

Mutagenic bypass of abasic DNA lesions during DNA and RNA synthesis in human cells

Dissertation zur Erlangung des Grades

„Doktor der Naturwissenschaften“

im Promotionsfach Pharmazie

am Fachbereich Chemie, Pharmazie, Geographie und Geowissenschaften

der Johannes Gutenberg-Universität in Mainz

Marta Rodríguez Alvarez

Geboren am 27 Mai 1990 in Oviedo, Spanien

July 2021

Abstract

1. Gutachter:

2. Gutachter:

Tag der mündlichen Prüfung

Table of contents

Abstract	1
Zusammenfassung	3
List of abbreviations.....	6
1 Introduction.....	7
1.1 Abasic DNA lesions: origin, structures and reactivity	7
1.2 Repair of abasic DNA lesions	8
1.3 Bypass of abasic DNA lesions during DNA synthesis: mechanistic insight and biological consequences	11
1.4 Transcriptional bypass of abasic DNA lesions.....	16
1.5 Methods for detection of mutagenic bypass of abasic DNA lesions.....	17
1.6 Scope of this work	19
2 Materials and methods.....	21
2.1 Materials.....	21
2.1.1 Bacteria	21
2.1.2 Human cell-lines	21
2.1.3 Media and supplements for cell culture.....	22
2.1.4 Expression vectors used in this report	23
2.1.5 Oligonucleotides for PCR, sequencing and reverse transcription of the EGFP coding region	23
2.1.6 Oligonucleotides for cloning of all pMR vectors used in this report.....	24
2.1.7 Oligonucleotides for screening of nucleotide substitutions that could lead to a non-fluorescent EGFP protein.....	25
2.1.8 Oligonucleotides for constructs preparation	26
2.1.9 Oligonucleotides for constructing vectors harboring single site-specific DNA lesions.....	26
2.1.10 Enzymes.....	27
2.1.11 Kits	28
2.1.12 Chemicals.....	28
2.1.13 Buffers and solutions	29
2.1.14 Software	29
2.2 Methods.....	30
2.2.1 Cloning of mutant EGFP pMR reporters containing point mutations that inactivate the protein fluorescence.....	30
2.2.2 Amicon filtration to eliminate excess of oligonucleotides	31
2.2.3 Preparation of ultracompetent bacteria	31
2.2.4 Transformation of ultracompetent bacteria.....	32
2.2.5 Mini and megaprep	32
2.2.6 Preparation of samples for sanger DNA sequencing.....	33

Abstract

2.2.7	Tandem nicking of plasmid DNA in transcribed DNA strand	33
2.2.8	Analytical ligation of nicked and gapped vectors	34
2.2.9	Gel electrophoresis.....	34
2.2.10	Strand depletion by incubation with competitor oligonucleotide.....	35
2.2.11	Preparative ligation	35
2.2.12	Thawing frozen human cells and cultivation.....	36
2.2.13	Transfection of human cells	37
2.2.14	Freezing human cell lines	37
2.2.15	Cell fixation with 1% formaldehyde	38
2.2.16	EGFP fluorescence analysis by flow cytometry.....	38
2.2.17	Detection of abasic sites in plasmid DNA by APE1 digestion	39
2.2.18	Characterisation of constructs containing uracil by UDG-Endo IV digestion assay.....	39
2.2.19	Characterisation of abasic DNA lesions by UDG-EndoIII/Fpg digestion assays	39
2.2.20	UDG treatment of uracil constructs to create natural abasic sites	39
2.2.21	DNA purification by phenol-chloroform.....	40
2.2.22	Detection of abasic sites in plasmid DNA by Endo IV digestion	40
2.2.23	Isolation of total RNA from mammalian cells.....	40
2.2.24	Verification of RNA integrity	41
2.2.25	Reverse transcription by gene specific primer (GFP)	41
2.2.26	PCR for quality control of cDNA samples.....	42
2.2.27	cDNA library preparation.....	42
2.2.28	Clean up of PCR products	43
2.2.29	Quality control of sequencing library samples	43
2.2.30	Pool preparation for sequencing	44
2.2.31	MiSeq system for library sequencing	44
2.2.32	Bioinformatic analysis	45
3	Results	47
3.1	EGFP reporters for sensitive detection of DNA lesions mutagenicity during DNA and RNA synthesis	47
3.1.1	Twenty-five different point mutations within the EGFP gene body lead to amino acid changes with the potential to abolish protein fluorescence.....	47
3.1.2	Successful insertion of the synthetic oligonucleotides with single base substitutions into the gapped pZAJ_5c vector.....	50
3.1.3	Eight of the base substitutions in the TS lead to a loss of the EGFP fluorescence	52
3.1.4	Validation of three EGFP mutants carrying inactivating single nucleotide substitutions.....	54

3.1.5	Secondary mutation events at the affected position reactivate EGFP fluorescence	56
3.1.6	Mutant pMR_S206* as an alternative reporter to study guanine modifications in the non-template strand.....	59
3.2	Mutagenic bypass of abasic DNA lesions during transcription	62
3.2.1	Generation of constructs accommodating BER-resistant AP lesion for detection of transcriptional mutagenesis in pZAJ_Q205*	62
3.2.2	Transcriptional bypass of repair resistant AP sites results in mutant mRNA	66
3.2.3	Transcriptional mutagenesis at AP sites is enhanced in the absence of nucleotide excision repair (NER).....	70
3.2.4	Lack of TM induced by physiological AP sites indicates that this lesion may be repaired within the cells.....	73
3.2.5	Transcriptional bypass analysis of uracil and abasic ribonucleotide lesions	81
3.3	Mutagenic bypass of abasic DNA lesions during DNA synthesis.....	85
3.3.1	Generation of reporter constructs to detect mutagenic TLS over an abasic lesion	85
3.3.2	Translesion synthesis over tetrahydrofuran abasic lesion is mutagenic .	89
3.3.3	Translesion synthesis over natural abasic lesion results in high mutation frequency.....	92
3.3.4	Although highly mutagenic, synthetic and natural abasic sites are bypassed in different manners during TLS.....	96
3.3.5	Mutagenic TLS over synthetic and natural abasic lesions shows higher adenine incorporation in absence of Rad18	105
3.3.6	Absence of Pol η and Rev1 TLS polymerases reduces the mutagenic bypass of synthetic abasic sites in human cells	108
3.4	Mutation profile of abasic DNA lesions during replicational and transcriptional bypass.....	110
3.4.1	RNA sequencing library preparation	110
3.4.2	Adenine is the most frequent ribonucleotide incorporated opposite synthetic abasic sites during transcriptional bypass	113
3.4.3	Mutagenic TLS over synthetic and natural AP lesions results mostly in adenine incorporation	119
3.4.4	Adenine is the most frequent nucleotide incorporated opposite synthetic AP sites during TLS followed by cytosine, thymine and guanine.....	121
3.4.5	The mutation profile of natural AP sites during TLS is A>G>C>T.....	123
4	Discussion	127
4.1	A novel approach for direct and sensitive detection of mutagenic bypass of DNA lesions	127
4.2	Transcriptional bypass of synthetic abasic lesion leads to adenine misincorporation into mutant mRNA	131

Abstract

4.2.1 Nucleotide excision repair (NER) works as a backup repair mechanism for synthetic AP sites in human cells.....	135
4.3 Translesion synthesis over abasic lesions mostly leads to adenine incorporation; however, mutation profile differs between synthetic and natural AP sites	137
References	145
Appendix I: Supplementary figures and tables	159
Appendix II: Sequences of plasmid vectors in FASTA format.....	167
Appendix III: Publications in peer-reviewed journals which included data from this work.	172
Index of figures	173
CURRICULUM VITAE	177
DECLARATION OF AUTHORSHIP	178

Abstract

DNA molecules are constantly damaged by exogenous agents and endogenous cellular processes. The balance between repair and tolerance of the resulting DNA lesions is essential for the maintenance of genome integrity. DNA damage tolerance mechanisms allow cells to continue replication and transcription in the presence of damage. Unavoidably, lesion bypass by RNA or DNA polymerases comes at the expense of fidelity. During translesion DNA synthesis (TLS), specialised DNA polymerases enable replication opposite and beyond DNA lesions. However, their elevated rates of nucleotide misincorporation turn TLS into the major source of spontaneous mutations in human cells. Erroneous bypass of DNA lesions may also occur during transcription, promoting the generation of mutant transcripts in a process termed transcriptional mutagenesis (TM). The repeated translation of the mutated mRNAs can lead to the production and accumulation of abnormal proteins with pathogenic properties.

Abasic DNA lesions are extremely frequent in the genome and bear a huge miscoding potential due to their inherent lack of genetic information. Despite the abundance of this type of damage, current knowledge about its mutagenic properties is far from complete and is generally based on biochemical data obtained in cell-free systems as well as genetic evidence in prokaryotes and lower eukaryotes.

This work describes an innovative system for the direct detection of mutagenic bypass events via reporter reactivation assays in human cells. We design and generate a set of mutated EGFP reporters that encode for a non-fluorescent protein. The erroneous bypass of a specific lesion placed in the mutated nucleotide leads to the synthesis of a fluorescent EGFP protein that serves to quantify the miscoding capacity of the lesion. This approach is further applied in the characterisation of the erroneous bypass of abasic DNA lesions during DNA and RNA synthesis. Once the lesion is classified as mutagenic, analysis of the transcripts via RNA sequencing allows the determination of its mutation profile.

By placing repair-resistant abasic site analog tetrahydrofuran in the transcribed strand of these newly designed reporters, regain of EGFP fluorescence indicated that transcriptional mutagenesis occurs at the site of damage. The percentage of cells showing green fluorescence was a direct indication of TM events and it varied from 11% to 79% depending on the DNA repair capacity of the cell. Subsequent RNA sequencing analysis revealed that RNA pol II exclusively incorporates adenine at the lesion site. By using this novel TM assay as a tool to investigate repair, NER involvement in the repair of synthetic abasic lesions was shown for the first time in human cells.

To detect mutations introduced during DNA synthesis opposite AP sites, this lesion was incorporated at the mutated position in the non-transcribed strand of the reporter and opposite to a gap. After transfection into several human cells, regain of EGFP fluorescence indicated that bypass of natural abasic sites, as well as their synthetic counterpart tetrahydrofuran, occurred mostly through adenine incorporation. Surprisingly, under the same conditions, natural abasic sites exhibited a different mutation profile as compared to tetrahydrofuran. While results showed that DNA polymerases mostly inserted adenine followed by guanine, cytosine, and thymine opposite natural AP sites, the incorporation of guanine was nearly absent opposite synthetic abasic lesions. This illustrates the importance of analyzing all chemical forms of a single adduct, which might lead to different mutagenic outcomes during DNA or RNA bypass events.

The miscoding property of abasic DNA lesions presents a challenge for the fidelity of replicational or transcriptional bypass processes. Most of these lesions arise from depurination reactions and cytosine deamination. Therefore, adenine leading the mutation pattern induced by these lesions implies that most AP sites are likely to be mutagenic. Thus, abasic lesions become a key source of mutagenesis in human cells, which entails important biological consequences. On one hand, it is likely that AP sites derived mutagenesis works as an important source of genetic diversity during evolution. On the other hand, mutations arising from the bypass of these lesions may play a critical role in cancer development and aggravation of degenerative diseases. In future studies, we can extend the established TM and TLS assays to investigate mutagenic bypass occurring at different DNA modifications. In addition, establishing the mutation pattern of additional DNA lesions can help to track the sources of mutations encountered in a broad range of human cellular backgrounds.

Zusammenfassung

Die DNA ist täglich exogene Agenzien und endogene zelluläre Prozessen ausgesetzt die Schäden im Genom verursachen können. Um die Integrität des Genoms aufrecht zu erhalten ist das Gleichgewicht zwischen Reparatur und Toleranz der resultierenden DNA-Läsionen essentiell. DNA-Schadenstoleranzmechanismen ermöglichen es den Zellen, Replikation und Transkription in Gegenwart von Schäden in der DNA-Matrize fortzusetzen, wobei eine Reduktion der Genauigkeit dabei unvermeidlich ist. Während der Translasiations-DNA-Synthese (TLS) ermöglichen spezialisierte DNA-Polymerasen die Replikation gegenüber von und über DNA-Läsionen hinweg. Die erhöhte Rate an Nukleotid-Fehleinbauten von TLS-Polymerasen eine Hauptquelle für spontane Mutationen in menschlichen Zellen. Eine fehlerhafte Umgehung von DNA-Läsionen kann auch während der Transkription auftreten. Dies fördert die Entstehung von Mutationen in RNA-Transkripten in einem Prozess der als transkriptionelle Mutagenese (TM) bezeichnet wird. Die wiederholte Translation der mangelhaften mRNAs kann zur Produktion und Akkumulation fehlerhafter Proteine mit pathogenen Eigenschaften führen.

Abasische DNA-Läsionen (AP-Stellen) sind im Genom extrem häufig und bergen aufgrund ihres Mangels an genetischer Information ein enormes Potential zur Fehlkodierung. Trotz ihrer Häufigkeit ist das derzeitige Wissen über die mutagenen Eigenschaften von AP-Stellen bei weitem nicht vollständig und basiert hauptsächlich auf biochemischen Daten die in zellfreien Systemen gewonnen wurden und auf genetischen Nachweisen in Prokaryoten und niederen Eukaryoten.

Das Ziel dieser Arbeit war daher die Etablierung eines Verfahrens zum direkten Nachweis der fehlerhaften Umgehung von DNA-Schäden in menschlichen Zellen am Beispiel von AP-Stellen. Dieser Studie beschreibt das Design und die Optimierung eines innovativen Reporter-Reaktivierungs-Assays zur Detektion und Quantifizierung der fehlerhaften Umgehung von DNA-Läsionen. Dafür generierten wir ein Set von mutierten EGFP-Reportern, die für ein nicht-fluoreszierendes Protein kodieren. Durch das spezielle Design der Reporter führt die fehlerhafte Umgehung einer DNA-Läsion innerhalb des mutierten Nukleotids zur Expression eines fluoreszierenden EGFP-Proteins, was zur Quantifizierung der Fehlkodierung der Läsion verwendet werden kann. Nach seiner Optimierung wurde der Assay dafür verwandt die fehlerhafte Umgehung von AP-Stellen während der DNA- und RNA-Synthese zu charakterisieren. Wird die Läsion als mutagen eingestuft kann mittels RNA-Sequenzierung das Mutationsprofil der Transkripte bestimmt werden.

Die Platzierung von Tetrahydrofuran, ein reparaturresistentes Analog von AP-Stellen, im transkribierten Strang dieser neu entworfenen Reporter führte zur Wiedererlangung der EGFP-Fluoreszenz was beweist, dass transkriptionelle Mutagenese an der DNA Läsion stattfand. Die Berechnung des Anteils grün fluoreszierender Zellen wies darauf hin, dass TM-Ereignisse an AP-Stellen mit einer Häufigkeit von 11% bis 79% stattfinden, in Abhängigkeit von der DNA-Reparaturkapazität der Zelle. Eine anschließende RNA-Sequenzierungsanalyse ergab, dass RNA pol II ausschließlich Adenin an der Läsionsstelle einbaute. Durch die Verwendung dieses neuartigen TM-Assays zur Untersuchung der DNA-Reparatur konnte außerdem zum ersten Mal die Beteiligung der Nukleotid-Exzisionsreparatur an der Behebung synthetischer AP-Stellen in menschlichen Zellen gezeigt werden.

Um Mutationen die während der DNA-Synthese gegenüber AP-Läsion eingeführt wurden detektieren zu können, wurde eine AP-Stellen an der mutierten Position in den nicht-transkribierten Strang des Reporters und gegenüber einer Lücke eingebaut. Die Transfektion menschlicher Zellen zeigte die Wiedererlangung der EGFP-Fluoreszenz und dadurch die mutagene Umgehung von natürlichen AP-Stellen und synthetischen Tetrahydrofurans unter Einbau von Adenin. Überraschenderweise was das Mutationsprofil natürlichen AP-Stellen unter diesen Bedingungen ein anderes als das Mutationsprofil von Tetrahydrofuran. Während die Ergebnisse zeigten, dass DNA-Polymerasen gegenüber natürlichen AP-Stellen hauptsächlich Adenin, gefolgt von Guanin, Cytosin und Thymin einbauten, war der Einbau von Guanin gegenüber synthetischen natürlichen AP-Stellen fast nicht vorhanden. Dies zeigt wie wichtig es ist alle Formen eines einzelnen Addukts zu analysieren die bei DNA- oder RNA-Umgehungsereignissen zu unterschiedlichen mutagenen Ergebnissen führen können.

Das Potential natürlicher AP-Stellen zur Fehlkodierung stellt eine Herausforderung für die Genauigkeit von Replikation und Transkription dar. Da die Mehrzahl der AP-Stellen durch Depurinierungsreaktionen und Cytosindesaminierung entstehen kann davon ausgegangen werden, dass die Umgehung dieser AP-Stellen unter vorrangigem Einsatz von Adenin zumeist mutagen ist. Dies macht natürlichen AP-Stellen zu einer Hauptmutationsquelle in menschlichen Zellen, was wichtige biologische Konsequenzen nach sich zieht. Einerseits ist es wahrscheinlich, dass die durch AP-Stellen eingeführten Mutationen als wichtige Quelle der genetischen Vielfalt während der Evolution fungierten. Andererseits können Mutationen, die durch die Umgehung dieser Läsionen entstehen, eine kritische Rolle bei der Krebsentwicklung und der Verschlimmerung degenerativer Krankheiten spielen. In zukünftigen Studien kann der etablierte TM- und

Zusammenfassung

TLS-Assay dazu verwendet werden die mutagene Umgehung verschiedenen DNA-Modifikationen zu untersuchen. Darüber hinaus kann die Bestimmung des Mutationsmusters zusätzlicher DNA-Läsionen dabei helfen, die Quellen von Mutationen zu verfolgen, die in einem breiten Spektrum menschlicher Zellen auftreten.

