verify its purity.

2.4. Analysis of degradation of crosslinked DNA

The stability of crosslinked DNA was assessed at 37 °C in 20 mM Tris, pH 7.4, 140 mM NaCl, 0.5 mM TCEP, 5% glycerol. At given time points, typically 1 day, 2 μl of a solution of crosslinked DNA was taken, 8 μl of

Table 1
DNA oligonucleotide sequences. Sequences of dsDNA oligonucleotides, and their names used in this study. Ap site is generated from uracil (U) on the position 2 and Ap-ICL forms with adenine on the position 6, the reaction confirming the specificity of Ap-ICL crosslinking reaction is shown in Supplementary Fig. S3.

4G	5′[HEX]-GTCGATGAAC$_1$U$_2$T$_3$AGACAGCT 3′-CAGCTACTTG$_4$A$_5$A$_6$TCTGTCGA	4I	5′[HEX]-GTCGATGAAC$_1$U$_2$T$_3$AGACAGCT 3′-CAGCTACTTI$_4$A$_5$A$_6$TCTGTCGA
4C	5′[HEX]-GTCGATGAAC$_1$U$_2$T$_3$AGACAGCT 3′-CAGCTACTTC$_4$A$_5$A$_6$TCTGTCGA	6I	5′[HEX]-GTCGATGAAC$_1$U$_2$T$_3$AGACAGCT 3′-CAGCTACTTG$_4$A$_5$I$_6$TCTGTCGA
4T	5′[HEX]-GTCGATGAAC$_1$U$_2$T$_3$AGACAGCT 3′-CAGCTACTTT$_4$A$_5$A$_6$TCTGTCGA		
4(5)A	5′[HEX]-GTCGATGAAC$_1$U$_2$T$_3$AGACAGCT 3′-CAGCTACTTA$_4$A$_5$A$_6$TCTGTCGA	TT	5′[HEX]-GTCAGATCAAA$_1$U$_2$T$_3$AAACCTG 3′-CAGTCTAGTTT$_4$G$_5$A$_6$TTTGGAC
5G	5′[HEX]-GTCGATGAAC$_1$U$_2$T$_3$AGACAGCT 3′-CAGCTACTTA$_4$G$_5$A$_6$TCTGTCGA	TC	5′[HEX]-GTCAGATCAAG$_1$U$_2$T$_3$AAACCTG 3′-CAGTCTAGTTC$_4$G$_5$A$_6$TTTGGAC
5C	5′[HEX]-GTCGATGAAC$_1$U$_2$T$_3$AGACAGCT 3′-CAGCTACTTA$_4$C$_5$A$_6$TCTGTCGA	GT	5′[HEX]-GTCAGATCCCA$_1$U$_2$T$_3$CCACCTG 3′-CAGTCTAGGGT$_4$G$_5$A$_6$GGTGGAC
5T	5′[HEX]-GTCGATGAAC$_1$U$_2$T$_3$AGACAGCT 3′-CAGCTACTTA$_4$T$_5$A$_6$TCTGTCGA	GC	5′[HEX]-GTCAGATCCCG$_1$U$_2$T$_3$CCACCTG 3′-CAGTCTAGGGC$_4$G$_5$A$_6$GGTGGAC

formamide was added, and the sample was frozen at −80 °C to terminate the degradation. After that, all samples were resolved on 20% denaturing polyacrylamide gel to validate the progress of degradation. The gel was scanned using a laser at 532 nm and a Cy3 570BP20 filter on the Amersham Typhoon Biomolecular Imager (GE Healthcare). The image was analyzed using the commercial ImageQuant TL. Lines were selected manually, and the background subtraction was performed using the rolling ball method. Bands with constant dimensions encompassing crosslinked DNA and degradation products were selected manually. For every line, the percentage of degradation of crosslinked DNA was calculated based on the ratio of the remaining crosslinked DNA and the degradation product.

3. Results

3.1. Design of the DNA substrates

Uracil, and thus the Ap site, was situated in the middle of the oligonucleotide duplex. All residues in the vicinity were named accordingly (Fig. 2A), and the position of the Ap site was numbered 2. The base to which the majority of crosslinks is formed was fixed to be adenine at position 6 [36]. Therefore, in our experimental setup, positions 2 and 6 were constant as there were involved in the Ap-ICL formation with the exception of a control experiment validating the preferential formation of these sites. Only natural DNA nucleobases were altered in other positions to obtain all possible combinations in the vicinity of the Ap site. Position 5 is opposite to the Ap site lesion. Because positions 1 and 4 were intrinsically linked by complementarity, thereafter the experiments were focused on positions 4 and 5. Subsequently, a wider area of the Ap site was investigated with the specific goal of comparing the AT-rich and GC-rich regions (Fig. 3).

3.2. Generation of the Ap site and crosslink formation

Purchased synthetic DNA oligonucleotides contained deoxyuridine and were labelled with 5′ fluorescent dye (HEX, 6-hexachloro-fluorescein). Uracils were introduced to easily generate a native Ap site. First, the oligos were annealed with a complementary DNA strand and