List of abbreviations

AP	Apurinic/aprimidinic
APE1	AP endonuclease 1
BER	Base excision repair
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DDT	DNA damage tolerance
DPC	DNA protein crosslink
Ds-DNA	Double-stranded DNA
EGFP	Enhanced green fluorescent protein
EndoIII	Endonuclease III
Endo IV	Endonuclease IV
FACS	Flow-activated cell sorting
Fpg	Formamidopyrimidine DNA Glycosylase
ICL	Intra strand crosslink
MMR	DNA mismatch repair
mRNA	Messenger RNA
NER	Nucleotide excision repair
NTS	Non-transcribed strand
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PIP	PCNA-interacting peptide
PNK	T4 Polynucleotide Kinase
Pol	Polymerase
ROS	Reactive oxygen species
RPA	Replication protein A
SSB	Single-strand break
ss-DNA	Single-stranded DNA
S-THF	5'-phosphorothioate tetrahydrofuran
THF	Tetrahydrofuran
TLS	Translesion synthesis
TM	Transcriptional mutagenesis
TS	Template/transcribed strand
UDG	Uracil DNA glycosylase

1 Introduction

1.1 Abasic DNA lesions: origin, structures and reactivity

Apurinic/aprimidinic (AP) lesions, also known as abasic sites, are the most abundant lesions formed in human DNA under physiological conditions reaching the astonishing number of 18000 per human cell per day (Friedberg, 2005). They arise by spontaneous or enzymatic hydrolysis of the N-glycosidic bond that connects the nitrogenous base with the sugar-phosphate backbone (Lindahl, 1993; Loeb & Preston, 1986). Chemical depurination/depyrimidination leads to the loss of purine or pyrimidine bases forming deoxyribose residues. Thereby the generation of AP lesions inherently results in the loss of genetic information. The rate of spontaneous base cleavage is enhanced at low pH and/or elevated temperatures (Lindahl & Andersson Annika, 1972) where purine bases are released with a higher frequency than pyrimidines, being guanine the most vulnerable base of all (Loeb & Preston, 1986; Schaaper et al., 1983). In addition, nitrogen in position 7 of guanosine residues is especially sensitive to alkylating agents forming a variety of adducts which repair considerably contributes to the formation AP sites within the cells (Lhomme et al., 1999). Thus, the specific DNA N-glycosylases removal of damaged bases forms AP sites as downstream repair intermediates in the base excision repair (BER) pathway (Boiteux & Guillet, 2004; Lindahl, 1993). For example, deamination of cytosine results in uracil formation that, as a non-canonical base present in the DNA, is efficiently removed by uracil-DNA glycosylases (UDGs) leading to the formation of AP sites (Krokan et al., 2002). UNG1 and UNG2, SMUG1, TDG, and MDB4 are the mammalian UDGs expressed in human cells, all of which excise uracil from double-stranded (ds-DNA). In addition, UNG2 and SMUG1 have also been shown to perform uracil excision in single-stranded DNA (ss-DNA) (Kavli et al., 2002).

Abasic DNA lesions exist in the cellular milieu in equilibrium between two main different forms, the predominant closed-ring form (furanose) and the open-ring form (aldehyde) which only comprises 1% of the total (**Figure 1-1A**). The latter form is the most reactive, especially at high pH, being prone to undergo β -elimination that results in DNA strand break with the formation of α,β unsaturated aldehydes (Lhomme et al., 1999). Abasic sites are susceptible to free radicals and reactive oxygen species (ROS) producing a variety of oxidised deoxyribose fragments; for instance, oxidation of the C1' carbon forms a common oxidised abasic lesion known as 2-deoxyribonolactone (Klungland Arne & Lindahl Tomas, 1997; Zheng & Sheppard, 2004). Additionally, recent studies have observed that the aldehyde form can generate inter-strand crosslinks (ICLs) (Dutta et al., 2007; Semlow et al., 2016) and DNA-protein crosslinks (DPCs) (Mohni et al., 2019;

Quiñones & Demple, 2016). The reactivity of these lesions complicates their study in water solutions and thereby, the need for a stable chemical analogue emerged. The reduced 2-hydroxymethyl-3-hydroxytetrahydrofuran or tetrahydrofuran (THF) is a widely used structural analogue of the prevalent furanose form of abasic sites whose lack of an aldehyde group at C1 (**Figure 1-1B**) increases its stability and prevents its repair by β -elimination (Takeshita et al., 1987; Wilson III David M et al., 1995).

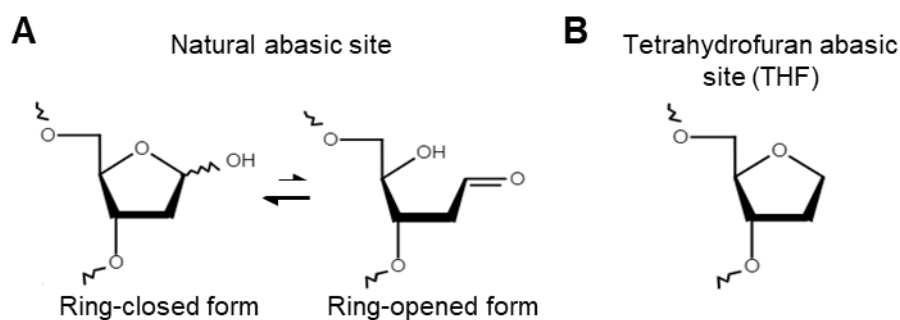


Figure 1-1: Chemical structure of abasic DNA lesions. (A) Equilibrium between the ring-closed (α,β isomer) (left) and ring-open form of a natural/physiological abasic site (right). (B) C1'-reduced abasic lesion known as tetrahydrofuran.

1.2 Repair of abasic DNA lesions

In order to maintain genome integrity, prokaryotic and eukaryotic cells have conserved a very efficient system to remove endogenous DNA damage from the DNA, which is known as the base excision repair pathway (BER). BER is a multistep repair mechanism where abasic sites are not only generated as intermediates of the repair process but also repaired. In humans, eleven DNA glycosylases have been described and account for the first step of BER that involves the search and removal of damaged DNA bases (Wallace, 2014). Specific DNA glycosylases are divided into two main groups, mono- or bi-functional DNA glycosylases depending on their catalytic mechanism.

Monofunctional glycosylases recognize thymine, uracil, and alkylated bases and cleave the N-glycosyl bond located between the base and the sugar (Lindahl, 1993; Loeb & Preston, 1986). The resulting AP site is further cleaved by AP endonuclease 1 (APE1) at the 5' phosphodiester bond creating a single-strand break (SSB) with 3'-OH and 5'-deoxyribose-5-phosphate (dRP) residues (Demple & Harrison, 1994) (**Figure 1-2**). From this point, BER may be further subdivided into two alternative pathways: short- and long-patch BER. In short-patch BER (SP-BER), the N-terminal DNA deoxyribosephosphodiesterase activity of DNA polymerase β (pol β) flips out and removes the sugar rest attached 5' to the nick (5'dRP). The additional polymerase activity of pol β synthesizes the missing nucleotide using the 3'-OH group (Matsumoto Y & Kim K, 1995) and the resulting nick is further sealed by DNA ligase III. During long-patch BER (LP-

Introduction

BER), displacement of 5'-dRP termini occurs and the 5'-flap endonuclease (Fen1) removes the 5'-overhanging structure replacing the dRP lyase activity of pol β . The DNA synthesis of the resulting 2-8 nucleotide patch takes place through pol β (or pol δ in the presence of PCNA) followed by ligation via DNA ligase I (Caldecott, 2020; Klungland Arne & Lindahl Tomas, 1997).

Alternatively, bifunctional DNA glycosylases have β -lyase activity in addition to the DNA glycosylase activity and remove oxidative bases from the DNA. They cleave the DNA backbone by incising 3' to the emerged AP site in a process designed as β -elimination resulting in the formation of a SSB with 3'- α,β unsaturated aldehyde (α,β). The 3'- α,β end can be further converted into a 3'-phosphate (P) terminus by a subsequent elimination reaction known as δ -elimination (Jacobs & Schär, 2012; Sun et al., 1995) (**Figure 1-2**). To obtain the canonical 3'-OH end required for further polymerization, either APE1 processes the 3'- α,β termini or polynucleotide phosphatase/kinase (PNKP) removes the 3'-P end resulting from AP lyase activity (Wiederhold et al., 2004). The remaining steps take place similarly as described for the monofunctional DNA glycosylases where DNA pol β mediates the gap-filling (Beard & Wilson, 2014) (**Figure 1-2**). *In vitro* excision assays performed with reconstituted SP- and LP- BER proteins have shown that LP-BER is strongly favored in the repair of oxidized and reduced AP sites (Klungland Arne & Lindahl Tomas, 1997). Particularly, the absence of the aldehyde group in THF abasic lesions after APE1 cleavage precludes dRP lyase activity of pol β and the resulting 5' blocking group is processed by FEN1. (Biade et al., 1998; Matsumoto Y & Kim K, 1995; Szczesny et al., 2008). On the contrary, natural abasic sites are generally repaired by short-patch BER where the N-terminal dRP lyase activity of pol β appears to be essential (Sobol RW et al., 2000).

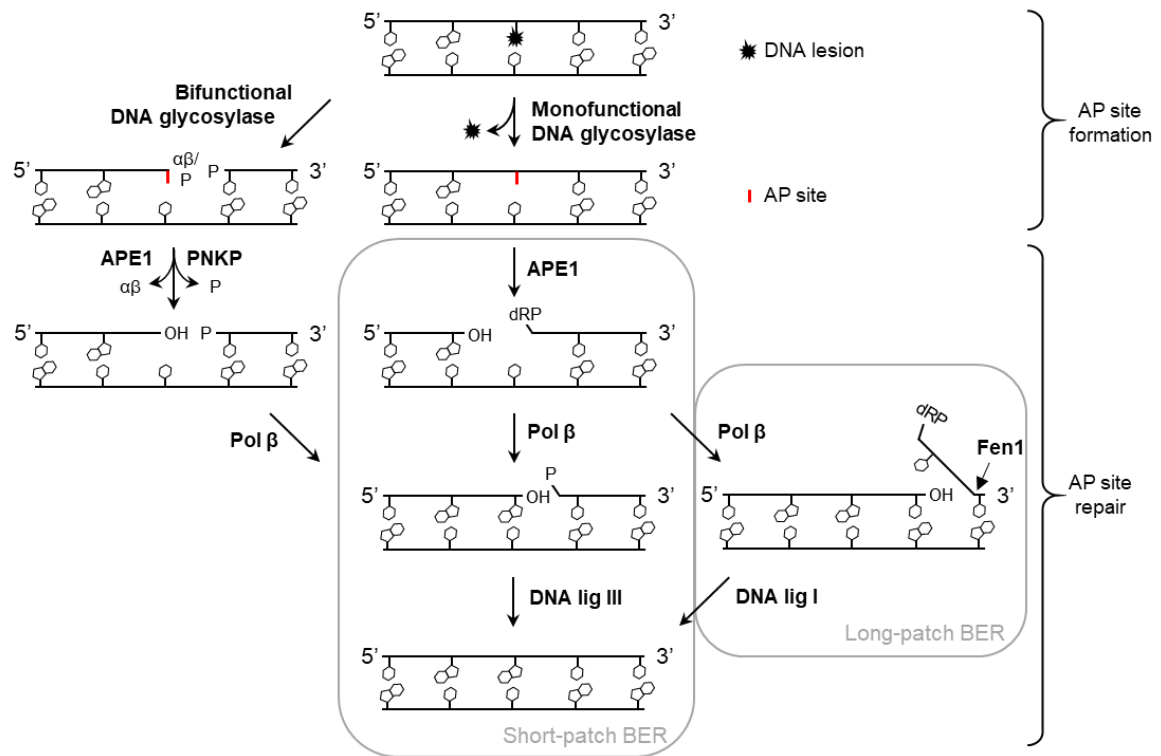


Figure 1-2: Abasic site formation and repair through mammalian short- and long-patch BER pathways.

Although BER is considered the primary defense against abasic DNA lesions, yeast strains depleted from major AP endonucleases activities are still viable (Swanson et al., 1999) and only showed mild sensitivity to AP site inducing agents such as MMS (Bennett, 1999; Torres-Ramos et al., 2000; Xiao W & Chow BL, 1998). Interestingly, depletion of APN1 and APN2 genes in combination with genes encoding for proteins involved in nucleotide excision repair (NER) pathway (either RAD1 or RAD10), are synthetic lethal thus suggesting a functional overlap between these pathways (Guillet Marie & Boiteux Serge, 2002). Since the bacterial multienzyme UvrABC complex functioning as NER was shown to incise AP lesions *in vitro* (Snowden A et al., 1990), several studies have proposed nucleotide excision repair (NER) pathway as a backup mechanism for the repair of oxidatively induced DNA damages in eukaryotes (Guillet Marie & Boiteux Serge, 2002; Kim & Jinks-Robertson, 2010). NER removes bulky lesions, such as those generated by UV irradiation, by introducing nicks 5' and 3' to the site of damage. Even though AP sites primarily do not alter the global B-form conformation of the DNA (Gelfand et al., 1998), it has been shown that NER is active in the removal of AP sites in yeast especially when converted into 3'-blocking ends (Guillet Marie & Boiteux Serge, 2002).

The XPA protein is a key factor in the NER pathway where it works as a scaffold protein that coordinates the action of the pre-incision complex to guarantee the proper elimination of the lesion (de Vries et al., 1995; Sugitani et al., 2016). The damage

Introduction

recognition step divides NER into two different sub pathways: global-genomic NER (GG-NER) that recognizes damage-induced distortions of the double helix via XPC protein (Sugasawa et al., 1998) and transcription-coupled NER (TC-NER) where CSB protein in cooperation with CSA recruits downstream TC-NER factors upon stalling of RNA polymerase II (RNA pol II) at the site of damage (Fousteri et al., 2006; Spivak, 2016). Based on the elevated rates of transcriptional associated mutagenesis observed in *S. Cerevisiae* due to unrepaired AP lesions in a NER deficient background, it was suggested that AP sites are substrates of NER but only when placed in the transcribed strand of active genes and therefore repaired by TC-NER (Kim & Jinks-Robertson, 2010). However, current evidence of the implication of NER in AP sites repair in mammalian cells remains elusive. The embryonic lethality of complete APE1 KO (Xanthoudakis et al., 1996) and the activation of apoptosis upon APE1 downregulation (Fung & Demple, 2005) along with the rapid action of BER have been major difficulties to study the potential overlapping of AP site repair pathways in human cells.

Considering that AP sites repair by BER or NER mechanism implies the formation of a SSB, it is reasonable that these pathways do not repair abasic sites in the context of ss-DNA, as that would lead to detrimental DSBs. The HMCES enzyme has been recently discovered to work as a sensor for AP sites located in ss-DNA. At the replication fork, HMCES crosslinks with the aldehyde in the ring-open form of an abasic site at 3' junction. Thus, the resulting DPC is further resolved by ubiquitin-mediated degradation preventing error-prone bypass mechanisms (Mohni et al., 2019).

1.3 Bypass of abasic DNA lesions during DNA synthesis: mechanistic insight and biological consequences

Even though AP lesions are rapidly removed from DNA, their abundance in the genome makes them likely to interfere with replication and transcription prior to their repair. Because AP sites generally induce strong blockage of DNA and RNA polymerases (Dean & Howard-Flanders, 1968; Locatelli et al., 2010; Tornaletti et al., 2006), organisms have developed strategies to replicate and transcribe opposite and beyond these lesions in order to avoid more detrimental damages. However, the non-coding character of AP sites confers a high mutagenic potential to these lesions whose bypass might lead to error-prone DNA or RNA synthesis, ultimately causing genome instability and thus, contribute to carcinogenesis (Brégeon & Doetsch, 2011; Friedberg, 2005; Lange et al., 2011).

Genomic DNA of eukaryotic cells is replicated by three members of the B-family polymerases, so-called replicative polymerases: alpha α , delta δ , and epsilon ϵ

(McCulloch & Kunkel, 2008). The pol α -primase complex initiates replication by providing an RNA primer of approximately 10 ribonucleotides long and synthesizes a DNA stretch of 20-30 deoxynucleotides (Copeland & Wang, 1993; Hübscher et al., 2002). Thus, pol α -primase complex supplies the 3'-end required for the elongation step that is performed by DNA pol δ and ϵ , in the lagging and leading strand respectively (Pursell & Kunkel, 2008). The tremendous accuracy of these polymerases relies on their extremely selective active site and their 3'-5' exonuclease activity that, as a proofreading mechanism, allows the removal of recently misincorporated nucleotides (Schmitt et al., 2009). Although replicative DNA polymerases commonly stall upon DNA damage encountering (Batista et al., 2009; Dean & Howard-Flanders, 1968), the replicative CMG helicase, unaware of the damage, continues to unwind the double helix leading to the formation of ss-DNA regions (Byun et al., 2005; Chang et al., 2006). Formation of ss-DNA coated by RPA (replication protein A) activates the cell cycle checkpoint and, if the stalled fork is not restored properly, might lead to the formation of deleterious DSBs and fork collapse ultimately leading to apoptosis (Batista et al., 2009; Zeman & Cimprich, 2014). To avoid such consequences, cells have developed two main DNA damage tolerance (DDT) strategies: template switching (TS) and translesion DNA synthesis (TLS). TS mechanism utilizes the newly synthesized sister chromatid as a template for the DNA synthesis over the lesion and therefore, it is considered an error-free bypass process. On the contrary, TLS enables the direct synthesis opposite and beyond DNA lesions employing specialised DNA polymerases. The relaxed fidelity of these polymerases and their lack of exonuclease activity might lead to an error-prone bypass that points to TLS as one of the main sources of spontaneous mutations in cells (Livneh et al., 2010; Prakash et al., 2005).

It is important to note that TS and TLS do not necessarily occur at the replication fork, instead, ss-DNA gaps left opposite DNA lesions and behind the replication fork might be eventually resolved by DDT mechanisms in a strategy termed post-replicative gap-filling (Daigaku et al., 2010). It was discovered in *S. cerevisiae* that TS and TLS pathways are regulated through different post-translational modifications of the proliferating cell nuclear antigen also known as PCNA. The E3 ubiquitin ligase Rad18 mediates the transfer of a ubiquitin protein from Rad6 (E2 conjugating enzyme) to the lysine 164 of PCNA. As a result, mono-ubiquitinated (mUb) PCNA interacts with Y-family TLS polymerases activating the TLS pathway (Stelter & Ulrich, 2003). Further polyubiquitylation of PCNA lys 164 by Ubc13 (E2) and Mms2 (E3) forms a K63-ubiquitin chain that activates TS (Hoege et al., 2002; Wilkinson et al., 2020) (**Figure 1-3**).

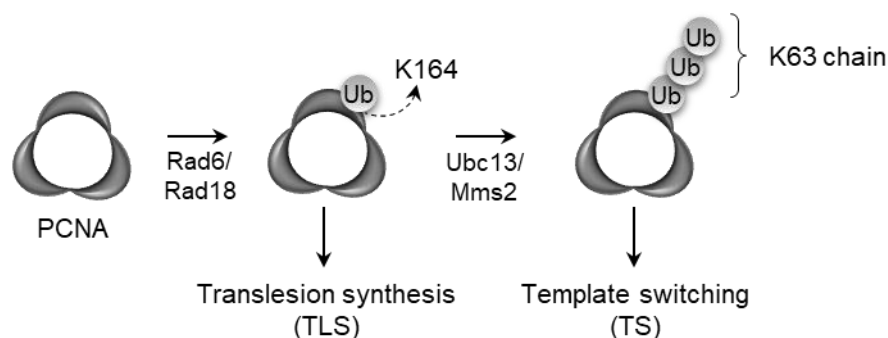


Figure 1-3: DTT regulation by post-translational modifications of PCNA. Modified from Hoege et al., 2002.

Even though both pathways (i.e TS and TLS) have been shown to perform oxidative damage bypass in eukaryotic cells, it has been proposed that a homology dependent repair, such as TS, is the bypass mechanism preferred by yeast (Swanson et al., 1999) while TLS is the DDT mechanism preferred by mammalian cells (Quinet et al., 2018). The need for mUb-PCNA via Rad18 for TLS activation in mammalian cells has been under debate during the past years since TLS events at photoproducts have been detected in absence of mUb-PCNA (Gervai et al., 2017; Hendel et al., 2011) and residual mUb-PCNA has been reported in absence of Rad18 (Simpson et al., 2006); however, following this research line, several studies have shown the essential role of PCNA monoubiquitination in the recruitment and stimulation of different TLS polymerases (Haracska et al., 2002; Watanabe et al., 2004; Wilkinson et al., 2020).

There are five mammalian DNA polymerases reported to be directly involved in TLS in human cells. The B-family polymerase zeta (pol ζ) and four Y-family polymerases: pol η (eta), ι (iota), κ (kappa) and Rev1 (Quinet et al., 2018; Vaisman & Woodgate, 2017). All the previously mentioned TLS polymerases have several common features: (1) due to their flexible catalytic sites that adapt to accommodate different DNA lesions these polymerases can synthesize DNA in presence of damage; (2) their lack of proofreading exonuclease activity contributes to the elongation from DNA mismatches and increases their mutagenic character and; (3) these TLS polymerases contain a short motif that directly binds to mUb-PCNA protein (PIP-boxes) which interaction allows their recruitment to the site of damage during replication (Diaz et al., 2003; Goodman, 2002; Lange et al., 2011). Interestingly, bypass of the lesion by TLS mechanism can occur in a one- or two-step process. In the one-step process, a single TLS polymerase inserts a nucleotide opposite to the lesion and continues the DNA synthesis. Alternatively, in the two-step process, one polymerase performs the base insertion and a second polymerase elongates the nascent DNA sequence starting from the inserted nucleotide (Prakash & Prakash, 2002). The choice between different DNA polymerases and one- or two-step TLS mechanisms

depends mostly on the type of damage (Quinet et al., 2016). For instance, Pol η , the best characterised of all TLS polymerases, accounts for the one-step accurate bypass of cyclobutene pyrimidine dimers (CPDs) induced by UV-radiation by inserting adenosine residues opposite to the lesion. Mutations in the POLH gene that encodes for Pol η , lead to the xeroderma pigmentosum variant (XP-V) syndrome where XP-V complementation group patients are more sensitive to sunlight exposure and susceptible to develop skin cancer (Johnson RE et al., 1999; Masutani et al., 1999). Even though pol η is capable of bypassing AP sites *in vitro* with a still unclear footprint (Choi et al., 2010; Patra et al., 2015), its role is not essential for this process as indicated by a study that reported the same level of AP site bypass for WT and XP-V cell-lines (Avkin et al., 2002). The B-family pol ζ possesses the highest fidelity among the TLS polymerases and thus, works mainly as an extender polymerase in the two-step TLS bypass (Johnson RE et al., 2000). The catalytic subunit of pol ζ , Rev3L, is considered a core player in the correct functioning of TLS and its impairment increases mutagenesis in mammalian cells (Shachar et al., 2009). Rev1 has a catalytic activity as deoxycytidine monophosphate (dCMP) transferase, which leads to the incorporation of cytosine residues opposite guanine, uracil, or AP sites (Chan et al., 2013; Gibbs & Lawrence, 1995; Lin et al., 1999; Nair et al., 2011). Still, this polymerase is unable to elongate DNA from the inserted nucleotide (Lange et al., 2011). The C-terminal domain of Rev1 interacts with other TLS polymerases including pol ι , pol κ , and pol ζ (Guo Caixia et al., 2003; Ohashi et al., 2004) suggesting that Rev1 might regulate the switch between inserter and extender polymerases (Friedberg, 2005; Sasatani et al., 2020). The absence of Rev1 and pol ζ leads to a significant, yet incomplete, reduction of AP sites bypass in human cells implying that both polymerases are important but not essential for this process (Weerasooriya et al., 2014). In contrast, pol κ was reported to be extremely ineffective in bypassing AP sites *in vitro* (Choi et al., 2010). Different studies suggested that this polymerase is specialised in the bypass of bulky adducts such as benzo(a)pyrene diolepoxide (BPDE) (Ogi et al., 2002) and N₂-N₂-guanine ICLs (Minko et al., 2008) where it also showed high processivity incorporating nucleotides to extend the resulting product (Vaisman & Woodgate, 2017). The remaining TLS polymerase, pol ι , is exclusively present in mammalian cells and it is considered the least accurate polymerase while reading through undamaged templates (Sharma et al., 2013). It can incorporate nucleotides opposite a high variety of DNA lesions including AP sites (McIntyre, 2020; Tissier A et al., 2000), yet with some discrepancies regarding the nucleotide insertion profile. Some studies reported thymine as the preferred nucleotide (Choi et al., 2010) while others reported a higher preference for guanine incorporation opposite abasic templates (Nair et al., 2009; Zhang et al., 2001) (**Table 1-1**).

Table 1-1: Human TLS polymerases and their implication in TLS over abasic DNA lesions. * (Patra et al., 2015); ** (Choi et al., 2010); *** (Nair et al., 2009; Zhang et al., 2001); **** (Lin et al., 1999).

Polymerase	Family	Main function	AP site bypass	Insertion opposite AP
zeta ζ	B	Extension polymerase from base mispairs	No	-
eta η	Y	dA insertion opposite CPDs and extension polymerase	Yes	A, G * T>A>C>G **
iota ι	Y	Insertion polymerase opposite uracil, AP sites and UV-radiation induced lesions	Yes	T ** G ***
kappa κ	Y	Insertion and extension polymerase opposite bulky adducts	No	-
rev1	Y	Scaffold protein for polymerase switching. dCMP activity	Yes	C ****

Since most chemotherapy treatments rely on the employment of DNA damaging agents to kill cancer cells, it has been proposed that TLS mechanisms can contribute to chemoresistance allowing the proliferation of cells in presence of damage (Knobel & Marti, 2011; Yamanaka et al., 2017). In this context, the search for inhibitors for TLS components, such as Rev1 or pol ζ, is emerging as a potential combinatorial treatment to increase tumor response to chemotherapy (Wenjie Wang et al., 2015; Wojtaszek et al., 2019).

It is worth mentioning that even though abasic sites have been reported to strongly block replicative polymerases (Batista et al., 2009; Dean & Howard-Flanders, 1968; Locatelli et al., 2010), numerous studies demonstrated the capacity of these polymerases to replicate through this lesion (Avkin et al., 2002; Choi et al., 2010; Mozzherin et al., 1997). Interestingly, a strong preference for adenine incorporation opposite to AP sites characterises the bypass via replicative polymerases in a mechanism known as “the A-rule” that seems to be conserved from bacteria to higher eukaryotes (Lawrence et al., 1990; Sagher & Strauss, 1983; Schaaper et al., 1983; Shibutani et al., 1997; Weerasooriya et al., 2014). Thus, treatment with aphidicolin, a well-known inhibitor of the B-family polymerases, in human cells transfected with a plasmid harboring a synthetic abasic site significantly impairs the overall bypass of the lesion (Avkin et al., 2002). Moreover, crystallization of the *Taq* polymerase from *Thermus aquaticus* incorporating adenine opposite to a synthetic AP site, led to the identification of a tyrosine residue in position 671 that is conserved among species and responsible for the geometric complementarity between the polymerase and the adenine nucleotide (Obeid et al., 2010). Although many studies extensively investigated the consequences of the

mutagenic bypass of AP lesions during DNA synthesis, the great amount of data described above is not sufficient to accurately characterise the mutation spectra of a single AP lesion undergoing mutagenic bypass nor the primary DNA polymerase responsible for this process in human cells.

1.4 Transcriptional bypass of abasic DNA lesions

As the main source of spontaneous mutations in cells, the bypass of DNA lesions during replication has been widely studied. Interestingly, lots of human cell types, such as neurons or myocytes, exist in a non-dividing state where transcriptional and translational events lay the ground for the viability and maintenance of the cell (Iyama & Wilson, 2013). However, the study of transcriptional bypass of abasic DNA lesions has captured much less attention. Erroneous bypass of DNA damage during transcription has been shown to have adverse consequences for cell physiology (Doetsch, 2002; Petrova et al., 2016; Viswanathan A et al., 1999). Upon encountering unrepaired lesions, RNA polymerase (RNA pol) arrest works as a damage recognition signal that activates a transcription-coupled response triggering the repair of the damage (Fousteri et al., 2006; Spivak, 2015). However, if the lesion does not lead to the persistent blockage of RNA pol II and activation of repair fails, transcriptional bypass might occur favoring the endurance of the lesion in the DNA. This response becomes critical in highly transcribed genes, where the access to the site of damage by the repair proteins is hampered by the constant activation of transcription (Liu et al., 1995; Saxowsky & Doetsch, 2006). Furthermore, the bypass of the damage by RNA polymerase might entail a nucleotide misincorporation opposite to the lesion leading to the generation of mutant transcripts in a process known as transcriptional mutagenesis (TM) (Doetsch, 2002). Repeated translation of these mutant transcripts would generate a whole population of abnormal proteins which, despite the transient and small scale of the events, bear the potential to disrupt the cellular physiology and have been linked to contributing to tumorigenesis in mammalian cells (Brégeon & Doetsch, 2011; Ezerskyte et al., 2018; Saxowsky et al., 2008).

Even though the mutagenic potential of AP sites is clear in the DNA synthesis context leading to permanent mutations, fewer studies have focused their research on the consequences of AP site bypass by transcriptional machinery. *In vitro* transcription assays have shown that T7 RNA pol (from the T7 bacteriophage), E.Coli RNA pol, as well as *S. cerevisiae* RNA pol II, are capable of synthesizing RNA through a template containing an AP site (Chen & Bogenhagen, 1993; Wei Wang et al., 2018; Zhou & Doetsch, 1993). However, conflicting results were obtained when using mammalian RNA polymerase II (RNAPII). While some studies stated that RNAPII is strongly blocked by

Introduction

an AP site placed in the transcribed DNA strand (Tornaletti et al., 2006), others reported a slow but efficient bypass of the lesion by the transcriptional machinery (Kuraoka et al., 2003). It is important to note that unlike the report by Tornaletti *et al.* where experiments were performed using RNAPII from rat liver cells and natural AP sites, Kuraoka *et al.* used purified RNAPII from HeLa cells and synthetic AP lesions (THF) in their template DNA which might explain the discrepancies within these results. Intriguingly, by using transcriptional mutagenesis assays in bacteria and yeast the ability of RNAP to bypass abasic DNA lesions was confirmed (Clauson et al., 2010; Kim & Jinks-Robertson, 2010). Unfortunately, a lack of consensus also exists when studying the mutagenic character of this lesion during TM (**Table 1-2**) since both adenine (Clauson et al., 2010; Wei Wang et al., 2018; Zhou & Doetsch, 1993) and cytosine (Kuraoka et al., 2003) incorporation have been reported. As a result, the molecular mechanism along with the mutation profile of TM through abasic sites demands further clarification.

Table 1-2: Summary of the literature presenting data about TM of abasic DNA lesions. AP: natural AP site; THF: tetrahydrofuran type AP site

RNA polymerase	Approach	Bypass	Type of AP site	Insertion opposite AP	Reference
T7 bacteriophage	<i>In vitro</i>	yes	THF	A > G	Chen & Bogenhagen, 1993
<i>E. Coli</i>	<i>In vitro</i>	yes	AP	A	Zhou & Doetsch, 1993
<i>E. Coli</i>	<i>In vivo</i>	yes	THF	A	Clauson et al., 2010
<i>S. cerevisiae</i>	<i>In vitro</i>	yes	THF	A	Wang et al., 2018
<i>S. cerevisiae</i>	<i>In vivo</i>	yes	AP	-	Kim & Jinks-Robertson, 2010
Rat hepatocytes	<i>In vitro</i>	no	AP	-	Tornaletti et al., 2006
HeLa	<i>In vitro</i>	yes	THF	C	(Kuraoka et al., 2003)

1.5 Methods for detection of mutagenic bypass of abasic DNA lesions

As stated in previous paragraphs, the study of the mechanism and biological consequences of AP sites bypass during both replication and transcription has been a scientific challenge. The main conflicting points are ascribed to a wide range of factors: First, even though *in vitro* studies have been very useful for mechanistic studies regarding the ability of DNA or RNA polymerases to bypass a specific DNA lesion (Choi et al., 2010; Kuraoka et al., 2003), they intrinsically exclude the resulting phenotypical outcomes derived from the bypass; second, the indispensable role of APE1 in higher

eukaryotes, strongly complicates the study of abasic sites prior to their repair. The chemical protection of the internucleotide linkage 5' to an abasic lesion with a phosphate backbone (phosphorothioate linkage), prevents its cleavage by AP endonucleases (Allgayer et al., 2016; Wilson III David M et al., 1995) and appeared as a promising tool to study AP sites repair in a BER-independent manner. Third, the use of either physiological or THF AP sites embedded in different sequence contexts leads to different patterns of nucleotide incorporation during TLS (Patra et al., 2015). Thus resulting in inconsistent data of the mutagenic profile of abasic lesions associated with their erroneous bypass; fourth, experiments in lower eukaryotes might not be the most accurate model to unravel the mechanism of TLS in mammalian cells which have at least two Y-family polymerases that are not present in yeast (Lange et al., 2011).

Several groups have used lesion-containing vectors transfected into mammalian cells to study TLS and TM of different DNA modifications (Avkin et al., 2002; Nagel et al., 2014; Petrova et al., 2016). The compelling method developed by *Avkin et al.* to quantify translesion synthesis in human cells demonstrated the importance of replicative polymerases for TLS of THF abasic sites and has been used in further TLS studies (Hendel et al., 2011; Shachar et al., 2009). At TM level, *Petrova et al.* designed an improved approach to study the phenotypical changes induced by transcriptional bypass over a 5-hydroxyuracil lesion. Thus, they demonstrated that reporter-based strategies are among the most promising techniques to study the mutagenic bypass of specific DNA lesions in mammalian cells. However, both methods require laborious techniques that involve the shift to a bacterial model and the need for a high yield of lesion-containing vectors. In order to find a simpler method that enables the study of not only TM but also TLS in human cells, we aimed to a fluorescence-based approach where mutagenic bypass of a site-specific DNA lesion would directly reactivate a fluorescent phenotype previously disrupted by a point mutation. This principle has been used to study the repair capacity of several human cell lines towards a defined lesion using transcriptional mutagenesis as a tool (Nagel et al., 2014). Since we aimed for a very sensitive approach, the stable EGFP (enhanced green fluorescent protein) appeared as a suitable candidate for the design of reporter reactivation assays. No cofactor is needed for excitation of EGFP protein whose fluorescence can be easily detected by microscopy or fluorescence-activated cell sorting (FACS) with a broad dynamic range (Cormack et al., 1996). Thus, EGFP reporters have been widely used in host cell reactivation (HCR) assays to study repair and transcriptional outcomes of specific DNA lesions (Allgayer et al., 2016; Khobta et al., 2010; Kitsera et al., 2011; Lühnsdorf et al., 2014). The finding of an EGFP reporter vector that contains a target sequence for two commercially available

Introduction

nicking enzymes, Nb.Bpu10I and Nt.Bpu10I, allowed the site specific insertion of modified bases within the protein coding region (Lühnsdorf et al., 2012) (**Figure 1-4**).

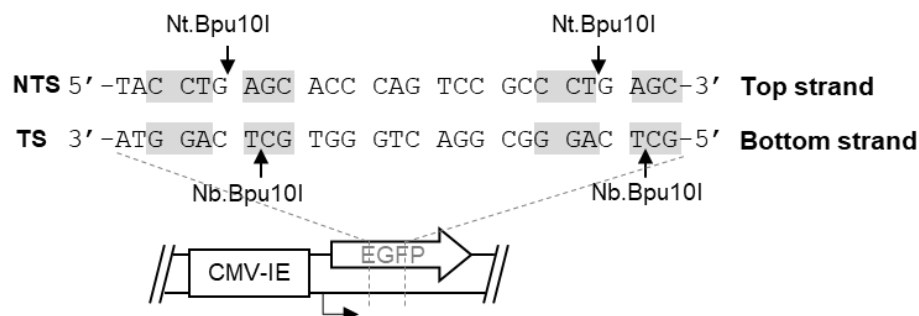


Figure 1-4: Schematic representation of EGFP reporter vector. Recognition nicking sites of the *Nb.Bpu10I* endonuclease (grey shadow in TS, bottom strand) and *Nt.Bpu10I* endonuclease (grey shadow in NTS, top strand), the arrows indicate the cutting position. CMV-IE is the strong promoter under whose control the reporter EGFP gene is located.

The possibility to modify the 18-nucleotides enclosed by the Bpu10I nicking sites opened the door to look for deleterious mutations that would affect the protein fluorescence in those positions. Later, the insertion of synthetic oligonucleotides (either in the transcribed (TS) or non-transcribed strand (NTS) of the gene) carrying abasic lesions at the mutated position, might serve as a suitable tool to detect mutagenic bypass of AP lesions in human cells.

1.6 Scope of this work

This investigation aimed to characterise the miscoding potential and define the mutation profile of abasic DNA lesions during replicational and transcriptional bypass in human cells. To this aim, it was necessary to generate a sensitive approach that allowed the direct detection of RNA and DNA synthesis errors caused by direct bypass of a specific DNA lesion. By analyzing the coding sequence of the EGFP gene, we hoped to find single point mutations that would encode for a non-fluorescent protein. It was essential that only that specific mutation eliminated the protein fluorescence while any alternative base substitution at the target position should revert to a fluorescent EGFP protein. Thus, we intended to employ the newly characterised mutations to design a set of mutated EGFP reporters and use them for reporter reactivation assays. After incorporating a site-specific abasic DNA lesion at the mutated nucleotide, subsequent detection of EGFP fluorescence would indicate that mutagenic bypass is occurring opposite to the lesion. Hence, allowing the direct visualization of mutations arising during DNA or RNA synthesis via FACS or microscopy analysis.

To detect erroneous transcriptional bypass of abasic lesions, BER-resistant AP sites will be incorporated at the mutated nucleotide on the transcribed strand of the newly design EGFP reporters. After transfection into human cells, the resulting EGFP fluorescence detected will be proportional to the transcriptional mutagenesis (TM) generated by the lesion. The mutagenicity of AP sites during DNA synthesis will be analyzed by generating constructs harboring the lesion at the mutated nucleotide in the non-transcribed strand of the mutated reporter opposite to a gap. Given that these reporters would not replicate in human cells, the gain of EGFP fluorescence observed in transfected cells would be a direct indication of the level of erroneous translesion synthesis (TLS) over the lesion.

By using this novel approach, we hope to define the miscoding capacity of AP sites in different human cell lines during replication and transcription. Subsequent analysis of transcripts resulting from TM and TLS events by RNA sequencing would help to clarify conflicting published data concerning the mutation pattern induced by this lesion.