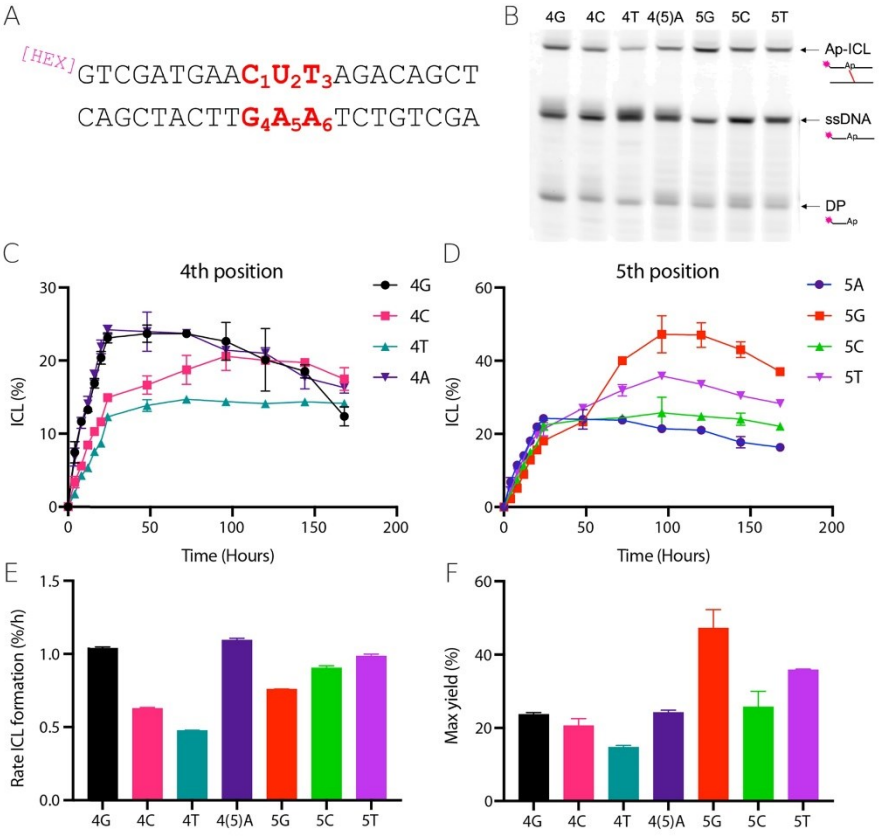

Fig. 2. Reaction kinetics of the Ap site DNA when varying bases at positions 4 and 5, respectively.*(A) Schematic design of the DNA duplex containing the Ap site and nomenclature of different positions. Position 2 is where the Ap site is formed by reaction with UDG, and ICL is formed between positions 2 and 6. Reactions were initiated by the formation of the Ap site by UDG and kept in dark at 37 °C (B) Reactions in each timepoint were stopped (here 24 h) and immediately resolved on 20% denaturing PAGE gel, ssDNA marks single-stranded DNA and DP donates degradation product. The gels were evaluated and the proportion of Ap-ICL formed from DNA oligos with different nucleobase in position 4(C) or position 5(D) were plotted in the graph. The initial rates of Ap ICL formation (E) and the maximum yield (F) of Ap ICL we plotted as bar charts. Error bars represent SEM from three independent repeats. Comparative gel at Tris-HCl pH = 7.4, 140 mM NaCl shown in the* Supplementary Fig. S4.

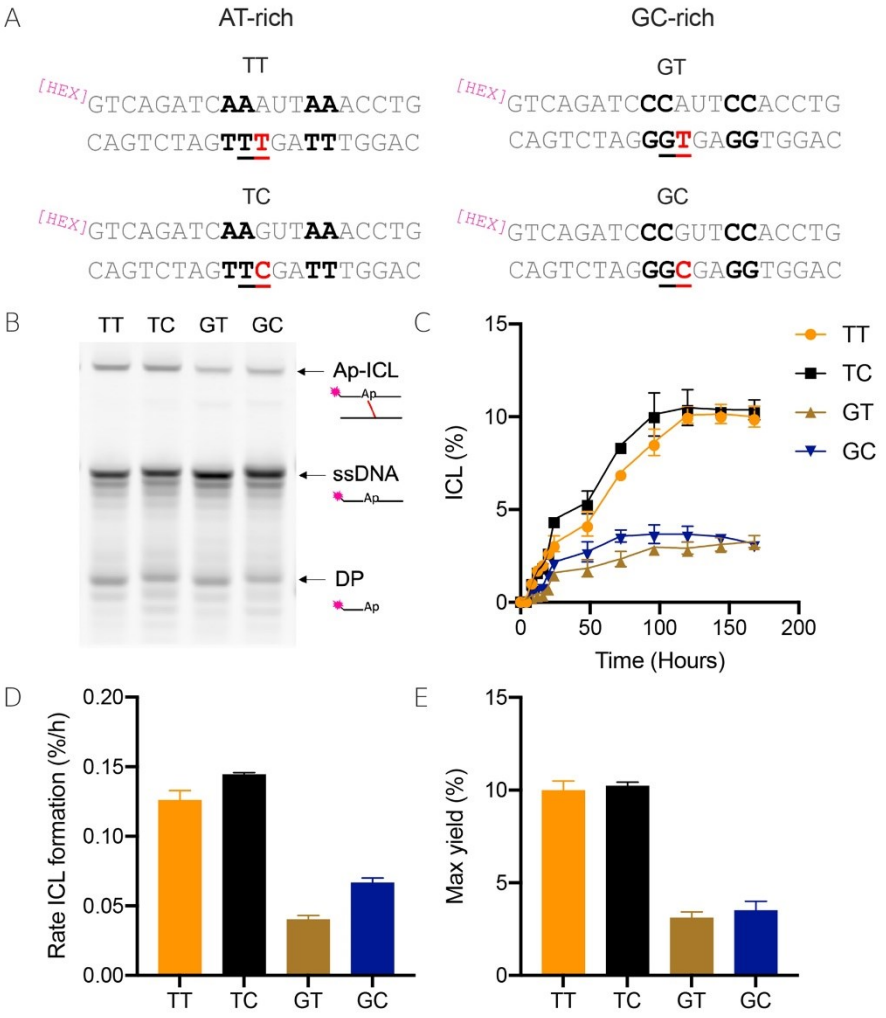

Fig. 3. Formation of Ap-ICL in DNA duplexes with AT-rich or GC-rich regions. *(A) Schematics of DNA duplexes containing i) AT-rich regions (left) or ii) GC-rich regions (right), in the vicinity of the Ap site. AT rich sequences where thymine and cytosine are interchanged in the 4th position are named TT or TC; similarly, GC rich sequences with bases in the 4th position are named GT or GC. (B) Reactions were resolved on 20% denaturing PAGE gel, ssDNA marks single-stranded DNA and DP donates degradation product. (C) Reaction kinetics of Ap-ICL formation from DNA oligos with AT-rich or GC-rich regions. The rate of ICL formation (D) and the maximum yield (E) were plotted for all four sequences. Error bars represent SEM from four independent repeats.*

incubated with an UDG [51,52]. The minimal amount of this enzyme was estimated by its titration into the reaction mixture. The presence of the generated Ap site was monitored on the PAGE gel after its sufficient hydrolysis by NaOH at an elevated temperature (Supplementary Fig. S1). Typically, a 5 min incubation with UDG was sufficient for a total conversion of the uracil to the Ap site.