2 Materials and methods

2.1 Materials

2.1.1 Bacteria

The bacteria strain used in this work was E.Coli SCS-8: *recA1 endA1 mcrAD(mcrBC-hsdRMS-mrr) Δ (argF-lac)U169 φ80dlacZ Δ M15, Tn10 (Tetr)*. Source: Agilent INC (Santa Clara, CA, USA)

2.1.2 Human cell-lines

Table 2-1: List of human cell lines used in this report.

Cell-line	Description	Origin
DLD-1	Human colorectal adenocarcinoma cells deficient in MSH6	Dr. Maja Tomcic-Christmann, Institute of Toxicology, Mainz
GM04312 (XP-A)	Human SV40 transformed. Skin fibroblasts, xeroderma pigmentosum complementation group A	NIGMS Human Genetic Cell Repository, Coriell Institute for Medical Research; Camden, New Jersey, USA
GM15876 (XP-A/ +XPA)	Human SV40 transformed. Skin fibroblasts, xeroderma pigmentosum complementation group A, transfected with the full-length cDNA of the XPAC gene	NIGMS Human Genetic Cell Repository, Coriell Institute for Medical Research; Camden, New Jersey, USA
GM15983 (XP-C)	Human SV40 transformed. Skin fibroblasts, xeroderma pigmentosum complementation group C	NIGMS Human Genetic Cell Repository, Coriell Institute for Medical Research; Camden, New Jersey, USA
GM16094 (CS-A)	Human SV40 transformed. Skin fibroblasts, Cockayne Syndrome type A	NIGMS Human Genetic Cell Repository, Coriell Institute for Medical Research; Camden, New Jersey, USA
GM16095 (CS-B)	Human SV40 transformed. Skin fibroblasts, Cockayne Syndrome type B	NIGMS Human Genetic Cell Repository, Coriell Institute for Medical Research; Camden, New Jersey, USA
HAP (c631)	Human haploid parental control cell line	Horizon Discovery Ltd, Cambridge Research Park, Waterbeach Cambridge, UK
HAP POLH KO	Human haploid POLH knockout cell line: 2bp deletion (Item number: HZGHC000655c010)	Horizon Discovery Ltd, Cambridge Research Park, Waterbeach Cambridge, UK
HAP POLK KO	Human haploid POLK knockout cell line: 4bp deletion (Item number: HZGHC000652c002)	Horizon Discovery Ltd, Cambridge Research Park, Waterbeach Cambridge, UK
HAP POLI KO	Human haploid POLI knockout cell line: 2bp deletion (Item number: HZGHC003483c012)	Horizon Discovery Ltd, Cambridge Research Park, Waterbeach Cambridge, UK

HAP Rev1 KO	Human haploid REV1 knockout cell line 28bp deletion (Item number: HZGHC000754c002)	Horizon Discovery Ltd, Cambridge Research Park, Waterbeach Cambridge, UK
HAP Rad18 KO	Human haploid RAD18 knockout cell line (Item number: HZGHC007527c001)	Horizon Discovery Ltd, Cambridge Research Park, Waterbeach Cambridge, UK
HeLa	Human cervical carcinoma cells	R. J. Wiesner (Institute for vegetables Physiology, University Cologne)
LoVo	Human colon derived from metastatic site cells deficient in MSH2	Maja Tomicic-Christmann and Markus Christmann
MRC5	Human SV40 transformed lung fibroblasts	NIGMS Human Genetic Cell Repository, Coriell Institute for Medical Research; Camden, New Jersey, USA
XP-F	Human XPF knockout cell line with complete loss of XPF using the CRISPR/Cas9 technique in fetal lung fibroblasts (MRC5Vi cells)	Lehmann et al., 2017

2.1.3 Media and supplements for cell culture

Material	Catalog number	Source	Location
DMEM (Dulbecco's Modified Eagle Medium), Gibco	41965-062	Life Technologies GmbH	Darmstadt, Germany
Effectene® Transfection Reagent	301425	Qiagen	
FBS (Fetal Bovine Serum), Gibco	10270-106	Life Technologies GmbH	Darmstadt, Germany
IMEM (Iscove's Modified Dulbecco's Medium), Gibco	12440-061	Life Technologies GmbH	Darmstadt, Germany
L-glutamin Solution (200mM)	G7513-100ML	Sigma Aldrich	
PBS (Phosphate Buffered Saline)	882126-12	Lonza	
Penicillin (10000U/ml)/ Streptomycin (10000µg/ml)	15140-122	Life Technologies GmbH	Darmstadt, Germany
Trypsin – EDTA Solution	T3924-100ML	Sigma Aldrich	Seelze, Germany

Materials and methods

2.1.4 Expression vectors used in this report

Vector	Description	Origin
pDsRed-Monomer-N1	Vector coding for DsRed-Monomer N1 used as a transfection marker	Clontech, Saint-Germain-en-Laye, France
pZAJ_5c	EGFP reporter vector containing Bpu10I nicking sites, kanamycin resistant.	From pEGFP-C3 vector backbone, (GenBank accession no. U57607)
pZAJ_Q205*	Modified pZAJ_5c with a point mutation in position c.613C>T coding for non-fluorescent EGFP, kanamycin resistant.	Johannes Burggraaff
pMR_Q205P	Modified pZAJ_5c with a point mutation in position c.614A>C coding for non-fluorescent EGFP, kanamycin resistant.	This Report
pMR_S206*	Modified pZAJ_5c with two subsequent point mutations in position c.617C>G c.618C>A coding for non-fluorescent EGFP, kanamycin resistant.	This Report
pMR_S206Y	Modified pZAJ_5c with a point mutation in position c.617C>A coding for non-fluorescent EGFP, kanamycin resistant	This Report
pMR_A207P	Modified pZAJ_5c with a point mutation in position c.619G>C coding for non-fluorescent EGFP, kanamycin resistant	This Report

2.1.5 Oligonucleotides for PCR, sequencing and reverse transcription of the EGFP coding region

Table 2-2: List of oligonucleotides used for PCR amplification of cDNA, Sanger sequencing (seq) and reverse transcription (RT). All oligonucleotides listed were ordered from Eurofins Genomics.

Oligonucleotide	Vector	Strand	Use	Sequence (5→3')
pEGFP-C3_1430-2	pZAJ, pMR	Reverse	Seq	TGTGGTATGGCTGATTATGA
pEGFP-ODC_ZA_1368-2	pZAJ, pMR	Reverse	RT	GCCTGTGCTTCTGCTAGGAT
PEGFP-C3_1153-1	pZAJ, pMR	Forward	PCR	GACCACTACCAGCAGAACAC
PEGFP-C3_1153-1 Rev	pZAJ, pMR	Reverse	PCR	GAACTCCAGCAGGACCATGT
pZA_Fwd_0	pZAJ, pMR	Forward	PCR	CTACACGACGCTCTTCCGATCTG ACCACTACC AGCAGAACAC

pZA_Fwd_1	pZAJ, pMR	Forward	PCR	CTACACGACGCTCTTCCGATCTA GGACCACTA CCAGCAGAACAC
pZA_Fwd_2	pZAJ, pMR	Forward	PCR	CTACACGACGCTCTTCCGATCTC TTGGACCAC TACCAGCAGAACAC
pZA_Fwd_3	pZAJ, pMR	Forward	PCR	CTACACGACGCTCTTCCGATCTT CATTGGACC ACTACCAGCAGAACAC
pZA_Rev_0	pZAJ, pMR	Reverse	PCR	GACGTGTGCTCTTCCGATCTGAA CTCCAGCAG GACCATGT
pZA_Rev_1	pZAJ, pMR	Reverse	PCR	GACGTGTGCTCTTCCGATCTTCG AACTCCAGC AGGACCATGT
pZA_Rev_2	pZAJ, pMR	Reverse	PCR	GACGTGTGCTCTTCCGATCTCGT TGAACTCCA GCAGGACCATGT
pZA_Rev_3	pZAJ, pMR	Reverse	PCR	GACGTGTGCTCTTCCGATCTATC GCGAACTCC AGCAGGACCATGT

2.1.6 Oligonucleotides for cloning of all pMR vectors used in this report

Table 2-3: List of oligonucleotides used for cloning of pMR_Q205P, S206*, S206A, S206Y and A207P mutant reporters after vector digestion with *Bpu10I* restriction enzyme. All oligonucleotides listed were ordered from Eurofins Genomics.

Oligonucleotide	Vector	Strand	Sequence (5→3')
nts.614C	pMR_Q205P	NTS	TGAGCACCCCGTCCGCC
ts.614G	pMR_Q205P	TS	TCAGGGCGGACGGGTGC
pMR_S205X NTS	pMR_S206*	NTS	TGAGCACCCAGTGAGCCC
pMR_S205X TS	pMR_S206*	TS	TCAGGGCTCACTGGGTGC
nts.616G	pMR_S206A	NTS	TGAGCACCCAGGCCGCC
ts.616C	pMR_S206A	TS	TCAGGGCGGCCTGGGTGC
nts.617A	pMR_S206Y	NTS	TGAGCACCCAGTACGCC
ts.617T	pMR_S206Y	TS	TCAGGGCGTACTGGGTGC
nts.619C	pMR_A207P	NTS	TGAGCACCCAGTCCCCC
ts.619G	pMR_A207P	TS	TCAGGGGGGACTGGGTGC

2.1.7 Oligonucleotides for screening of nucleotide substitutions that could lead to a non-fluorescent EGFP protein

Table 2-4: List of oligonucleotides used for the screening of nucleotide substitutions placed in the template strand of pZAJ_5c vector incubated with the strand-specific endonuclease *Nb.Bpu10I*. Each single nucleotide substitution is indicated with a bold letter. All oligonucleotides listed were ordered from Eurofins Genomics.

Oligonucleotide	Sequence (5→3')
ts.610C	TCAGGGCGGACTGGG C GC
ts.610G	TCAGGGCGGACTGGG G GC
ts.611T	TCAGGGCGGACTGG T TGC
ts.611C	TCAGGGCGGACTGG C TGC
ts.611A	TCAGGGCGGACTGG A TGC
Q204_T	TCAGGGCGGACT T GGTGC
Q204_C	TCAGGGCGGACT C GGTGC
Cloning EGFP_Q204 BTM	TCAGGGCGGACT A GGTGC
ts.Δ613	TCAGGGCGGACTGGTGC
ts.614C	TCAGGGCGGAC C GGGTGC
ts.614G	TCAGGGCGGAC G GGGTGC
ts.614A	TCAGGGCGGAC A GGGTGC
pZA_Bpu10Ibottom	TCAGGGCGGACTGGGTGC
ts.615G	TCAGGGCGG A TGGGTGC
ts.615A	TCAGGGCGG A ATGGGTGC
ts.616T	TCAGGGCGG T CTGGGTGC
ts.616C	TCAGGGCGG C CTGGGTGC
ts.616G	TCAGGGCGG G CTGGGTGC
ts.617T	TCAGGGCG T ACTGGGTGC
ts.617C	TCAGGGCG C ACTGGGTGC
ts.617A	TCAGGGCG A ACTGGGTGC
ts.619T	TCAGGG T GGACTGGGTGC
ts.619G	TCAGGG G GGACTGGGTGC
ts.619A	TCAGGG A GGACTGGGTGC
ts.620T	TCAGG T CGGACTGGGTGC
ts.620C	TCAGG C CGGACTGGGTGC
ts.620A	TCAGG A CGGACTGGGTGC

2.1.8 Oligonucleotides for constructs preparation

Table 2-5: List of oligonucleotides including controls for construct preparation and complementary oligos for strand depletion experiments. All oligonucleotides listed were ordered from Eurofins Genomics.

Oligonucleotide	Vector	Strand	Sequence (5→3')
pZA Bpu10I bottom	pZAJ_5c	TS	TCAGGGCGGACTGGGTGC
pZA Bpu10I top	pZAJ_5c	NTS	TGAGCACCCAGTCCGCCC
Cloning EGFP_Q204X BTM	pZAJ_Q205*	TS	TCAGGGCGGACTAGGTGC
Cloning EGFP_Q204X TOP	pZAJ_Q205*	NTS	TGAGCACCTAGTCCGCCC
Q204X NT compl	pZAJ_Q205*	NTS	GGGCGGACTAGGTGCTCA
Q204X NB compl	pZAJ_Q205*	TS	GCACCTAGTCCGCCCTGA
ts.614G	pMR_Q205P	TS	TCAGGGCGGACGGGGTGC
nts.614C	pMR_Q205P	NTS	TGAGCACCCCGTCCGCCC
ts.619G	pMR_A207P	TS	TCAGGGGGGACTGGGTGC
nts.619C	pMR_A207P	NTS	TGAGCACCCAGTCCCCC
pMR_A207P NTS compl	pMR_A207P	NTS	GGGGGACTGGGTGCTCA
pMR_A207P TS compl	pMR_A207P	TS	GCACCCAGTCCCCCCTGA
pMR_S205X NTS	pMR_S206*	NTS	TGAGCACCCAGTGAGCCC
pMR_S206X NTS compl	pMR_S206*	NTS	GGGCTCACTGGGTGCTCA
pMR_S206X TS compl	pMR_S206*	TS	GCACCCAGTGAGCCCTGA
nts.617A	pMR_S206Y	NTS	TGAGCACCCAGTACGCC

2.1.9 Oligonucleotides for constructing vectors harboring single site-specific DNA lesions

Table 2-6: List of oligonucleotides harboring site-specific DNA modification and HPLC purified. *: phosphorothioate linkage; X: tetrahydrofuran abasic site; F: 2' fluorinated; r: ribose sugar; TS: template strand; NTS: non-template strand

Name	Vector	Strand	Modification	Origin	Sequence (5→3')
1 THF	pZAJ_Q205X	TS	THF	Biospring	TCAGGGCGGACT X GGTGC
2 SF	pZAJ_Q205X	TS	S-THF	Biospring	TCAGGGCGGACT *X GGTGC
(1) dU	pZAJ_Q205X	TS	dU	Biospring	TCAGGGCGGACT U GGTGC
(2) SdU	pZAJ_Q205X	TS	S-dU	Biospring	TCAGGGCGGACT *U GGTGC
(3) SdUS	pZAJ_Q205X	TS	S-dU-S	Biospring	TCAGGGCGGACT *U* GGTGC
NTS THF	pZAJ_Q205X	NTS	THF	Biospring	TGAGCACCC X AGTCCGCCC
NTS SF	pZAJ_Q205X	NTS	S-THF	Biospring	TGAGCACCC *X AGTCCGCCC
pZAJ_q204x	pZAJ_Q205X	NTS	dU	Biospring	TGAGCACCC U AGTCCGCCC
c.613U					
c.613fU	pZAJ_Q205X	NTS	FU	Biospring	TGAGCACCC FU AGTCCGCCC

Materials and methods

pMR_Q205P c.614C>THF	pMR_Q205P	NTS	THF	Biospring	TGAGCACCCC X GTCCGCCC
c.614fU	pMR_Q205P	NTS	FU	Biospring	TGAGCACCCC f UGTCCGCCC
c.614U	pMR_Q205P	NTS	U	Biospring	TGAGCACCCC U GTCCGCCC
pMR_S206x c.617G>THF	pMR_S206X	NTS	THF	Biospring	TGAGCACCCCAGT x AGCCC
c.617fU- c.618A	pMR_S206X	NTS	FU	Biospring	TGAGCACCCCAGT f UAGCCC
c.617U- c.618A	pMR_S206X	NTS	U	Biospring	TGAGCACCCCAGT U AGCCC
c.617fU	pMR_S206Y	NTS	FU	Biospring	TGAGCACCCCAGT f UCGCCC
c.617U	pMR_S206Y	NTS	U	Biospring	TGAGCACCCCAGT U CGCCC
ts.619rG	pMR_A207P	TS	riboG	Gene Link	TCAGGG r GGGACTGGGTGC
ts.619SrG	pMR_A207P	TS	S-riboG	Gene Link	TCAGGG* r GGGACTGGGTGC
ts.619rTHF	pMR_A207P	TS	riboTHF	Gene Link	TCAGGG r XGGACTGGGTGC
ts.619rSF	pMR_A207P	TS	riboS- THF	Gene Link	TCAGGG* r XGGACTGGGTGC
ts.619THF	pMR_A207P	TS	THF	Biospring	TCAGGG X GGGACTGGGTGC
ts.619SF	pMR_A207P	TS	S-THF	Biospring	TCAGGG* X GGGACTGGGTGC
pMR_A207P c.619C>THF	pMR_A207P	NTS	THF	Biospring	TGAGCACCCCAGTCC X CCC
c.619fU	pMR_A207P	NTS	FU	Biospring	TGAGCACCCCAGTCC f UCCC
c.619U	pMR_A207P	NTS	U	Biospring	TGAGCACCCCAGTCC U CCC

2.1.10 Enzymes

Enzyme	Catalog number	U/ μ l	Source
Antartic phosphatase	M0289S	5	New England Biolabs
AP Endonuclease 1	M0282S	10	New England Biolabs
Bpu10I	R0649S	5	New England Biolabs
Endonuclease III	M0268S	10	New England Biolabs
Endonuclease IV	M0304S	10	New England Biolabs
Fpg	M0240S	8	New England Biolabs
Nb.Bpu10I	ER1681	5	ThermoFisher Scientific
Nt.Bpu10I	ER2011	5	ThermoFisher Scientific
PNK	EK0032	10	ThermoFisher Scientific
Q5 DNA polymerase	M0491S	2	New England Biolabs
rDNase	740963	lyophilized	Macherey-Nagel
RNase H	EN0201	5	ThermoFisher Scientific

T4 DNA ligase	EL0013	30	ThermoFisher Scientific
Uracil DNA glycosylase	M0280S	5	New England Biolabs

2.1.11 Kits

KIT	Catalog number	Origin
Amicon® Ultra - 0.5 ml Centrifugal Filter Devices 30k	UFC503096	Merck
Effectene Transfection Reagent 4x1 ml	301427	Qiagen GmbH
GeneElute™ HP Plasmid Miniprep Kit	PLN10	Sigma-Aldrich
GFX™ PCR DNA and Gel Band Purification Kit	GE28-9034-70	Merck
NucleoSpin® RNA	740955.50	Macherey-Nagel
Qiagen ® Plasmid Mega Kit	12181	Qiagen GmbH
Qubit™ dsDNA HS Assay Kit	Q32851	Thermofisher
Agilent High Sensitivity DNA kit	5067-4626	Agilent

2.1.12 Chemicals

Chemical	Origin
2-propanol ≥ 99.8%	Carl Roth GmbH
50x TAE buffer	Carl Roth GmbH
Agarose (ultrapure)	Invitrogen
Beta-mercaptoethanol (50 mM)	PanReac AppliChem
BSA (Bovine Serum Albumin) 10 mg/ml	New England Biolabs (NEB)
CaCl ₂ ≥ 99% (Calcium chloride)	Sigma Aldrich
DMSO 99.5% (Dimethylsulfoxid)	Carl Roth GmbH
EDTA (Ethylenediaminetetraacetic acid)	Sigma Aldrich
Ethanol 70%	Carl Roth GmbH
Ethanol 96%	Carl Roth GmbH
Ethidium bromide solution 1% (30 µg/ml)	Carl Roth GmbH
Formaldehyde 37%	Sigma Aldrich
Glycerol ≥ 98%	Carl Roth GmbH
HEPES ≥ 99.5% (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)	Sigma Aldrich
KCl ≥ 99.5% (Potassium chloride)	Carl Roth GmbH
MgCl ₂ ≥ 98.5% (Magnesium chloride)	Carl Roth GmbH
MgSO ₄ ≥ 98.5% (Magnesium sulfate)	Carl Roth GmbH
MnCl ₂ (Manganese chloride)	PanReac AppliChem

Materials and methods

NaCl \geq 99.5% (Sodium chloride)

Phenol:chloroform-isoamyl alcohol

PIPES

Carl Roth GmbH

Sigma Aldrich

PanReac AppliChem

2.1.13 Buffers and solutions

Buffers/Working solutions	Composition
1x BEH-BSA	495 μ l BEH buffer + 5 μ l BSA 100x
1x LB	12.5 g LB, dissolve in 500 ml nanopure water and autoclave to sterilize
1x LB 1.5% Agar	25 g LB, 15 g Agar, dissolve in 1000 ml nanopure water and autoclave to sterilize.
1x TAE buffer	20 ml TAE 50x + 980 ml ddH ₂ O
2% formaldehyde	2.2 ml formaldehyde (37%) + 37.8 PBS
2x LB	6 g LB, dissolve in 200 ml nanopure water and autoclave to sterilize.
BEH	10 mM HEPES pH 7.5 + 200 mM NaCl + 1 mM EDTA
SOB	0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10m MgCl ₂ , 10 mM MgSO ₄ . Dissolve in nanopure water and autoclave to sterilize.
TB	10 mM PIPES, 15 mM CaCl ₂ , 250 mM KCl. Dissolve in nanopure water and adjust pH to 6.7 with NaOH or HCl, then add MnCl ₂ to 55 mM and adjust to final volume. Filter, sterilize and store at 4°C.
TE	10 mM Tris pH 8.0 + 1 mM EDTA

2.1.14 Software

Software	Source	Location
CellQuest™ Pro (Flow cytometry)	Becton Dickinson GmbH	Heidelberg Germany
Image Lab™ (gel band quantifications)	Bio-Rad Laboratories, Inc.	Hercules, USA
Cutadapt version 2.4	(Martin, 2011)	

2.2 Methods

2.2.1 Cloning of mutant EGFP pMR reporters containing point mutations that inactivate the protein fluorescence

To create new plasmid reporters, pZAJ_5c vector was always used as the backbone for the cloning procedure (**Figure 2-1**): 5µg of pZAJ_5c vector were incubated with 20.8U Bpu10I/µg DNA 1h at 37°C, followed by 20min at 80°C to heat-inactivate the enzyme. The restricted vector was further incubated with 2U Antarctic phosphatase/µg to dephosphorylate the excised oligonucleotides placed in between the nicking positions. To remove the original oligonucleotides, amicon filtration was performed as described in 2.2.2.

In a parallel reaction, annealing and phosphorylation of the synthetic oligonucleotides (**Table 2-3**) selected for cloning (inserts) were performed. For this reaction, 0.19 nmol of each oligonucleotide were used and incubated in 1x ligase buffer with 16.7 U of PNK according to the following program:

Step	Temp.	Time	Function
1	4°C	∞	
2	37°C	30 min	phosphorylation
3	95°C	10 min	
4	95°C-40°C	rate 0.1°C/s	annealing
5	50°C	1 min	
7	4°C	∞	

Then 50 ng from the linearized and filtered vector were used and incubated with 2 different concentrations of insert 1:1 (0.185 ng) or 1:3 (0.37 ng) and 8 U of ligase according to the following program:

Step	Temp.	Time
1	4°C	∞
2	22°C	1h
3	60°C	20 min
4	4°C	∞

Materials and methods

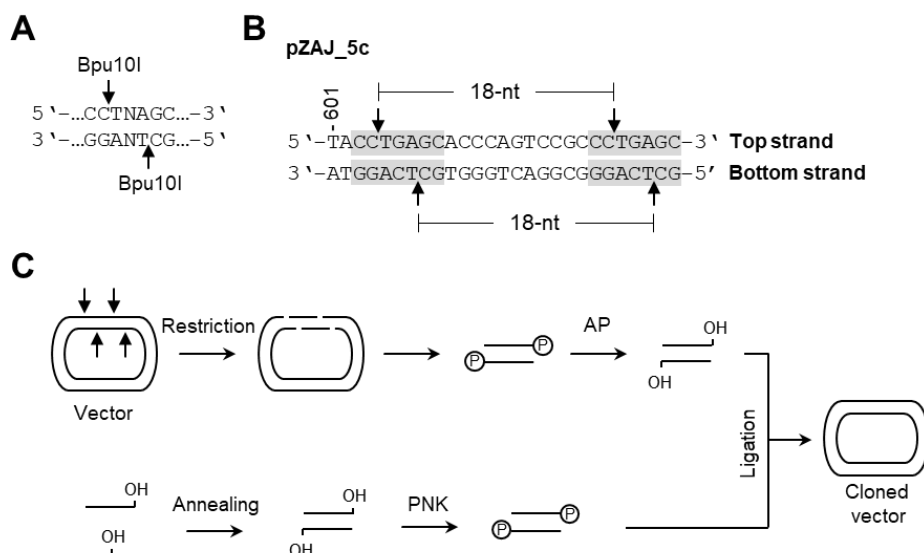


Figure 2-1: Cloning procedure followed in this report. (A) Recognition sites of Bpu10I restriction enzyme. **(B)** Recognition sites of Bpu10I in pZAJ_5c vector are shown with grey shadow, the arrows indicate the cutting position. The stretch of nucleotides between the cutting sites is 18 nucleotides long. **(C)** Out of scale scheme of the cloning procedure. AP (Antarctic phosphatase).

2.2.2 Amicon filtration to eliminate the excess of oligonucleotides

Amicon® Ultra-0.5 device was inserted into provided microcentrifuge tubes. First, the filter was washed with 500 µl sterile water (14.000 g. for 10 min) and the filtrate was discarded. A second wash with 500 µl of BEH buffer was performed (14.000 g. for 10 min) and the filtrate was discarded. Afterward, the Amicon® Ultra filter device was placed upside down into the microcentrifuge tube, spun at 1000 g. for 2 min, and the flow-through discarded. Afterward, up to 500 µl of the sample were added to the Amicon® Ultra filter device, filtrated (14.000 g. for 10 min) and the filtrate that now contains the oligonucleotides was discarded. The addition of 470 µl BEH into the filter was performed and mixed by pipetting up and down. Another centrifugation step (14.000 g. for 10 min) was performed and the filtrate was discarded. Then, Amicon® Ultra filter device was placed upside down into a new microcentrifuge tube and spun for 2 minutes at 1000 g. to transfer the concentrated sample from the filter to the tube. To get an approximate concentration of 100-200 ng/µl TE buffer was added to the concentrate and the concentration of the plasmid was measured by Nanodrop.

2.2.3 Preparation of ultracompetent bacteria

To prepare chemically competent bacteria glycerol stock of *Escherichia coli* SCS-8 was scratch to inoculate an overday culture 1x LB (1-2 ml) supplemented with tetracycline to a final concentration of 12.5 µg/ml in a sterile 50 ml falcon tube. This culture was inoculated into 120 ml SOB medium and grew at room temperature (18-22°C) until $OD_{600} = 0.4-0.6$ which usually takes approximately 16 hours. After incubation, the flask is

transferred to the ice for 10 minutes while shaking gently to cool the culture. The culture was further transferred to four 50 ml falcon tubes (30 ml x 4 = 120 ml) and bacteria were collected by centrifugation at 1800 g. for 10 minutes at 4°C. The supernatant was discarded and the pellet disrupted by adding 40 ml of ice-cold TB buffer and vortex gently. Buffer was added directly to one of the falcon tubes, when disrupted the suspension was transferred to the next falcon tube to compile the bacteria, this procedure was repeated in all falcon tubes. After a second centrifugation step, the bacteria were suspended in 10 ml of ice-cold TB buffer. DMSO was added to a final concentration of 7%, mixed gently and placed on ice for 10 minutes. The bacterial suspension was finally aliquoted: 550 µl into 2ml Eppendorf tubes, froze in liquid nitrogen, and stored at -80°C in a paper box.

2.2.4 Transformation of ultracompetent bacteria

Competent *E. coli* SCS-8 cells were transformed with the ligation reaction products. Ultracompetent bacteria were slowly thawed on ice for approximately 30 minutes. The transforming DNA obtained by cloning was added in a concentration up to 25ng per 50µl of competent cells. Bacteria and DNA were properly mixed by pipetting up and down previous of placing the tube/s on ice for another 20-30 minutes. Afterward, the suspension was transferred to a circulating water bath previously heated to 42°C (heat shock) and incubated exactly for 90 seconds. The suspension was further transferred to the ice to allow the cells to cool for 1-2 minutes. The suspension was transferred to 600 µl of 2xLB medium and incubated at 37 °C on a shaking incubator at 4 g. for 45-60 min. 50 µl of transformed competent cells were plated onto an agar plate and distribute equally with glass beads. Invert the plates and incubate them at 37°C, transformed colonies should appear within 12-16 hours. For cloned constructs, the efficiency of the transformation procedure was performed by comparison of 1:1 and 1:3 (vector:insert) ratio samples and clones were selected only from plates with the highest number of individual colonies.

2.2.5 Mini and megaprep

For the plasmid DNA miniprep, 50 ml falcon tubes containing 5 ml 1x LB medium supplemented with kanamycin (30 µg/ml) antibiotic were prepared. Overnight cultures with three single colonies from each agar plate of transformed bacteria were prepared. The colony was taken with a pipette tip which was directly thrown into the 50 ml falcon tubes (one for each colony) and incubated for 16 hours (overnight) at 37°C in a rotary shaker at 2 g. The actual mini preparation of the Plasmid DNA was then performed using the Sigma Aldrich: GenElute™ Plasmid Miniprep Kit (maximum output 15 µg) and

Materials and methods

following the manufacturer's protocol with exception of the last step where enclosed elution buffer was substituted by 10 mM Tris-HCl solution (pH 8).

For the mega-preparation, two single colonies were selected directly from the agar plate using a pipette tip. The pipette tip was thrown into a 50 ml falcon tube containing 5 ml 1x LB medium supplemented with kanamycin (30 µg/ml; 1:1000). Incubate the culture for 8 hours on a rotary shaker at 120 rpm. Check the density of the cultures by eye and use the densest culture for the inoculation of 500ml 1x LB medium in a 2-liter flask. Incubate for a minimum of 16 hours at 37°C on a rotary shaker with ~200 rpm (overnight). The plasmid DNA was extracted using the Plasmid Mega-prep Kit from Qiagen using the manufacturer-recommended protocol. At the end of the preparation, dissolve the DNA pellet in 2 ml TE buffer.

The same procedure was followed for all pZAJ and pMR vectors used in this report.

2.2.6 Preparation of samples for Sanger DNA sequencing

The reaction mix was prepared into 200 µl PCR tubes RNase and DNase free with flat lids. 450 ng of plasmid DNA from the miniprep was mixed with 10 pmol of pEGFP-C3_1430-2 primer [table 2-2], filled up to a total volume of 7 µl (if needed with sterile water). To protect the tubes, suitable firm boxes were used and sent to the sequencing facility in a stable padded envelope. In our case, all sequenced plasmids were sent to StarSEQ GMBH, Johann-Joachim-Becher-Weg 30a, 55099 Mainz, Germany.

2.2.7 Tandem nicking of plasmid DNA in transcribed DNA strand

All pZAJ and pMR vectors were nicked in the template (bottom) strand with the help of the site-specific Nb.Bpu10I nicking endonuclease, as described previously (Kitsera 2019, Luhnsdorf B (2012) Anal Biochem). 50 µg of plasmid DNA were incubated with 1 U/µg of Nb.Bpu10I nicking endonuclease in a total reaction volume of 250 µl containing 1x R Buffer supplemented by the company and according to the following program:

Step	Temperature	Time
1	4°C	∞
2	37°C	2h
3	80°C	20 min
4	4°C	∞

This reaction can be scaled up to 200 µg of plasmid DNA if needed. For tandem nicking of plasmid DNA in the non-template strand (NTS), use the same procedure with site-specific Nt.Bpu10I nicking endonuclease.

2.2.8 Analytical ligation of nicked and gapped vectors

Analytical ligation experiment consisted of two consecutive steps, first, the phosphorylation and annealing of the synthetic oligonucleotides to the vector was performed, and second, the ligation of the oligonucleotides into the nicked or gapped vector. The annealing reaction was assembled on ice, in four PCR 200 μ l tubes containing 15 μ l of total reaction volume. 100 ng of the previously nicked (or gapped) plasmid were added to each tube together with 1x T4 ligase buffer (1.5 μ l). Then, synthetic oligonucleotides were added to the reaction (0.0575 pmol/ng of plasmid DNA*) which correspond to 180x excess into tubes number 3 and 4. Then, T4 Polynucleotide Kinase (PNK) was added in a concentration of 0.05 U/ng to tube number 4. All tubes are filled up with sterile water to a total volume of 15 μ l and incubated as indicated in the following program:

Step	Temperature	Time
1	4°C	∞
2	37°C	30min
3	80°C	10 min
4	4°C	∞

While annealing runs, diluted T4 DNA ligase buffer (1x) and 0.4U/ μ l of T4 DNA ligase were prepared. After annealing is completed, 5 μ l of T4 DNA ligase buffer (1x) were added to tube number 1. 2 U (5 μ l) of diluted T4 DNA ligase were added to tubes 2, 3 and 4. All samples were incubated as indicated in the following program:

Step	Temperature	Time
1	4°C	∞
2	37°C	1h
3	65°C	15 min
4	4°C	∞

*If analytical ligation of a gapped plasmid is performed, synthetic oligonucleotides were added in a concentration of 0.0144 pmol/ng of plasmid DNA which corresponded to 45x excess.

2.2.9 Gel electrophoresis

Electrophoresis gels for the analysis of plasmid DNA were always prepared with a concentration of 0.8% agarose. The agarose powder (0.8 gr) was dissolved in 100 ml 1X TAE buffer by heat incubation; once the agarose mix acquired a transparent color, it was cool down at room temperature for 5 minutes before adding 0.5 mg/L ethidium bromide. The solution was poured into the gel tray and polymerized for approximately 45min. 1X TAE buffer also served as a running buffer during the electrophoresis and contained 0.5

Materials and methods

mg/l ethidium bromide. Samples were prepared using between 50-100 ng of DNA mixed with 6x loading dye and charged into the gel wells where one of the wells contains GeneRuler 1 kb DNA ladder from Thermo Fisher. The gel runs for 50 minutes at an applied voltage of 80V. When the run was completed, the gel was briefly swirled in deionized water before taking the image with the Molecular Imager Gel Doc™XR+ (Bio-Rad) recorded and analyzed with the Image Lab software from Bio-Rad.

Exceptions: Samples that did not undergo ligation steps or samples from PCR reactions were run in ethidium bromide free gels (gels for PCR products have a concentration of 1.5% agarose) and buffers. After running was completed, the gel was stained for 10min in 1X TAE buffer containing 0.5 mg/l ethidium bromide.

2.2.10 Strand depletion by incubation with competitor oligonucleotide

10 µg of double-nicked plasmid were incubated with 0.2875 nmol/µg of plasmid (920x molar excess) of competitor synthetic oligonucleotide in a total volume of 500 µl and 1x T4 DNA ligase buffer according to the following program:

Step	Temperature	Time
1	4°C	2min
2	80°C	10min
3	4°C	∞

The reaction was followed by Amicon filtration as explained in 2.2.2 to get rid of the duplexes formed between the complementary synthetic oligonucleotides and the native DNA strand. This reaction can be scaled up if needed.

2.2.11 Preparative ligation

Each preparative ligation reaction contained 10 µg of gapped plasmid, 45x molar excess of synthetic oligonucleotides, 1x T4 DNA ligase buffer and 12.5U PNK/µg of gapped plasmid in a total volume of 700 µl. Before adding the PNK to the reaction, an aliquot of 100ng was separated to further verify the ligation reaction via agarose electrophoresis. Phosphorylation of the synthetic oligonucleotides was achieved as indicated in the following program:

Step	Temperature	Time
1	4°C	∞
2	37°C	30min
3	80°C	10 min
4	4°C	∞

Reactions were transferred to ice and supplemented with T4 DNA ligase in a concentration of 5 U/ μ g of the gapped plasmid. Ligation of phosphorylated oligonucleotides was achieved following the program below:

Step	Temperature	Time
1	4°C	∞
2	37°C	1h
3	65°C	15 min
4	4°C	∞

This reaction can be scaled up if necessary.

2.2.12 Thawing frozen human cells and cultivation

To thaw frozen aliquots of human cells (**Table 2-1**) from the liquid N₂, the cryovials were placed rapidly in a 37°C water bath. Once they acquired the liquid phase, the tubes were carefully cleaned with 70% ethanol before transferring them under the flow hood. 50 ml falcon tubes were labeled with each cell line and passage, 3 ml of pre-warmed fresh medium (DMEM or IMDM) were taken with 5 ml pipet and the same pipet was used to aspirate the cells. The medium with the cells was transferred to the already labeled 50 ml falcon tube and centrifuged at 1000 rpm for 5 minutes at 23°C. In the meantime, one T25 small flask per cell line was labeled with cell line and passage. After centrifugation is finished, the supernatant was discarded with a pipet without disturbing the cell pellet to get rid of the freezing medium. 5 ml of pre-warmed fresh medium were used to gently resuspend the cells and transferred to the T25 flask. Then the cells were placed in the incubator for 24h before exchanging the medium.

Incubation of cells was always at 37°C and 5% CO₂.

Table 2-7: Splitting conditions and the number of plated cells in 6-well plates 24h prior to transfection.