The crosslinking reaction was allowed to proceed in a dark incubator at 37 °C. Aliquots were taken at different time points to monitor the time course. The fluorescently labelled samples were analyzed on denaturing PAGE gels, imaged, quantified and the percentage of generated crosslink was plotted in a graph. Initial rates were calculated as a linear fit to the data and maximum yields of Ap-ICL were determined as the maximum measured value.

3.3. The opposite nucleotide to the Ap site influences the rate of formation of Ap-ICL

The chemical nature of the DNA is as such that Ap sites can be generated spontaneously at any given DNA context [30]. In our first set of experiments, the single base opposing the Ap site was varied whilst the sequence of the entire remaining DNA duplexed was identical. In this setup, all four bases were tested on this position that we named the 5th

position (Fig. 2A), The rate of formation of Ap-ICLs were followed so as the maximum yields. The data were plotted in the graph and to bar charts. The observed rates were 1.1%/h for 5 A, 0.76%/h for 5 G, 0.91%/h for 5 C and 0.99%/h for 5 T (we took the liberty to express the rates in the percentage of original Ap sites converted in one hour to Ap-ICL). The maximum yield for all these oligos was 24.25% for 5 A, 47.2% for 5 G, 25.8% for 5 C and 35.9% for 5 T of the total DNA (For further rates please refer to Supplementary Tables ST1 – ST3). Later, with an increasing amount of degradation product, the amount of Ap-ICL started to decline. There were significant differences in the rate of formation as shown in the graph. Overall, these changes were relatively small and regardless of the nucleobase present opposite to the Ap site the Ap-ICL formed to a similar extent (Fig. 2).

3.4. Varying base-pair adjacent to Ap site on its 5′ end

Similar to the variation of the bases opposite to the AP site, the adjacent base-pair occupying positions 1 and 4 were varied. All four base pair combinations were tested for these positions and DNA oligonucleotides were named 4 G, 4 T, 4 C, and 4(5)A. Sequences 4 A and 5 A are identically co-named 4(5)A as 4 A was kept constant at 5th position and *vice-versa* with sequence 5 A. Yet again, significant differences in the formation of the Ap-site crosslink were observed. Surprisingly enough, the differences in rates of formation were not extraordinary, but the Ap-ICL was yet again formed in all possible base combinations. Overall, the rates vary approximately 2-fold (Supplementary Table ST1). The maximum yield varies slightly, and it was reached at different timepoints (Fig. 2 & ST1). The initial rate was the lowest for both pyrimidine bases. Both thymine and cytosine containing duplexes reached their maximum yield of crosslinks later but the percentages were not dissimilar with the remaining two bases at this position. Once again the amount of crosslinks starts to decline due to the irreversible decomposition to smaller degradation products (Fig. 2). To confirm published work and demonstrate the specificity of Ap-ICL forming from Ap with adenine residues, two control experiments were carried out. For sequence 4 G, hypoxanthine was placed at positions 4 and 6. No visible Ap-G crosslink was formed when adenine at position 6 was mutated to hypoxanthine. In contrast, when hypoxanthine was placed at position 4, however, Ap-A crosslink was formed as readily as when guanine was present (Fig. SI 3).

3.5. AT-rich regions facilitate Ap-ICL formation

Investigating the sequence surrounding the Ap site has revealed the differences in the rate of Ap-ICL formation, although a more complete picture requires investigation of further sequence variation that may influence the rigidity of the entire region encompassing the Ap site. We modified the sequence in the wider surrounding of the Ap site by a set of two AT or GC pairs from both sites (Fig. 3A). We set out to determine how AT-rich and GC-rich sequences influence the rate of formation of Ap-ICL. These sequences were tested in non-identical duplicates for which the base at 4th position contained either cytosine or thymine and the oligos were named accordingly (TT, TC, GT, GC), *e.g.* TC is an AT-rich oligonucleotide with C at 4th position (Fig. 3A). These bases were chosen to extend the AT-rich or GC-rich part around the Ap-site. The reaction set-up was identical as above, where reaction progress was monitored on a denaturing PAGE gel. The amount of formed Ap-ICL was apparent from the gel figures (Fig. 3B). Next, the gels were evaluated, and initial reaction rates were determined from the slopes of the data points.

When AT- and GC-rich regions were swapped, we observed significant differences in the rate of formation (Fig. 3C). The effect of the cytosine or thymine base at the 4th position on the rate of the reaction is rather small and the rates for oligos containing cytosine were faster than those with thymine. This was observed for both AT-rich and GC-rich DNA duplexes, and this trend was consistent in experiments shown in Fig. 2 and Fig. 3 confirming the robustness of our approach. Finally, the rate of formation for the AT-rich sequence with a cytosine at 4th position (TC) had a nearly 4 times higher rate of formation than the one with GC-rich sequence with thymine at 4th position (GT). This trend was observed for both pairs of sequences (TC > TT >>> GC > GT), and it was also reflected in a maximum yield reaching an approximate 2.5-fold difference (Fig. 3E).

3.6. Stability of Ap-ICL

Finally, the stability of Ap-ICL was tested to determine how persistent the crosslink could be at physiological conditions and without the intervention of other proteins and enzymes. The Ap-ICLs were purified by isolation from the gel. Large volumes of the crosslinking reactions stopped approximately at the timepoint for the maximum yield of the Ap-ICL. After their isolation, the pure Ap-ICLs were buffer-exchanged into a physiological buffer and subjected to their spontaneous decomposition. Our initial experiments as well as published literature have suggested that the decomposition of Ap-ICL is a lengthy process [53]. Therefore, the subsamples were taken in 24 h intervals, and after two weeks, all reaction time points were resolved on denaturing PAGE gels (Fig. 4 A). The proportions of degrading Ap-ICL were plotted as a function of time, the initial reaction rates of their decomposition were fitted with single exponential decay and plotted into a bar chart (Fig. 4B & 4 C). The ICL band disintegrated into a band the size of the original substrate. With a delay, more products appeared, with the nature of these substrates arising from the spontaneous hydrolysis of the Ap site. Sequences with AT- and GC-rich regions were used in this experiment, and all reactions had a similar rate of decomposition, except for an AT-rich sequence with a T at the 4th position (TT) that decomposed at a slightly slower rate. Less than 1/5 of the initial material was decomposed in the first 10 days (Fig. 4).