Cell-line	Medium	Splitting conditions	Plated cells
DLD-1	DMEM	1:3 every 2-3 days	4x10 ⁵
GM04312 (XP-A)	DMEM	1:3 every 2 days	4x10 ⁵
GM15876 (XP-A+XPA)	DMEM	1:3 every 2 days	4x10 ⁵
GM15983 (XP-C)	DMEM	1:3 every 2 days	4x10 ⁵
GM16094 (CS-A)	DMEM	1:3 every 2 days	4x10 ⁵
GM16095 (CS-B)	DMEM	1:3 every 2 days	4x10 ⁵
HAP (c631)	IMDM	1:10 every 2-3 days	3x10 ⁵
HAP POLH KO	IMDM	1:5 every 2-3 days	4x10 ⁵
HAP POLK KO	IMDM	1:5 every 2-3 days	4x10 ⁵
HAP POLI KO	IMDM	1:5 every 2-3 days	4x10 ⁵

Materials and methods

HAP Rev1 KO	IMDM	1:5 every 2-3 days	4x10 ⁵
HAP Rad18 KO	IMDM	1:3 every 3 days	4x10 ⁵
HeLa	DMEM	1:10 every 2-3 days	3x10 ⁵
LoVo	DMEM	1:3 every 2-3 days	4x10 ⁵
MRC5	DMEM	1:5 every 2-3 days	3x10 ⁵
XP-C	DMEM	1:3 every 2 days	4x10 ⁵

2.2.13 Transfection of human cells

The transfection reagent used for all transfections was Effectene® from Qiagen. Cells were propagated in DMEM high glucose medium supplemented with 10% fetal bovine serum (FBS). Twenty-four hours before transfection, cells were seeded on 6-well plates (**Table 2-7**) so they were 70-80% confluent by the time of transfection. Since the number of plated cells differs from one cell line to another, the specific number of plated cells is indicated in table 2-1. By the time of transfection, the medium in all 6-wells is replaced by 1.5 ml fresh medium. The transfection mixture for a single well contained:

Material	For diploid cell-lines	For HAP cell-lines
Construct	400 ng	400 ng
dsRed	400 ng	400 ng
EC Buffer	up to 100 µl	up to 100 µl
Enhancer	3.2 µl	3.2 µl
Effectene	4 µl	10 µl
Fresh medium	500 µl	500 µl

The transfection mix was added dropwise to the wells containing fresh medium and cells were incubated at 37°C, 5% CO₂ until fixation 24 hours later.

*HAP cells were transfected following the same protocol with two exceptions: First, the medium used was IMEM medium. Second, by the moment of transfection FBS-free IMEM medium was used, 4 hours after transfection, the medium was exchanged for 2 ml of IMEM medium supplemented with 10% FBS (**Supplementary figure 1**).

2.2.14 Freezing human cell lines

Cells were washed with PBS and trypsinized using 1.5 ml trypsin per T75 flask. When the cells were detached (checked under the microscope), 10 ml of pre-warmed fresh medium were used for resuspension. Cells were transferred to a 50 ml falcon tube previously labeled and centrifuges 1000 rcf for 5 minutes at 4°C. In the meantime, cryovial tubes were labeled with date and passage. After centrifugation, the supernatant was carefully discarded with a pipet without disrupting the cells pellet. Cells were resuspend in 7.5 ml of freezing medium (10% DMSO= 18 ml medium + 2 ml DMSO). 1.8

ml of cells were added to each cryovial tube. The tubes stayed for 2 days at -80°C; then they were transferred to liquid nitrogen and the excel sheet with relevant information was filled (AKKhobta; LAB; MATERIALS; StorageN2).

2.2.15 Cell fixation with 1% formaldehyde

The protocol was performed under the hood if possible (not essential). 2% Formaldehyde solution in PBS was prepared and placed on ice. Cells within the 6-well plates were taken from 37°C and checked under the microscope (shape, density and fluorescence). The medium of the cells was discarded and cells were washed with 2 ml cold PBS. Afterward, 0.5-0.7 ml of trypsin per well were added (to ensure even coverage of the flask bottom, gently tilt the flask to all sides). The cells were incubated at 37°C, 5% CO₂ for 2-5 min until they acquired rounded shape and were detached from the well. Cells were resuspended with 1 ml of medium, transferred into a new 2 ml Eppendorf tube previously labeled and left on ice immediately.

For the following steps, work was performed on the bench and as often on ice as possible. Cells were centrifugated for approximately 40 seconds (mini centrifuge, 2000 rcf). The supernatant was discarded, cells were resuspended in 1 ml of ice-cold PBS and vortex carefully. The centrifugation step was repeated and the supernatant was further discarded. The final volume for the fixation was approximately 1ml of 1% Formaldehyde. The pellet was resuspended in 0.5 ml ice-cold PBS, mixed carefully and then 0.5 ml 2% Formaldehyde were added before vortexing the solution. Then, cells can be used directly for FACS analysis or stored at 4°C until needed (storage is possible for several months).

2.2.16 EGFP fluorescence analysis by flow cytometry

To remove the fixing medium, the cells were centrifuged at 1.1 g. for 3 minutes and the supernatant was discarded. Cells were resuspended in cold PBS and placed on ice until FACS analysis. The expression levels of DsRed-Monomer and enhanced green fluorescent protein (EGFP) were measured by flow cytometry (FACSCalibur™ and CellQuest Pro software (Beckton Dickinson, GmbH, Heidelberg, Germany)). The flow cytometer instrument was calibrated using non-transfected cells and high expression of DsRed-Monomer protein applying an FL-2 threshold level of 30 as a marker for transfected cells. Therefore, the median EGFP fluorescence was determined by the cells expressing both red fluorescence in the FL-2 channel and green fluorescence in the FL-1 channel.

Materials and methods

2.2.17 Detection of abasic sites in plasmid DNA by APE1 digestion

To confirm the presence of AP lesions, 100 ng of the construct were incubated with 1 U of AP endonuclease 1 (APE1) in 1x NEB Buffer in a total reaction volume of 15 μ l. Samples were incubated for 30 minutes at 37°C followed by a heat-inactivation of the enzyme for 20 minutes at 65°C. All reactions were assembled on ice and the results were further analyzed by gel electrophoresis.

2.2.18 Characterisation of constructs containing uracil by UDG-Endo IV digestion assay

Detection of uracil in plasmid DNA was performed using Uracil DNA glycosylase (UDG) and endonuclease IV (Endo IV) in 1x BEH-BSA reaction buffer. 3 PCR-clean tubes containing 100 ng of plasmid DNA were prepared. The first sample contained only the reaction buffer, the second sample contained 2 U of Endo IV and the third sample contained 2 U of both UDG and Endo IV. Samples were incubated in a total reaction volume of 15 μ l for 30 minutes at 37°C followed by a heat-inactivation of the enzymes for 20 min at 65°C. All reactions were assembled on ice and the results were further analyzed by gel electrophoresis.

2.2.19 Characterisation of abasic DNA lesions by UDG-EndoIII/Fpg digestion assays

Characterisation of abasic DNA lesions in plasmid DNA was performed using Uracil DNA glycosylase (UDG) and Endonuclease III (Endo III) or Formamidopyrimidine DNA Glycosylase (Fpg) in 1x BEH-BSA reaction buffer. 3 PCR-clean tubes per construct containing 100 ng of plasmid DNA were prepared. The first sample contained only the reaction buffer, the second sample contained 2U of either EndoIII or Fpg and the third sample contained 2 U of both UDG and EndoIII/Fpg. Samples were incubated in a total reaction volume of 15 μ l for 60 minutes at 37°C followed by a heat-inactivation of the enzymes for 20 minutes at 65°C. All reactions were assembled on ice and the results were further analyzed by gel electrophoresis.

2.2.20 UDG treatment of uracil constructs to create natural abasic sites

Constructs with natural abasic sites were generated by in vitro incubation of uracil-containing constructs with uracil DNA glycosylase (UDG). 8 μ g of each construct were incubated together with 150 U of UDG into a final volume of 1000 μ l in 1x BEH-BSA buffer. The mix was distributed into 200 μ l PCR-clean tubes and incubated for 30 minutes at 37°C followed by a heat-inactivation of the enzymes for 20 minutes at 65°C. DNA was further purified by phenol-chloroform protocol.

2.2.21 DNA purification by phenol-chloroform

Each sample (V) was transferred to a 2 ml Eppendorf tube. 1 V of phenol: chloroform (1:1) was added to the sample, the lid was tightly closed and vortex for 20-30" until the mixture was uniform. The suspension was centrifuged for 4 minutes at 750 rpm to separate the three phases. The upper phase that contains the DNA was transferred into a new 2ml Eppendorf tube. To each sample we added: 0.1 V NaCl 5 M or Sodium acetate pH 5.2 + 2.5 V EtOH 100%. The solution was mixed by turning the tube upside down and left in the freezer at -20°C for 20 minutes. Afterward, the samples were centrifuged at maximum speed for 25 minutes at 4°C. The supernatant was carefully transferred to another 2 ml Eppendorf tube and 1 ml cold EtOH 70% was added to the pellet. The centrifugation step at maximum speed was repeated for 15 minutes at 4°C. Then, the supernatant was carefully removed with a pipette and the DNA pellet left to dry at room temperature for 5-10 minutes. In order to get the desired concentration TE buffer was added and the DNA was resuspended by vortexing (do not use a pipette). The sample was placed on ice for 10 minutes, and DNA concentration was measured by nanodrop and stored at -20°C.

2.2.22 Detection of abasic sites in plasmid DNA by Endo IV digestion

To confirm the presence of abasic lesions, 100 ng of each construct were incubated with 5 U of Endonuclease IV (Endo IV) in 1x NEB Buffer 3 in a total reaction volume of 15 µl. The samples were incubated for 30 min at 37°C followed by a heat-inactivation of the enzyme for 20 minutes at 65°C. All reactions were assembled on ice and the results were further analyzed by gel electrophoresis.

2.2.23 Isolation of total RNA from mammalian cells

RNA isolation was performed 24 hours post-transfection in six-well plates. The monolayer of cells was washed with cold PBS twice before adding the lysis buffer. 350 µl of lysis buffer supplemented with a reducing agent (β-mercaptoethanol) were added in a concentration 1:100 directly into the cells, the tip of the 1000 µl pipette was cut and used to suck out all viscous liquid within the well moving the pipette in circles. The viscous sample was transferred to a NucleoSpin® Filter Column supplemented in the RNA isolation kit NucleoSpin® RNA from Macherey-Nagel. Protocol for RNA purification was followed according to the supplier's instructions, including in-column treatment with DNase.

Materials and methods

2.2.24 Verification of RNA integrity

RNA integrity was verified by denaturing agarose gel electrophoresis. 1% agarose TAE gel was prepared in DEPC water. 4 µl of 2x loading dye were added to 400 ng of RNA sample in a safe-lock Eppendorf tube and mixed by pipetting up and down. Samples were denatured at 65°C for 15 minutes using a thermoblock. The samples were loaded into the gel and run 50 min, 80 V (new wide Bio-Rad chamber). For good quality samples: 28S RNA migrates between 2 and 3 kb; 18S between 1-1.5 kb; 28S/18S band intensity ratio should be 2:1.

2.2.25 Reverse transcription by gene-specific primer (GFP)

To synthesize cDNA from previously isolated RNA samples, we used the Superscript III first-strand synthesis system for RT-PCR from Invitrogen. Starting from 150 ng of RNA mixed with 10 mM dNTP Mix and 2 pmol of the gene-specific primer (pEGFP-ODC_ZA_1368-2). The mixture was heated to 65°C for 5 minutes to disassemble secondary structures in the mRNA and incubated at 4°C for at least 1 minute to allow the primers to anneal to the mRNA. In the meantime, two master mixes were prepared, one containing the reverse transcriptase (RT) necessary for the reverse transcription and another one (noRT) with all the components but lacking reverse transcriptase (noRT) to check for DNA contamination in our samples. The RT master mix was prepared according to purchaser recommendations:

Component	1 sample	10 samples
10X RT buffer	2 µl	20 µl
25mM MgCl ₂	4 µl	40 µl
0.1M DTT	2 µl	20 µl
RNaseOUT (40 U/ µl)	1 µl	10 µl
SuperScript III RT (200 U/µl)	1 µl	10 µl

Step 1 samples containing RNA, dNTPs, and primer (total volume: 12 µl) were mixed with correspondent master mix (10 µl) by pipetting gently up and down and collection at the bottom of the tube was performed by brief centrifugation. The samples were incubated at 55°C for 50 minutes and the enzyme heat-inactivated at 85°C for 5 minutes.

* This reverse transcription protocol has been also performed using random hexamers instead of GFP with successful amplification of the EGFP mRNA (**Supplementary figure 2**).

** For the preparation of samples included in the second library (**Supplementary table 2**) the reverse transcription was performed using random hexamers and 400 ng of RNA template.

2.2.26 PCR for quality control of cDNA samples

For this PCR, different dilutions from the cDNA samples synthesized during RT were used as templates. For RT samples, 1:30; 1:300; 1:3000 dilutions and for noRT samples, 1:3; 1:30; 1:300 dilutions. The forward primer was PEGFP-C3_1153-1 and the reverse primer used was PEGFP-C3_1153-1 Rev (**Table 2-2**). Then, 5 μ L of the PCR product were mixed with 5 μ L of Gel Loading Dye (6x) without SDS and run in a 1.5% agarose gel.

The PCR program runs as follows:

Step	Temperature ($^{\circ}$ C)	Time	Action	Function
1	98	5 min		polymerase activation
2	98	40 s		denaturation
3	60	40 s		annealing
4	70	1 min	to step 2, 30x	extension
5	70	2 min		

2.2.27 cDNA library preparation

The library preparation consists of the amplification of the cDNA samples by two subsequent PCR runs aiming to generate amplicons with different lengths and to add the barcodes (indexes) to identify each individual sample during the sequencing run.

In order to increase library complexity, the first PCR was prepared using 1:3 dilutions of the cDNA previously synthesized as a template and a mixture of four different forward (pZA_Fwd_0, pZA_Fwd_1, pZA_Fwd_2, pZA_Fwd_3) and for different reverse primers (pZA_Rev_0, pZA_Rev_1, pZA_Rev_2, pZA_Rev_3) that anneal with all templates (**Table 2-2**). The generation of shorter and longer amplicons helped to avoid the overlapping of sequences that might be further misinterpreted by the sequencing system.

Reagent	Final concentration
Q5 reaction buffer (10x)	5x
dNTPs [10mM]	200 μ M
Primer forward [10 μ M]	0.5 μ M
Primer reverse [10 μ M]	0.5 μ M
Q5 polymerase [2U/ μ l]	0.02U
H ₂ O up to a final volume of 25 μ l	

The second PCR was prepared by using 1 ng of purified amplicons obtained in the first PCR as a template. Master mixes were prepared for each different combination of indexes. In our case, 8 master mixes were prepared with different forward indexes and 12 master mixes with different reverse indexes supplied by the core facility of genomics

Materials and methods

in IMB. Up to 96 index combinations were possible and thus, each sample was amplified with an individual barcode (**Figure 2-2**).

	D701	D702	D703	D704	D705	D706	D707	D708	D709	D710	D711	D712
D501	o	o	o	o	o	o	o	o	o	o	o	o
D502	o	o	o	o	o	o	o	o	o	o	o	o
D503	o	o	o	o	o	o	o	o	o	o	o	o
D504	o	o	o	o	o	o	o	o	o	o	o	o
D505	o	o	o	o	o	o	o	o	o	o	o	o
D506	o	o	o	o	o	o	o	o	o	o	o	o
D507	o	o	o	o	o	o	o	o	o	o	o	o
D508	o	o	o	o	o	o	o	o	o	o	o	o

Figure 2-2: Indexes combinations for second PCR amplification. D50X: Correspond to eight forward primers including a different index indicated by the number placed instead of the X; D7XX: Correspond to 12 reverse primers including a different index indicated by the number placed instead of the XX;.o: Represents each different sample.

The PCR program for the 1st and 2nd PCR runs as follows:

Step	Temperature (°C)	Time	Action	Function
1	98	5 min		polymerase activation
2	98	40 s		denaturation
3	60	40 s		annealing
4	70	1 min	to step 2, 15x	extension
5	70	2 min		

2.2.28 Clean up of PCR products

For purification of PCR products, the GFXTM PCR DNA and gel band purification kit from GE healthcare (Merck) was used following the manufacturer instructions. Elution of samples that were used for sequencing was always performed using 20 µL of DNase/RNase free water.

2.2.29 Quality control of sequencing library samples

Quality control of the cDNA library comprised the measurement of two parameters: the high-sensitive accurate measure of the DNA concentration by Qubit and the fluorescent profiles of the content of the DNA sample by Bioanalyzer. 1 µL of each purified sample obtained after the 2nd PCR run was used for the QubitTM ds-DNA HS assay kit and measured by a Qubit® 2.0 fluorometer from thermofisher following the supplier instructions. To obtain the electropherogram profiles, 1 µL of each purified sample was used for the Agilent high sensitivity DNA kit and measured by the Agilent 2100 Bioanalyzer following the supplier instructions. Samples that depicted a good profile (**Figure 2-3A**) were used directly for the pool preparation while samples depicting any

kind of contamination (**Figure 2-3B**), as adapter dimer peaks, were additionally purified as described in 2.2.28.

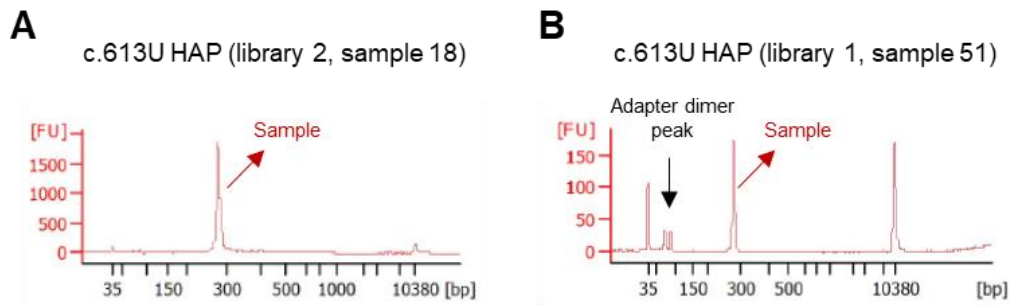


Figure 2-3: Electropherogram profiles of two representative samples obtained by the Agilent Bioanalyzer. (A) The ideal profile obtained from sample 18 of the 2nd library preparation depicting a high sample peak and two small peaks at 35 and 10380 bp that correspond to the lower and upper marker. **(B)** Profile contaminated with adapter dimers obtained from sample 51 of the 1st library preparation, this sample demanded a second round of purification.

2.2.30 Pool preparation for sequencing

To prepare the cDNA library pool all samples were mixed into a 1.5 ml Eppendorf tube and adjusted to a final concentration of 4 nM. The dilution was performed in a total volume of 200 μ l of RNase/DNase-free water.

2.2.31 MiSeq system for library sequencing

The MiSeq system from Illumina used for RNA sequencing was based on the so-called sequencing by synthesis explained in **Figure 2-4**. This system was chosen because it allowed the sequencing of fragments up to 300 bp and the amplicons of our library samples obtained after the PCR runs were between 250 and 270 bp long. A nano FC (flow cell) was used to load the samples of each library into the MiSeq. Samples were tagged with dual indexes as explained in **Figure 2-2** and the primers were pair-ended to allow the device to perform two reads from the same molecule (one forward and one reverse).

Materials and methods

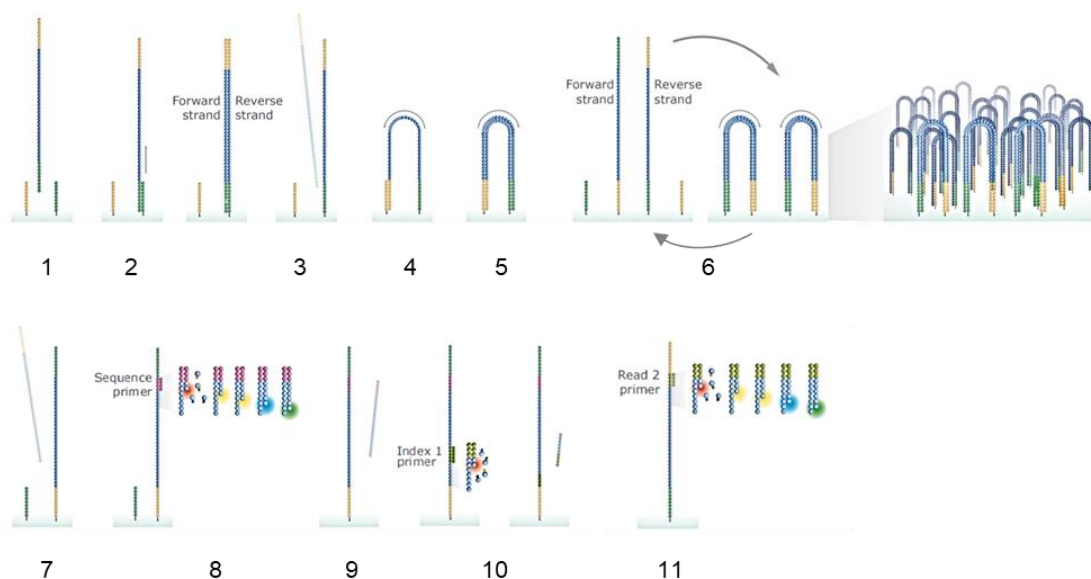


Figure 2-4: MiSeq sequencing by synthesis from Illumina. 1. Adapter hybridizes to flow cell; 2. Reverse strand synthesis; 3. Remove forward strand; 4. Fold over and hybridize to second primer; 5. Synthesizes second strand; 6. Bridge amplification where thousands of molecules are amplified in parallel; 7. The reverse strand is cleaved and washed away; 8. In each cycle incorporation of fluorescently tagged nucleotides from the forward primer occurs; 9. The read product is washed away; 10. Sequence index 1 and the read product is washed away; 11. The same process is followed with the reverse primer and the second read is sequenced. Modified from <https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/ForAllYouSeqMethods.pdf>.

2.2.32 Bioinformatic analysis

Libraries were sequenced in two flow cells in a MiSeq sequencer, outputting in total 4,216,022 of 150 base pair-long read pairs. These additional nucleotides had to be trimmed off along with library adapters. Reads were trimmed using Cutadapt version 2.4 (Martin, 2011), removing from both ends sequences corresponding to library adapters (<https://support.illumina.com/bulletins/2016/12/what-sequences-do-i-use-for-adapter-trimming.html>) along with the additional nucleotides. Adapter-free reads were then aligned to the reference amplicon sequence using bwa (Li & Durbin, 2009) mem's algorithm for paired-end reads. After alignment, a postprocessing step of the samples with internal mutation (**Supplementary table 4**) was carried out for each library: reads containing an unexpected mutation at positions 72 or 78 were filtered out, as they were attributable to EGFP contamination. Finally, for each library, the amplicon's per-base nucleotide composition was determined using the `coverage` function implemented in Bioconductor's (Huber et al., 2015) Rsamtools package (Morgan et al., 2021).

In addition, we tried to estimate the error rate introduced by the sequencer and the PCR amplification step. For this, we calculated the empirical distribution of the error rate in the first 62 bases of the amplicon, prior to the a-basic modification area, for which we don't expect any deviation from the amplicon's sequence. Libraries with the same lesion were added up to one unique distribution, which was then used to estimate the mean error

rate for every 10K bases sequenced, and the standard error of the mean. The sequencing results of both libraries performed in this work together with the raw data are in **Supplementary table 1**, **Supplementary table 2**, **Supplementary table 3** and **Supplementary table 4**.

3 Results

3.1 EGFP reporters for sensitive detection of DNA lesions mutagenicity during DNA and RNA synthesis

This part of the work aimed to generate a set of reporter vectors that encode for a non-fluorescent EGFP protein due to a single point mutation in the coding region of the gene.

3.1.1 Twenty-five different point mutations within the EGFP gene body lead to amino acid changes with the potential to abolish protein fluorescence

To find possible nucleotide changes within the EGFP gene that would lead to an alteration in the amino acid sequence of the protein, we used the pZAJ_5c reporter vector that harbors the EGFP coding sequence from the pEGFP-C3 expression vector. The essential component of this reporter is that contains Bpu10I restriction sites at the coding region of the gene, which facilitates the selective exchange of the enclosed 18 nucleotides in a strand-specific manner by the Bpu10I-derived nicking endonucleases. The Nb.Bpu10I nicking sites are located at the bottom strand and the Nt.Bpu10I nicking sites at the top strand starting from the nucleotide 607 of the EGFP coding region (**Figure 3-1**). Furthermore, this particular region encodes for a stretch of amino acids comprising the S203-L208 where T204 and S206 are directly involved in the transition between the neutral and ionized state of the fluorophore needed for fluorescence activation (Brejc et al., 1997). Thus, mutations within this sequence were likely to affect the protein fluorescence. In an ideal case, we hoped to find several point mutations involving different nucleotides (G, T, C or A) that would lead to a loss of EGFP fluorescence.

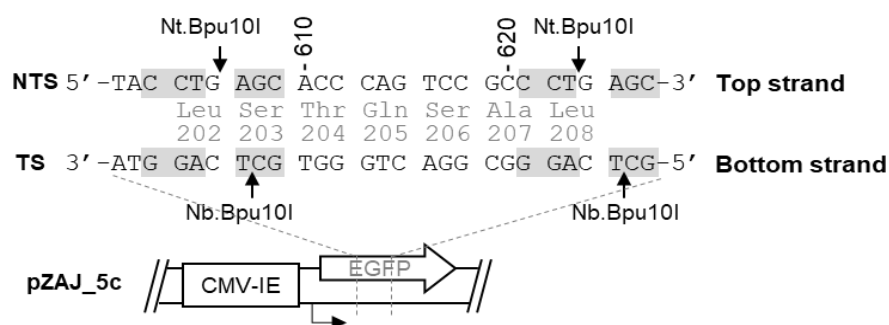


Figure 3-1: Schematic representation of pZAJ_5C vector encoding for EGFP protein. Recognition nicking sites of the Nb.Bpu10I (grey shadow in TS, bottom strand) and Nt.Bpu10I endonucleases (grey shadow in NTS, top strand), arrows indicate the cutting sites. The position in the coding strand of the first and last nucleotide analyzed are indicated with numbers above the sequence. The broken arrow shows the start of the transcription, rectangle indicates the CMV-IE promoter under whose control the reporter EGFP gene is located. The amino acid sequence is indicated between the strands.

Altogether, we studied 11 nucleotide positions (c.610-c.620) within 4 codons preventing the modification of the sequence recognized by the Bpu10I-derived nicking

endonucleases. Of the 33 potential single nucleotide substitutions, we predicted that 7 would not confer an amino acid change (silent mutations). By excluding these silent mutations and the c.613C>T nonsense transition mutation identified previously in the group (Master Thesis of Johannes Burggraaff, 2016), we selected 25 missense mutations for experimental screening. In summary, the screening involved seven cytosine, seven adenine, six thymine and five guanine substitutions within the 18-nucleotide stretch limited by the Bpu10I nicking sites. Among these potential mutations, we hoped to find one substitution for each different nucleotide that would lead to an amino acid change that would ultimately eliminate the protein fluorescence (**Table 3-1**).

Results

Table 3-1: Analysis of the EGFP coding sequence fragment flanked by Bpu10I sites to identify point mutations that lead to changes in the amino acid sequence of the EGFP protein. (*) Termination codon; (=) no amino acid change expected (silent mutations). Table adapted from Rodriguez-Alvarez et al., 2020.

Codon No. and Amino Acid	Codon 5'→3'	Nucleotide position	DNA (NTS)	Expected amino acid change	Selected for screening
204 Thr	ACC	610	<u>G</u> CC	T204A	√
			C <u>C</u> C	T204P	√
			<u>I</u> CC	T204S	
		611	A <u>A</u> C	T204N	√
			A <u>G</u> C	T204S	√
			A <u>T</u> C	T204I	√
		612	A <u>C</u> A	=	
			A <u>C</u> G	=	
			A <u>C</u> T	=	
		205 Gln	CAG	613	<u>A</u> AG
<u>G</u> AG	Q205E				√
<u>I</u> AG	Q205*				√
614	<u>C</u> GG			Q205R	√
	<u>C</u> CG			Q205P	√
	<u>C</u> TG			Q205L	√
615	<u>C</u> AA			=	
	<u>C</u> AC			Q205H	√
	<u>C</u> AT			Q205H	√
	<u>A</u> CC			S206T	√
206 Ser	TCC	616	<u>G</u> CC	S206A	√
			<u>C</u> CC	S206P	√
			<u>T</u> AC	S206Y	√
		617	<u>T</u> GC	S206C	√
			<u>T</u> TC	S206F	√
			618	<u>T</u> CA	=
		<u>T</u> CG		=	
		<u>T</u> CI		=	
		<u>A</u> CC		A207T	√
		207 Ala	GCC	619	<u>C</u> CC
<u>I</u> CC	A207S				√
<u>G</u> AC	A207D				√
620	<u>G</u> GC			A207G	√
	<u>G</u> TC			A207V	√
	<u>T</u> CC				
TOTAL			33		25

To perform a phenotypic screening of the candidate mutants, we designed a simplified approach starting from the functional EGFP reporter (pZAJ_5c). Essentially, this approach relies on the generation of a gap in the transcribed strand (TS) of the pZAJ_5c reporter. Then, the excised 18-nucleotide fragment will be substituted by 25 synthetic oligonucleotides each harboring one of the single nucleotide substitutions identified in **Table 3-1**. After transfection into mismatch repair-deficient cells (to prevent the repair of the nucleotide substitutions), we will analyze the relative EGFP fluorescence by FACS 24 hours after transfection. The nucleotide substitutions that would result in a loss of

EGFP fluorescence can be selected for cloning and further genetic and phenotypic validation (**Figure 3-2**).

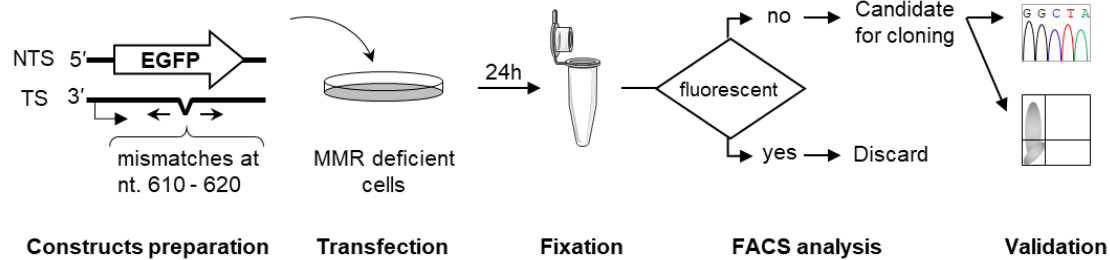


Figure 3-2: Schematic representation of the novel screening strategy followed in this section. 1. Preparation of 25 constructs containing single substitutions in the TS of the EGFP coding gene. 2. Transfection of the constructs into MMR deficient cells to avoid repair of single mismatches created after inserting the substitutions into the TS. 3. Fixation of cells 24 hours after transfection. 4. Selection of candidates for cloning by FACS analysis. Only the samples that lack EGFP fluorescence are selected. 5. Genetic (DNA sequencing) and phenotypic (FACS analysis) validation of the selected clones.

3.1.2 Successful insertion of the synthetic oligonucleotides with single-base substitutions into the gapped pZAJ_5c vector

To examine the potential of twenty-five point mutations to eliminate the EGFP fluorescence, we generated a set of hybrid EGFP constructs. Starting from the pZAJ_5c vector containing the innate non-template strand (NTS), we substituted a single nucleotide in the TS by inserting a synthetic oligonucleotide that created a mismatched pair (**Table 2-4**). By incubation with the strand-specific Nb.Bpu10I nicking endonuclease as described in methods section [2.2.7], we generated tandem nicks separated by 18 nucleotides at the TS of the gene comprising the nucleotides 608-625 of the EGFP coding sequence (**Figure 3-1**).

To verify that plasmid DNA was nicked at both sites, we performed an analytical ligation experiment [2.2.8] which was further analyzed by agarose gel electrophoresis [2.2.9] (**Figure 3-3A**). The previously nicked plasmid was incubated in four different samples: without reagents, in presence of T4 DNA ligase alone, T4 DNA ligase + synthetic oligonucleotide, or T4 DNA ligase + oligonucleotide + T4 Polynucleotide Kinase (PNK). Due to the single-strand breaks generated by the nicking endonuclease, the vector DNA was present in an open circular (oc) form (**Figure 3-3A, lane 1**). Upon incubation with T4 DNA ligase, the excised 18-nucleotide fragment religated with the original plasmid backbone reconstituting a covalently closed (cc) form (**Figure 3-3A, lane 2**). Incubation with a 180-fold excess of synthetic oligonucleotides led to the annealing of these oligonucleotides into the nicked vector displacing the original fragment. Because these oligonucleotides lack 5' phosphorylated ends, the ligation reaction was inhibited showing again an open circular (oc) form of the plasmid (**Figure 3-3A, lane 3**). If only one site

Results

was cut, the original fragment would religate in presence of the synthetic oligonucleotides and T4 DNA ligase and migrate as a cc form of the plasmid. Since that was not the case, this specific result confirmed the presence of two nicks in the original vector backbone. Subsequent incubation of the nicked pZAJ_5c vector with T4 DNA ligase, synthetic oligonucleotides, and PNK led to a covalently closed (cc) form of the plasmid (**Figure 3-3A, lane 4**). Thereby indicating that the synthetic oligonucleotides were correctly integrated into the vector DNA in presence of ligase and PNK (**Figure 3-3B**).

Once verified the pZAJ_5c vector was double nicked, the 18-mers fragment had to be depleted in order to generate a gap enclosed by the nicking sites in the TS of the vector [2.2.10]. This was achieved by melting out the excised fragment in the presence of a 920-fold molar excess of complementary synthetic oligonucleotides and rapid annealing. During this step, the excised 18-nucleotide single-stranded DNA fragments formed duplexes with the complementary synthetic oligonucleotides. As the hybrid duplexes and the remaining excess of complementary oligonucleotides were removed by Amicon filtration, a vector bearing an 18-nucleotide gap at the bottom strand was obtained and further verified by analytical ligation (**Figure 3-3C**). The vector remained in an open circular form even in the presence of T4 DNA ligase (**Figure 3-3C, line 2**), indicating that no original excised fragment remained in the preparation that could religate under these conditions and therefore, that the plasmid was perfectly gapped.

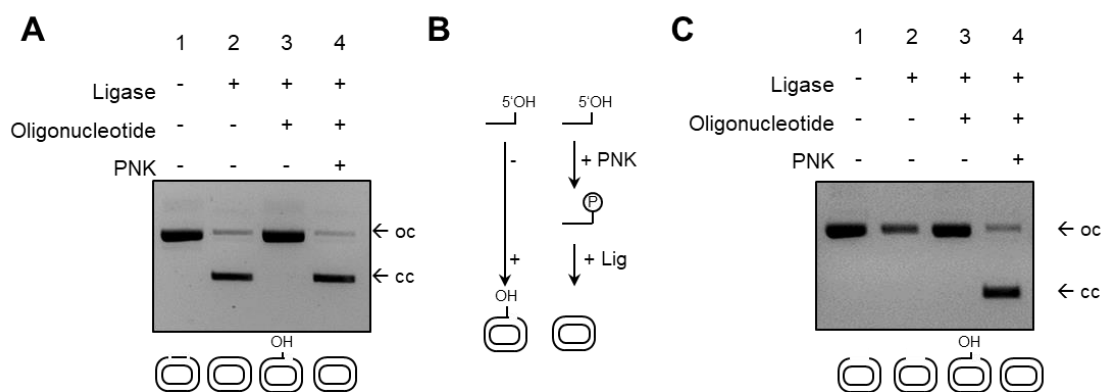


Figure 3-3: Verification of nicking reaction of the pZAJ_5c at both *Nb.Bpu10I* sites and further generation of a gap in the template strand (TS). (A) Analytical ligation after incubation of pZAJ_5c with nicking endonuclease enzyme *Nb.Bpu10I*. (B) Phosphorylation reaction of the synthetic oligonucleotides by PNK and ligation into the nicked plasmid. (C) Analytical ligation of pZAJ_5c bottom strand gapped. 100ng DNA per well.

The following step included the insertion of the 25 different synthetic oligonucleotides chosen previously and listed under [2.1.7] into the gapped pZAJ_5c vector. In this respect, we incubated the gapped vector with each specific oligonucleotide, T4 DNA ligase, and PNK [2.2.11]. While preparing the samples, aliquots containing 100 ng DNA before adding PNK into the reaction mix were taken to follow the efficiency of the reaction

via agarose gel electrophoresis analysis. After the incubation with all components was completed, we collected an additional 100 ng sample from the reaction mix and run them in parallel with the PNK-free samples. The agarose gel analysis revealed an open circular (oc) form of the plasmids in absence of PNK indicating a complete inhibition of the ligation reaction. This result confirmed, on the one hand, the absence of the original 18-nucleotide DNA fragment and on the other hand, the successful annealing of the synthetic oligonucleotides that cannot fully religate into the vector due to the absence of 5' phosphorylation. Thus, the covalently closed (cc) form of the plasmids appeared only in the presence of PNK, which indicated successful incorporation of the 2 control oligonucleotides (613A for the nonsense mutation and 613G as a positive control for fluorescent EGFP protein) and the 25 synthetic oligonucleotides into the gapped pZAJ_5c vector (**Figure 3-4**).

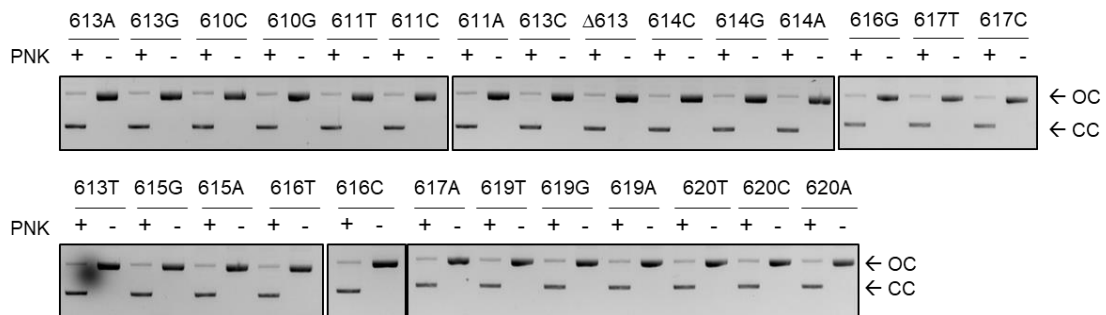


Figure 3-4: Verification of the insertion of 25 synthetic oligonucleotides into the template strand gapped pZAJ_5c vector. Oligonucleotide insertion reactions were performed in the presence of 45-fold molar excess of the corresponding synthetic oligonucleotides and incubated with or without PNK before ligation. Ligation results in a covalently closed (cc) form of the plasmid, which migrates faster than the open circular (oc) form in the presence of ethidium bromide.