4. Discussion

The first studies of specific Ap-ICL outlined two possible ways by which the Ap-ICL can form [2,37]. They demonstrated that within a duplex within this exemplary sequence 5′-dC-Ap-dT, Ap-ICL can be formed by crosslinking Ap with guanine or adenine in the opposite DNA strand. The latter study showed that an Ap-G crosslink is not stable unless reduced with $NaBH_3CN$. This is presumably due to its sterically unfavourable arrangement, and the crosslink reverted back to the Ap site. Importantly, and perhaps coincidentally, the identical sequence motif was used to determine the NMR structure of Ap-ICL within the DNA duplex. In this study, only one type of crosslink was observed, making a 5′Ap-dT "hotspot" of Ap site crosslink formation *in vivo* (Fig. 1) [36]. Recently, another Ap-ICL where Ap site linking amino group of cytosine from opposite DNA strand (Ap-C) was observed. This Ap-ICL was preferentially formed within where C was mispaired with A. This crosslink was only formed at quite acidic conditions pH = 5 and was shown not to form under physiological conditions in nonmispaired oligonucleotide [54].

In this manuscript, we focused on Ap-ICL formed within this 5′-Ap-dT "hot-spot". We have a synthesized and annealed panel of DNA oligonucleotides with different DNA sequences surrounding the Ap site. The entire study used this non-enzymatic Ap-ICL formation, and by measuring its kinetics, we revealed sequence circumstances and influences leading to the formation of ICLs. To investigate the generation of the Ap-ICL, uracil was converted to the Ap site by an enzymatic reaction with UDG. We then set up the reactions to measure the rates of ICL formation for individual oligonucleotides with different sequences. Reactions were terminated at different time points, resolved on PAGE gels and the fractions of crosslinks formed were plotted against time. This allowed us to follow the kinetics of ICL formation. Our data has revealed that different nucleotides surrounding the Ap site affected the rate of the Ap-ICL formation. Although the rates differed to an extent,

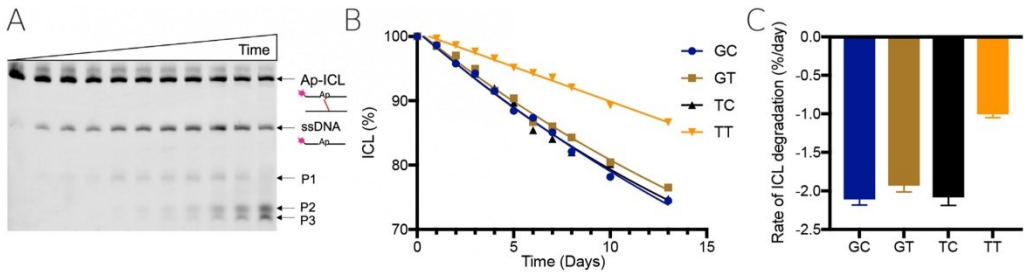

Fig. 4. Stability and degradation of Ap-ICL with AT-rich or GC-rich regions. *Gel purified Ap-ICLs from DNA oligos with AT-rich or GC-rich regions flanking the Ap site were subjected to spontaneous degradation at physiological conditions. (A) An illustrative 20% denaturing PAGE gel of Ap-ICL degradation in time is shown, where P1, P2 and P3 are products of degradation. HEX labelled fluorescent imaging. (B) Degradation of Ap-ICL was plotted over time and fitted with exponential decay (C) The rate of Ap-ICL degradation was plotted and is represented as a bar chart. The half-lives for GC, GT, TC and TT were 28.8; 31.7; 29.3 and 60.6 days.*

what we found most striking was that the Ap-ICLs form consistently within all of these sequences and that the resulting crosslinks were very stable with a half-life longer than 30 days. This is in agreement with the literature where different Ap-ICL were already been reported to be relatively stable [47,53–55].

The data show that the rate of formation of the Ap-ICLs is only slightly affected by different single nucleotide or one base pair substitutions in the vicinity of the Ap site. The maximum yield of Ap-ICLs for all of these sequences was roughly 15%. Meaning that on average 15% of Ap sites are converted into Ap-ICLs. On the other hand, the formation of ICLs is significantly affected by the presence of AT-rich or GC-rich regions flanking the Ap site. In the case of AT-rich sequences, we observed significantly higher rates of formation. We speculate that this may be because AT-rich regions have a lower melting temperature and thus allow for "DNA breathing". We predict that weaker base-pairing, in turn, would allow for crosslinking reactions to proceed with less difficulty and faster. Although our study is focused purely on the *in vitro*, non-enzymatic formation of Ap-ICLs, the study aimed to broaden understanding of which circumstances and how often the crosslinks can form *in vivo*. Below, we attempt to make a rough estimate on the number of Ap-ICLs in one human cell, as our work shed light on a rough number of Ap-ICLs formed per 100 Ap sites.

Spontaneous deamination and depurination generate 10,000 Ap/sites/day in a cell [13,14]. In addition to this, the DNA damage and action Ap site generation by various DNA glycosylases further increase this number [30]. To accurately count this number can be challenging. Our rough estimate disregards that DNA repair enzymes are present, and they eliminate the majority of Ap sites. Then, observable Ap sites at any given moment may only represent 0.25–4% of the sites that would form the Ap-ICL during the day. We may predict that the majority of these Ap sites are repaired or protected by the formation of DNA protein crosslinks, and therefore only a small proportion of Ap sites are present [53,56–58]. In recent work, a relative number of Ap-sites were estimated and compared against the knockdown of human Ap endonuclease APEX1. It has been observed that an extra 77% more Ap sites were present in the knockdown, while 90% of the enzyme was removed [59]. It is difficult to estimate the original numbers of Ap sites from this relative comparison, but these numbers suggest that significant numbers still persist in the genome.

With DNA fibre analysis, it was predicted that the number of Ap sites to be more than 35,000 Ap sites in one human cell [56]. On the other hand, further methods for determining the presence of Ap sites have been developed. A mass spectroscopy method using a chemical O-(pyridine-3-yl-methyl) hydroxylamine (PMOA) is reportedly a very sensitive technique for the detection of Ap sites directly within the duplex [60]. This technique is quantitative and allows for the detection of up to 2 Ap sites in 10^8 nt. With such sensitivity, a rough number of the steady-state Ap sites detected by this method is 2 Ap sites per 10^7 nt [61]. Considering that the human genome has 3.2×10^9 nt and is diploid, then one human cell would have only 160 Ap sites. It is noteworthy, that experiments using aldehyde reactive probes might be biased as the probe may react with other aldehydes in the DNA. The Ap site undergoes a mutarotation as do other sugar moieties, and generally, only less than 1% of the sugars are in open aldehyde form [33,35]. Therefore, the rate of derivatization of the Ap site with an aldehyde reactive probe will be determined by the ring-opening reaction of the ribose into the aldehyde form. The rate of ring-opening can be noticed in mass spectroscopy studies where PMOA reagent was used. Here, almost the full conversion of small DNA oligonucleotide containing Ap site to a PMOA adduct was shown on HPLC chromatograms and confirmed by mass spectroscopy [62]. Additionally, a biochemical study with an aldehyde reactive probe conjugated with biotin showed that the majority of Ap sites were conjugated with the probe if sufficient time and reagents were in supply [63]. Both studies with PMOA, including one with mammalian samples, used excessive amounts of PMOA allowing first-order kinetics where ring-opening of Ap site is a rate-limiting reaction. Thereafter, we believe that HPLC-MS detection of conjugated Ap-sites is specific and validated by the MS method and might be the most accurate. This leads to a steady-state value of only about 100–200 Ap sites present in the genome. Moreover, product inhibited bifunctional glycosylases may bind and crosslink the Ap-sites [32,64,65]. Other DNA repair enzymes and DNA binding proteins interacting with Ap sites could impede the Ap-ICL formation thus protecting the genome integrity [57,66,67]. Additionally, it has been also shown that Ap-ICL crosslink can be reverted beside the repair by NEIL3 glycosylase [7,53]. On the other hand, human Ap endonuclease and exonuclease APEX1 was shown to cleave the Ap-ICL using its exonuclease activity on degraded Ap-G crosslink. These protective processes could explain the relatively low number of Ap sites detected using PMOA and mass spectroscopy. Based on the aforementioned predictions and our data, we estimate the amount of Ap-ICL crosslink to be 1–5 Ap-ICLs present or formed in any human cell.

Author contribution

BL and AH designed, optimized and carried out experiments, AH designed figures, EB and JS supervised the study and wrote the manuscript.

Acknowledgements

This work was supported by the Czech Science Foundation (17–21649Y), and we thank the Academy of Sciences Czech Republic for support in the form of a J.E. Purkyne Fellowship awarded to JS. The

project was also supported by the Academy of Sciences Czech Republic (RVO: 61388963). We cordially acknowledge Edward Curtis and Tamara Jenkins for the critical reading and proofreading of the manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.dnarep.2022.103300.

References

[1] J. Zhang, J.C. Walter, Mechanism and regulation of incisions during DNA interstrand cross link repair, DNA Repair 19 (2014) 135–142, https://doi.org/10.1016/j.dnarep.2014.03.018 ((Amst)).

[2] F. Coste, J.-M. Malinge, L. Serre, W. Shepard, M. Roth, M. Leng, C. Zelwer, Crystal structure of a double-stranded DNA decamer containing a cisplatin interstrand cross link adduct, Nucleic Acids Res. 27 (1999) 1837–1846.

[3] D.M. Noll, T. McGregor Mason, P.S. Miller, Formation and repair of interstrand cross-links in DNA, Chem. Rev. 106 (2006) 277–301, https://doi.org/10.1021/cr040478b.

[4] K.M. Johnson, N.E. Price, J. Wang, M.I. Fekry, S. Dutta, D.R. Seiner, Y. Wang, K. S. Gates, On the formation and properties of interstrand DNA DNA cross-links forged by reaction of an abasic site with the opposing guanine residue of 5′-CAp sequences in duplex DNA, J. Am. Chem. Soc. 135 (2013) 1015–1025, https://doi.org/10.1021/ja308119q.

[5] J. Nakamura, M. Nakamura, DNA protein crosslink formation by endogenous aldehydes and AP sites, DNA Repair 88 (2020), 102806, https://doi.org/10.1016/j.dnarep.2020.102806 ((Amst)).

[6] M. Räschle, P. Knipscheer, M. Enoiu, T. Angelov, J. Sun, J.D. Griffith, T. E. Ellenberger, O.D. Schärer, J.C. Walter, Mechanism of replication-coupled DNA interstrand crosslink repair, Cell 134 (2008) 969–980, https://doi.org/10.1016/j.cell.2008.08.030.

[7] D.R. Semlow, J. Zhang, M. Budzowska, A.C. Drohat, J.C. Walter, Replication-dependent unhooking of DNA interstrand cross-links by the NEIL3 Glycosylase, e14, Cell 167 (2016) 498–511, https://doi.org/10.1016/j.cell.2016.09.008.

[8] R.A. Wu, D.R. Semlow, A.N. Kamimae-Lanning, O.V. Kochenova, G. Chistol, M.R. G. Hodskinson, R. Amunugama, J.L. Sparks, M. Wang, L. Deng, C.A. Mimoso, E. Low, K.J. Patel, J.C. Walter, TRAIP is a master regulator of DNA interstrand crosslink repair, Nature 567 (2019) 267–272, https://doi.org/10.1038/s41586-019-1002-0.

[9] P.R. Martin, S. Couvé, C. Zutterling, M.S. Albelazi, R. Groisman, B.T. Matkarimov, J.L. Parsons, R.H. Elder, M.K. Saparbaev, The human DNA glycosylases NEIL1 and NEIL3 excise psoralen induced DNA DNA cross links in a four stranded DNA structure, Sci. Rep. 7 (2017) 1–13, https://doi.org/10.1038/s41598-017-17693-4.

[10] P. Knipscheer, M. Räschle, A. Smogorzewska, M. Enoiu, T.V. Ho, O.D. Schärer, S. J. Elledge, J.C. Walter, The fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair, Science 326 (2009) 1698–1701, https://doi.org/10.1126/science.1182372.

[11] M.R.G. Hodskinson, J. Silhan, G.P. Crossan, J.I. Garaycoechea, S. Mukherjee, C. M. Johnson, O.D. Schärer, K.J. Patel, Mouse SLX4 is a tumor suppressor that stimulates the activity of the nuclease XPF-ERCC1 in DNA crosslink repair, Mol. Cell. 54 (2014) 472–484, https://doi.org/10.1016/j.molcel.2014.03.014.

[12] M.R.G. Hodskinson, A. Bolner, K. Sato, A.N. Kamimae-Lanning, K. Rooijers, M. Witte, M. Mahesh, J. Silhan, M. Petek, D.M. Williams, J. Kind, J.W. Chin, K. J. Patel, P. Knipscheer, Alcohol-derived DNA crosslinks are repaired by two distinct mechanisms, Nature 579 (2020) 603–608, https://doi.org/10.1038/s41586-020-2059-5.

[13] T. Lindahl, B. Nyberg, Rate of depurination of native deoxyribonucleic acid, Biochemistry. 11 (1972) 3610–3618. (http://www.ncbi.nlm.nih.gov/pubmed/4626532).

[14] T. Lindahl, a Andersson, Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid, Biochemistry. 11 (1972) 3618–3623. (http://www.ncbi.nlm.nih.gov/pubmed/4559796).

[15] D. Svilar, E.M. Goellner, K.H. Almeida, R.W. Sobol, Base excision repair and lesion-dependent subpathways, Antioxid. Redox Signal. 14 (2011) 2491–2507.

[16] C.D. Mol, A.S. Arvai, R.J. Sanderson, G. Slupphaug, B. Kavli, H.E. Krokan, D. W. Mosbaugh, J.A. Tainer, Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: Protein mimicry of DNA, Cell 82 (1995) 701–708, https://doi.org/10.1016/0092-8674(95)90467-0.

[17] A.L. Ross, J.E. Sale, The catalytic activity of REV1 is employed during immunoglobulin gene diversification in DT40, Mol. Immunol. 43 (2006) 1587–1594, https://doi.org/10.1016/j.molimm.2005.09.017.

[18] C. Rada, G.T. Williams, H. Nilsen, D.E. Barnes, T. Lindahl, M.S. Neuberger, Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice, Curr. Biol. 12 (2002) 1748–1755, https://doi.org/10.1016/S0960-9822(02)01215-0.

[19] E.J. Steele, Mechanism of somatic hypermutation: critical analysis of strand biased mutation signatures at A:T and G:C base pairs, Mol. Immunol. 46 (2009) 305–320, https://doi.org/10.1016/j.molimm.2008.10.021.

[20] B. Demple, T. Herman, D.S. Chen, Cloning and expression of APE, the cDNA encoding the major human apurinic endonuclease: definition of a family of DNA repair enzymes, Proc. Natl. Acad. Sci. USA 88 (1991) 11450–11454, https://doi.org/10.1073/pnas.88.24.11450.

[21] D. Lu, J. Silhan, J.T. MacDonald, E.P. Carpenter, K. Jensen, C.M. Tang, G. S. Baldwin, P.S. Freemont, Structural basis for the recognition and cleavage of abasic DNA in Neisseria meningitidis, Proc. Natl. Acad. Sci. 109 (2012) 16852–16857, https://doi.org/10.1073/pnas.1206563109.

[22] M. Bhagwat, J. a Gerlt, 3′- and 5′-Strand cleavage reactions catalyzed by the Fpg protein from Escherichia coli occur via successive beta- and delta-elimination mechanisms, respectively, Biochemistry 35 (1996) 659–665, https://doi.org/10.1021/bi9522662.

[23] H.M. Nash, S.D. Bruner, O.D. Schärer, T. Kawate, T. a Addona, E. Spooner, W. S. Lane, G.L. Verdine, Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily, Curr. Biol. 6 (1996) 968–980. (http://www.ncbi.nlm.nih.gov/pubmed/8805338).

[24] B. Demple, L. Harrison, Repair of oxidative damage to DNA: enzymology and biology, Annu. Rev. Biochem. 63 (1994) 915–948, https://doi.org/10.1146/annurev.bi.63.070194.004411.

[25] V. Bailly, W.G. Verly, AP endonucleases and AP lyases, Nucleic Acids Res. 17 (1989) 3617–3618. (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC317828/).

[26] C.C. RICHARDSON, I.R. LEHMAN, A. KORNBERG, A deoxyribonucleic acid phosphatase-exonuclease from Escherichia coli, J. Biol. Chem. 239 (1964) 251–258.

[27] D.S. Chen, T. Herman, B. Demple, Two distinct human DNA diesterases that hydrolyze 3′-blocking deoxyribose fragments from oxidized DNA, Nucleic Acids Res. 19 (1991) 5907–5914, https://doi.org/10.1093/nar/19.21.5907.

[28] L. Wiederhold, J.B. Leppard, P. Kedar, F. Karimi-Busheri, A. Rasouli-Nia, M. Weinfeld, A.E. Tomkinson, T. Izumi, R. Prasad, S.H. Wilson, S. Mitra, T. K. Hazra, AP endonuclease-independent DNA base excision repair in human cells, Mol. Cell. 15 (2004) 209–220, https://doi.org/10.1016/j.molcel.2004.06.003.

[29] J. Silhan, Q. Zhao, E. Boura, H. Thomson, A. Förster, C.M. Tang, P.S. Freemont, G. S. Baldwin, Structural basis for recognition and repair of the 3′ phosphate by NExo, a base excision DNA repair nuclease from Neisseria meningitidis, Nucleic Acids Res. 46 (2018) 11980–11989, https://doi.org/10.1093/nar/gky934.

[30] D.E. Barnes, T. Lindahl, Repair and genetic consequences of endogenous DNA base damage in mammalian cells, Annu. Rev. Genet. 38 (2004) 445–476, https://doi.org/10.1146/annurev.genet.38.072902.092448.

[31] A. Álvarez-Quilón, J.L. Wojtaszek, M. Mathieu, T. Patel, C.D. Appel, N. Hustedt, S. E. Rossi, B.D. Wallace, D. Setiaputra, S. Adam, Y. Ohashi, H. Melo, T. Cho, C. Gervais, I.M. Muñoz, E. Grazzini, J.T.F. Young, J. Rouse, M. Zinda, R. S. Williams, D. Durocher, Endogenous DNA 3′ blocks are vulnerabilities for BRCA1 and BRCA2 deficiency and are reversed by the APE2 nuclease, Mol. Cell. (2020) 1–14, https://doi.org/10.1016/j.molcel.2020.05.021.

[32] R. Gilboa, D.O. Zharkov, G. Golan, A.S. Fernandes, S.E. Gerchman, E. Matz, J. H. Kycia, A.P. Grollman, G. Shoham, Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA, J. Biol. Chem. 277 (2002) 19811–19816, https://doi.org/10.1074/jbc.M202058200.

[33] K.-S. Ryu, C. Kim, I. Kim, S. Yoo, B.-S. Choi, C. Park, NMR application probes a novel and ubiquitous family of enzymes that alter monosaccharide configuration, J. Biol. Chem. 279 (2004) 25544–25548, https://doi.org/10.1074/jbc.M402016200.

[34] C. De Los Santos, M. El khateeb, P. Rege, K. Tian, F. Johnson, Impact of the C1′ configuration of abasic sites on DNA duplex structure, Biochemistry 43 (2004) 15349–15357, https://doi.org/10.1021/bi048400c.

[35] J.A. Wilde, P.H. Bolton, A. Mazumder, M. Manoharan, J.A. Gerlt, Characterization of the equilibrating forms of the aldehydic abasic site in duplex DNA by 17O NMR, J. Am. Chem. Soc. 111 (1989) 1894–1896, https://doi.org/10.1021/ja00187a062.

[36] S. Dimitrov, Y. Koleva, T.W. Schultz, J.D. Walker, O. Mekenyan, Interspecies quantitative structure-activity relationship model for aldehydes: aquatic toxicity, Environ. Toxicol. Chem. 23 (2004) 463–470, https://doi.org/10.1897/02-579.

[37] I.G. Minko, I.D. Kozekov, T.M. Harris, C.J. Rizzo, R.S. Lloyd, M.P. Stone, Chemistry and biology of DNA containing 1,N2 deoxyguanosine adducts of the α,β-unsaturated aldehydes acrolein, crotonaldehyde, and 4-hydroxynonenal, Chem. Res. Toxicol. 22 (2009) 759–778, https://doi.org/10.1021/tx9000489.

[38] H. S. Yu, T. Oyama, T. Isse, K. Kitagawa, T. T. P. Pham, M. Tanaka, T. Kawamoto, Formation of acetaldehyde-derived DNA adducts due to alcohol exposure, Chem. Biol. Interact. 188 (2010) 367–375, https://doi.org/10.1016/j.cbi.2010.08.005.

[39] E. Freese, M. Cashel, Crosslinking of deoxyribonucleic acid by exposure to low pH, BBA Spec. Sect. Nucleic Acids Relat. Subj. 91 (1964) 67–77, https://doi.org/10.1016/0926-6550(64)90171-9.

[40] J. Burnotte, W.G. Verly, Crosslinking of methylated DNA by moderate heating at neutral pH, Biochim. Biophys. Acta Nucleic Acids Protein Synth. 262 (1972) 449–452, https://doi.org/10.1016/0005-2787(72)90488-1.

[41] C. Goffin, W.G. Verly, Interstrand DNA crosslinks due to AP (apurinic/apyrimidinic) sites, FEBS Lett. 161 (1983) 140–144, https://doi.org/10.1016/0014-5793(83)80747-9.

[42] C. Goffin, S. Bricteux-Grégoire, W.G. Verly, Some properties of the interstrand crosslinks in depurinated DNA, Biochim. Biophys. Acta Gene Struct. Expr. 783 (1984) 1–5, https://doi.org/10.1016/0167-4781(84)90071-X.

[43] A.S. Prakash, N.W. Gibson, Sequence-selective depurination, DNA interstrand cross-linking and DNA strand break formation associated with alkylated DNA, Carcinogenesis 13 (1992) 425–431, https://doi.org/10.1093/carcin/13.3.425.

[44] S. Dutta, G. Chowdhury, K.S. Gates, Interstrand cross-links generated by abasic sites in duplex DNA, J. Am. Chem. Soc. 129 (2007) 1852–1853, https://doi.org/10.1021/ja067294u.

[45] A.H. Kellum, D.Y. Qiu, M.W. Voehler, W. Martin, K.S. Gates, M.P. Stone, Structure of a stable interstrand DNA cross-link involving a β-N-Glycosyl Linkage between an N6-dA amino group and an abasic site, Biochemistry 60 (2021) 41–52, https://doi.org/10.1021/acs.biochem.0c00596.

[46] N.E. Price, K.M. Johnson, J. Wang, M.I. Fekry, Y. Wang, K.S. Gates, Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex dna, J. Am. Chem. Soc. 136 (2014) 3483–3490, https://doi.org/10.1021/ja410969x.
[47] N.E. Price, M.J. Catalano, S. Liu, Y. Wang, K.S. Gates, Chemical and structural characterization of interstrand cross-links formed between abasic sites and adenine residues in duplex DNA, Nucleic Acids Res. 43 (2015) 3434–3441, https://doi.org/10.1093/nar/gkv174.
[48] M. Imani Nejad, K. Housh, A.A. Rodriguez, T. Haldar, S. Kathe, S.S. Wallace, B. F. Eichman, K.S. Gates, Unhooking of an interstrand cross-link at DNA fork structures by the DNA glycosylase NEIL3, DNA Repair 86 (2020), 102752, https://doi.org/10.1016/j.dnarep.2019.102752 ((Amst)).
[49] B. Cihlova, A. Huskova, J. Böserle, R. Nencka, E. Boura, J. Silhan, High throughput fluorescent assay for inhibitor screening of proteases from rna viruses, Molecules 26 (2021), https://doi.org/10.3390/molecules26133792.
[50] A.M. Maxam, W. Gilbert, A new method for sequencing DNA, Proc. Natl. Acad. Sci. 74 (1977) 560–564, https://doi.org/10.1073/pnas.74.2.560.
[51] T. Lindahl, An N-Glycosidase from Escherichia coli That Releases Free Uracil from DNA Containing Deaminated Cytosine Residues, Proc. Natl. Acad. Sci. 71 (1974) 3649–3653, https://doi.org/10.1073/pnas.71.9.3649.
[52] S.R.W. Bellamy, G.S. Baldwin, A kinetic analysis of substrate recognition by uracil-DNA glycosylase from herpes simplex virus type 1, Nucleic Acids Res. 29 (2001) 3857–3863.
[53] S.J. Admiraal, P.J. O'Brien, Base excision repair enzymes protect abasic sites in duplex DNA from interstrand cross-links, Biochemistry 54 (2015) 1849–1857, https://doi.org/10.1021/bi501491z.
[54] M.J. Catalano, S. Liu, N. Andersen, Z. Yang, K.M. Johnson, N.E. Price, Y. Wang, K. S. Gates, Chemical structure and properties of interstrand cross-links formed by reaction of guanine residues with abasic sites in duplex DNA, J. Am. Chem. Soc. 137 (2015) 3933–3945, https://doi.org/10.1021/jacs.5b00669.
[55] J.G. Varela, L.E. Pierce, X. Guo, N.E. Price, K.M. Johnson, Z. Yang, Y. Wang, K. S. Gates, Interstrand cross-link formation involving reaction of a mispaired cytosine residue with an abasic site in duplex DNA, Chem. Res. Toxicol. 34 (2021) 1124–1132, https://doi.org/10.1021/acs.chemrestox.1c00004.
[56] P.D. Chastain, J. Nakamura, S. Rao, H. Chu, J.G. Ibrahim, J.A. Swenberg, D. G. Kaufman, Abasic sites preferentially form at regions undergoing DNA replication, FASEB J 24 (2010) 3674–3680, https://doi.org/10.1096/fj.09-145276.
[57] K.N. Mohni, S.R. Wessel, R. Zhao, A.C. Wojciechowski, J.W. Luzwick, H. Layden, B. F. Eichman, P.S. Thompson, K.P.M. Mehta, D. Cortez, HMCES maintains genome integrity by shielding abasic sites in single-strand DNA, Cell 176 (2019) 144–153, https://doi.org/10.1016/j.cell.2018.10.055, e13.
[58] S. Boiteux, F. Coste, B. Castaing, Repair of 8-oxo 7,8-dihydroguanine in prokaryotic and eukaryotic cells: properties and biological roles of the Fpg and OGG1 DNA N-glycosylases, Free Radic. Biol. Med. 107 (2017) 179–201, https://doi.org/10.1016/j.freeradbiomed.2016.11.042.
[59] N. Wang, H. Bao, L. Chen, Y. Liu, Y. Li, B. Wu, H. Huang, Molecular basis of abasic site sensing in single-stranded DNA by the SRAP domain of E. coli yedK, Nucleic Acids Res. 47 (2019) 10388–10399, https://doi.org/10.1093/nar/gkz744.
[60] K.P. Roberts, J.A. Sobrino, J. Payton, L.B. Mason, R.J. Turesky, Determination of Apurinic/Apyrimidinic Lesions in DNA with High-Performance Liquid Chromatography and Tandem Mass Spectrometry, Chem. Res. Toxicol. 19 (2006) 300–309, https://doi.org/10.1021/tx0502589.
[61] J. Guo, H. Chen, P. Upadhyaya, Y. Zhao, R.J. Turesky, S.S. Hecht, Mass SPectrometric Quantitation of Apurinic/Apyrimidinic Sites in Tissue DNA of Rats Exposed to Tobacco Specific Nitrosamines and in Lung and Leukocyte DNA of Cigarette Smokers and Nonsmokers, Chem. Res. Toxicol. 33 (2020) 2475–2486, https://doi.org/10.1021/acs.chemrestox.0c00265.
[62] H. Chen, L. Yao, C. Brown, C.J. Rizzo, R.J. Turesky, Quantitation of apurinic/apyrimidinic sites in isolated DNA and in Mammalian tissue with a reduced level of artifacts, Anal. Chem. 91 (2019) 7403–7410, https://doi.org/10.1021/acs.analchem.9b01351.
[63] S. Bennett, J. Kitner, Characterization of the aldehyde reactive probe reaction with AP-sites in DNA: Influence of AP-lyase on adduct stability, Nucleosides Nucleotides Nucleic Acids 25 (2006) 823–842, https://doi.org/10.1080/15257770600726133.
[64] K. Nagorska, J. Silhan, Y. Li, V. Pelicic, P.S. Freemont, G.S. Baldwin, C.M. Tang, A network of enzymes involved in repair of oxidative DNA damage in Neisseria meningitidis, Mol. Microbiol. 83 (2012) 1064–1079, https://doi.org/10.1111/j.1365 2958.2012.07989.x.
[65] B. Landova, J. Silhan, Conformational changes of DNA repair glycosylase MutM triggered by DNA binding, FEBS Lett. 594 (2020) 3032–3044, https://doi.org/10.1002/1873-3468.13876.
[66] L.T. Da, J. Yu, Base flipping dynamics from an intrahelical to an extrahelical state exerted by thymine DNA glycosylase during DNA repair process, Nucleic Acids Res. 46 (2018) 5410–5425, https://doi.org/10.1093/nar/gky386.
[67] P.S. Thompson, K.M. Amidon, K.N. Mohni, D. Cortez, B.F. Eichman, Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link, Nat. Struct. Mol. Biol. 26 (2019) 613–618, https://doi.org/10.1038/s41594-019-0255-5.