3.1.3 Eight of the base substitutions in the TS lead to a loss of the EGFP fluorescence

All synthetic oligonucleotides except one carried a nucleotide substitution at the potential mutated position that created a mismatch once religated into the TS of the gapped pZAJ_5c vector. The $\Delta 613$ construct did not contain any nucleotide substitution; instead, it lacked the guanine nucleotide originally located at position 613. To examine the potential of these 25 nucleotide substitutions to encode for an amino acid that eliminates the EGFP fluorescence, we transfected the constructs into mismatch repair-deficient DLD1 cell-line as a suitable transfection host for the EGFP expression analysis [2.2.13].

DLD1 is a human colorectal adenocarcinoma cell line deficient in MSH6. This gene encodes for the Msh6 protein that together with Msh2 forms a heterodimer responsible for the mismatch recognition step in the mismatch repair (MMR) pathway (Tamura et al., 2019). Thereby, the inability of the cells to recognize the mismatches at the hybrid EGFP

Results

constructs will lead to the direct transcription of the template strand carrying the specified base substitution (**Figure 3-2**). This strategy simplifies the analysis of the resulting fluorescence and prevents the need of an elaborate cloning procedure. 24 hours after transfection of the MMR deficient cells, we quantified the relative EGFP expression by FACS analysis [2.2.16] and classified the fluorescence of the different constructs into 3 groups: <20% (impaired EGFP fluorescence), 20-60% (mildly impaired EGFP fluorescence) and >60% (regular EGFP fluorescence). Thus, we identified eight potentially interesting base substitutions within the <20% fluorescence group: 610A:G; 613C:A; 614A:G; 616T:C; 616T:G; 617C:T; 617C:A; 619G:G (**Figure 3-5**). They displayed more than a 5-fold reduction in the EGFP fluorescence levels and thereby were considered as potential candidates for mutations that could eliminate the EGFP fluorescence. The repertoire of the identified base substitutions with a loss-of-function potential covered the whole nucleotide spectrum with four C:G (NTS:TS) mutants, one with T:A (in addition to c.613C>T mutant that was already available) and one each with G:C and A:T base pairs. One of the candidates (ts.610A>T) has unintentionally dropped out of the screening. However, further phenotypic characterisation of this nucleotide substitution turned out to be no longer needed for the following reasons: first, the equivalent T204S amino acid substitution (conferred by the 611C:C mismatch) showed a residual EGFP signal, and second, the target T:A (NTS:TS) nucleotide pair was already covered by the available c.613C>T mutant. For subsequent validation of the candidate mutants, we prioritized those in which secondary nucleotide substitutions at the affected position were more likely to reconstitute the EGFP fluorescence. This could be estimated from reasonably high signals displayed by the respective neighbors in the preliminary screen, especially at positions 613C:A, 614A:G and 619G:G.

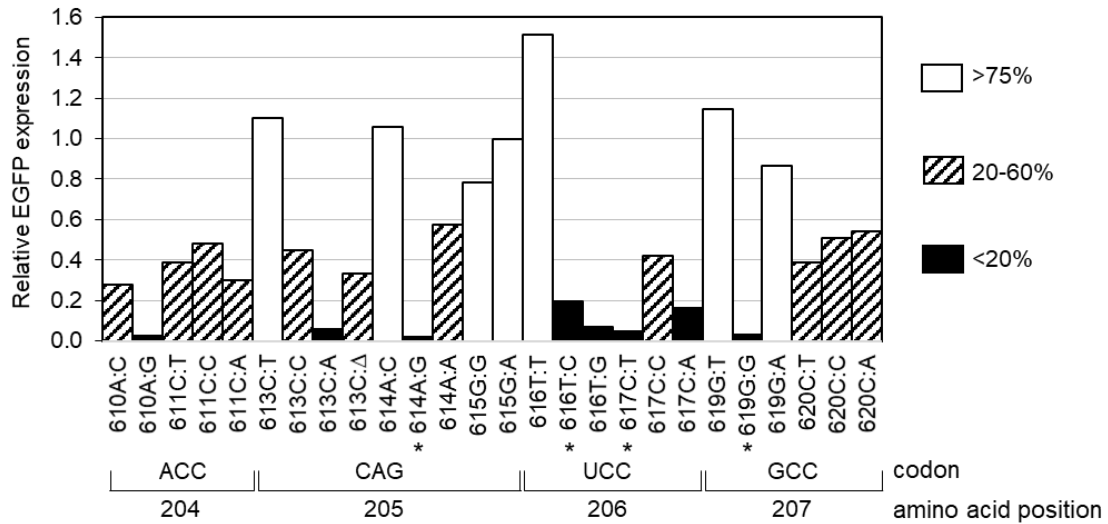


Figure 3-5: Phenotypic screening of mismatches with the indicated nucleotide substitutions in the TS to identify potential point mutations that would lead to loss of EGFP fluorescence. Relative EGFP expression was measured in the DLD1 cells transfected with 25 different constructs containing the specified single-nucleotide mismatches (NTS:TS). The codons indicated underneath the graph display the original codon from the EGFP gene that is affected by the substitutions. (*) Positions selected for cloning and characterisation. Figure adapted from Rodriguez-Alvarez et al., 2020.

3.1.4 Validation of three EGFP mutants carrying inactivating single nucleotide substitutions

The following step was the cloning of the mutant vectors carrying the previously identified mutations: ts.614G (Q205P), ts.616C (S206A), ts.617T (S206Y) and ts.619G (A207P). The cloning procedure started by incubating pZAJ_5c reporter with the restriction enzyme Bpu10I that creates double nicks in both TS and NTS strand comprising positions c.604 and c.622 (**Figure 3-1**). This reaction created a linearized vector and a stretch of 18-mers long double-stranded DNA that was further eliminated via Amicon filtration. On a parallel reaction, we annealed and phosphorylated the synthetic oligonucleotides via incubation with PNK. These two complementary oligonucleotides contained the previously designed point mutation, in both the top and the bottom strand. Incubation with T4 DNA ligase was necessary for permanent ligation of the linearized vector with the recently phosphorylated pair of synthetic oligonucleotides [2.2.1] (**Figure 2-1**). Amplification of the cloned vectors was performed by transforming bacteria [2.2.4] and a miniprep was carried out to obtain enough purified DNA material to continue with further steps [2.2.5]. After cloning four different vectors (Q205P, S206A, S206Y and A207P) carrying the specific point mutation in both strands of the EGFP coding sequence, the presence of the mutated bases was verified by Sanger sequencing [2.2.6]. The chromatograms confirmed the presence of the desired mutations and the success of the cloning procedure (**Figure 3-6**). Once verified, we aimed for big-scale amplification

Results

of the plasmids via megaprep [2.2.5] to produce enough DNA template to perform further experiments.

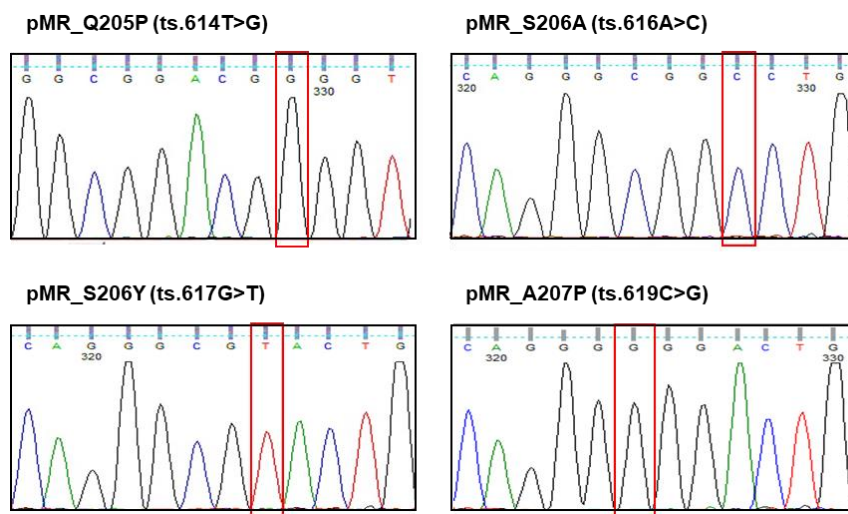


Figure 3-6: Sanger sequencing of cloned vectors. Chromatograms of the sequencing results obtained from newly generated mutant plasmids showing the sequence of the template strand (TS).

To phenotypically validate the loss-of-function mutations, we transfected the set of four mutated vectors into HeLa cells and analyzed the EGFP fluorescence 24 hours after transfection. The resulting EGFP expression was compared with the parental pZAJ_5c as a fully fluorescent control and pZAJ_Q205* as a control for a non-fluorescent vector. Relative EGFP fluorescence indicated that three of the four mutant reporters coded for an amino acid that hindered the EGFP fluorescence: pMR_Q205P (ts.614G), pMR_S206Y (ts.617T) and pMR_A207P (ts.619G). Because pMR_S206A (ts.616C) showed 31% of residual EGFP fluorescence relative to the functional EGFP, it was further eliminated as a potential mutant candidate (**Figure 3-7**). Here, the substitution of a serine amino acid by alanine in position 206 partially (but not completely) affected the fluorescence of the protein. In summary, we successfully cloned and validated three new mutant EGFP reporters whose main feature is the absence of EGFP fluorescence as a result of a single-point mutation.

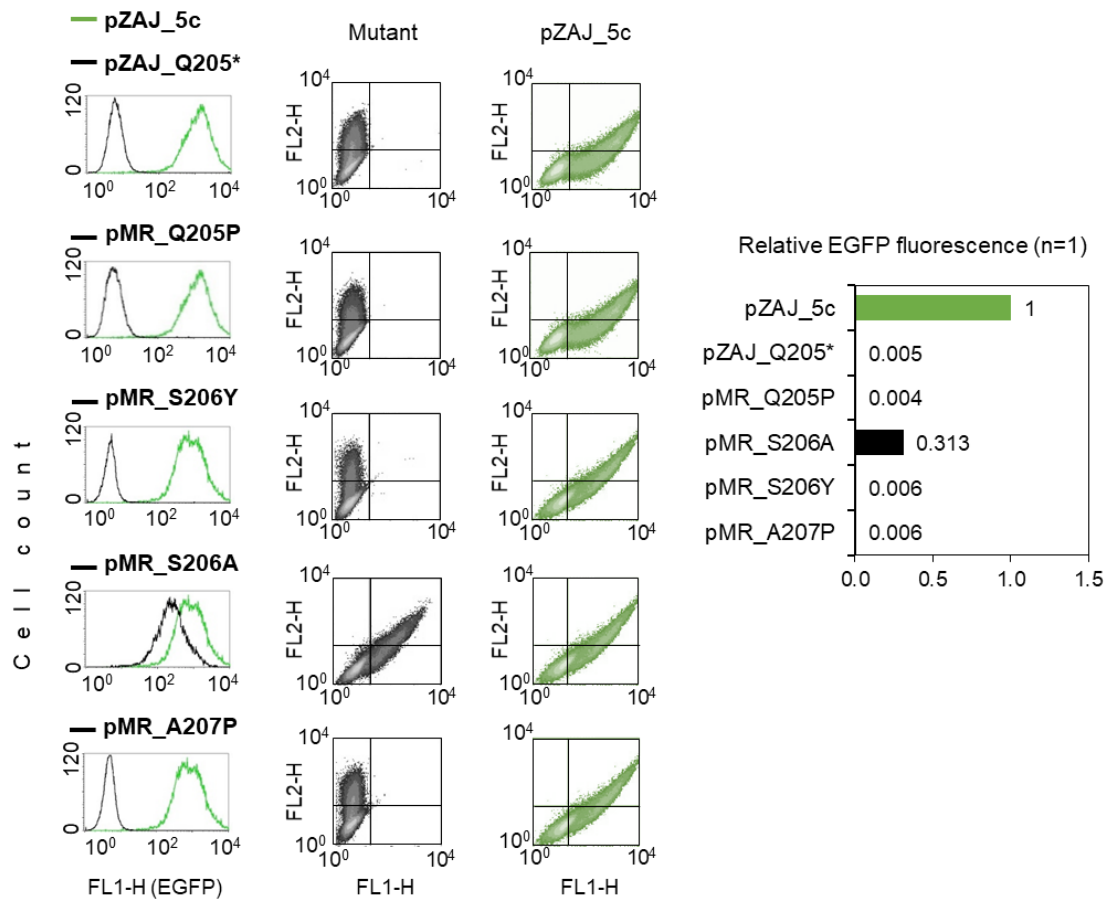


Figure 3-7: Phenotypic characterisation of the specified EGFP mutant vectors. Flow cytometry analyses of different candidates identified in the screening 24h after transfection into HeLa cell-line. EGFP fluorescence distribution plots (left) show expression data of each candidate (black) overlaid with a fluorescent EGFP plasmid as a positive control (green), scatter plots (middle) and quantification are shown on the right graph. Figure adapted from Rodriguez-Alvarez et al., 2020.

3.1.5 Secondary mutation events at the affected position reactivate EGFP fluorescence

For efficient detection of mutagenesis occurring at the target position, random nucleotide substitutions at the mutated base must reactivate the EGFP fluorescence. In order to test all sequence variants, three mismatched constructs were generated from every single mutant (twelve in total). We included also the previously created Q205* (ts.613G>A) mutant which characterisation was missing at this level. Thereby, we substituted the 18-mers fragment flanked by the Nb.Bpu10I nicking sites of the template strand in each reporter with three synthetic oligonucleotides. For the mutant EGFP vector pZAJ_Q205* (ts.613A) we generated: ts.613G, ts.613C and ts.613T; for pMR_Q205P (ts.614G) we generated: ts.614A, ts.614C and ts.614T; for pMR_S206Y (ts.617T) we generated: ts.617A, ts.617G and ts.617C; and for pMR_A207P (ts.619G) we generated: ts.619A, ts.619C and ts.619T. The idea behind this experiment was that if the mutated nucleotide was substituted by a DNA lesion such as an AP site, mutagenic bypass events

Results

occurring at that position would lead to the insertion of a nucleotide different from the original. In order to directly detect these mutagenic events, the misincorporated nucleotide should code for an amino acid that reactivates the protein fluorescence. Thus, each of the constructs carried one of the three alternative nucleotides with the potential to be inserted at the target position. Electrophoresis agarose gel in the presence and absence of PNK showed the successful incorporation of the synthetic nucleotides into the plasmids (**Figure 3-8**).

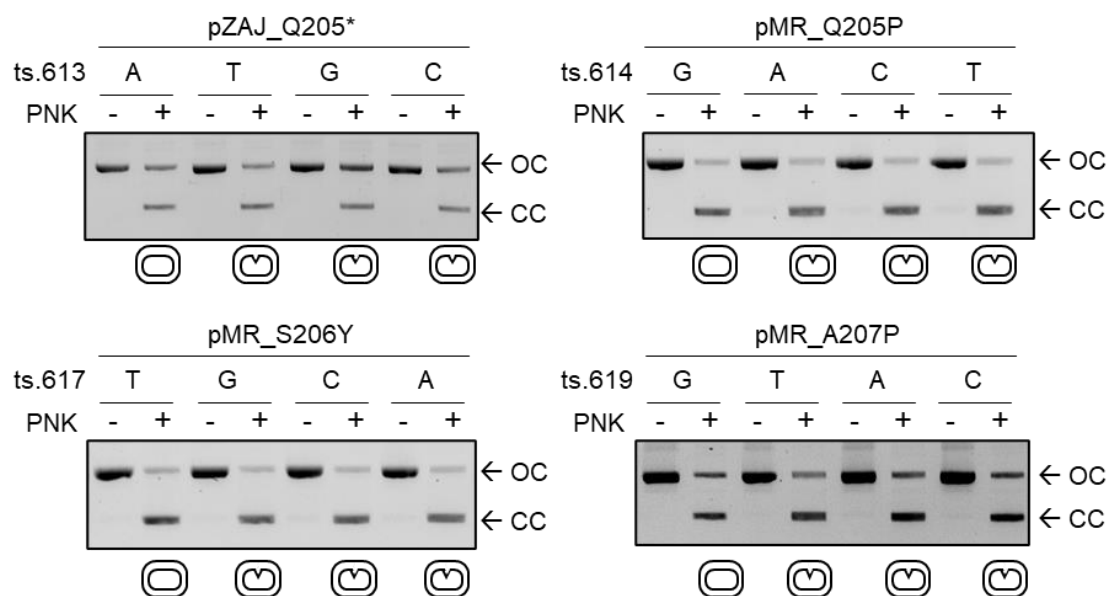


Figure 3-8: Verification of insertion of the synthetic oligonucleotides carrying all possible single mismatches into four different mutant reporters. Ligation results in a covalently closed (cc) form of the plasmid, which migrates faster than the open circular (oc) form in the presence of ethidium bromide. The first construct of each plasmid represents the original base pair, the three additional constructs contain a mismatched pair due to the insertion of a synthetic oligonucleotide carrying an alternative nucleotide in the specified position of the TS.

To analyze whether any alternative nucleotide would revert into a fluorescent EGFP protein, we transfected each construct into the MMR deficient cell line DLD1. As mentioned previously, DLD1 cells cannot repair the mismatch inserted into the plasmid and transcription occurs directly on the template strand carrying each nucleotide substitution. Analysis of the EGFP fluorescence 24 hours after transfection indicated that only adenine eliminates the protein fluorescence in pZAJ_Q205*, whilst any alternative nucleotide (guanine, cytosine, or thymine) would efficiently reactivate the EGFP fluorescence (**Figure 3-9**). In the same manner, only guanine maintained the non-fluorescent status in the pMR_Q205P and pMR_A207P, whereas any alternative substitution at these positions rescued the EGFP fluorescence. However, we observed a different pattern in the pMR_S206Y mutant reporter, cytosine (C) and adenine (A) substitutions showed only partial recovery of the fluorescent signal (**Figure 3-9**, A:A and A:C versus A:T). Still, they showed increased fluorescence compared to the originally

mutated one by the factors of >40 and >6, for ts.617T>C and ts.617T>A, respectively. As expected, guanine substitution showed maximal fluorescence at this position since it reverts the mutation to the wild-type EGFP sequence. This result indicates, that pMR_S206Y has a lower sensitivity for detection of nucleotide misincorporations at position 617 than the other mutants.

Considering these minor limitations, we have successfully created a set of four EGFP mutants that allow the direct detection of nucleotide misincorporations at the target positions. The mutated nucleotides that are potential targets for the incorporation of DNA lesions are A, C and T within the transcribed strand and A, G and T within the non-template strand. Nevertheless, we still need a reporter carrying a G:C (NTS:TS) mutation in order to obtain a complete collection of mutants carrying each possible nucleotide substitution.

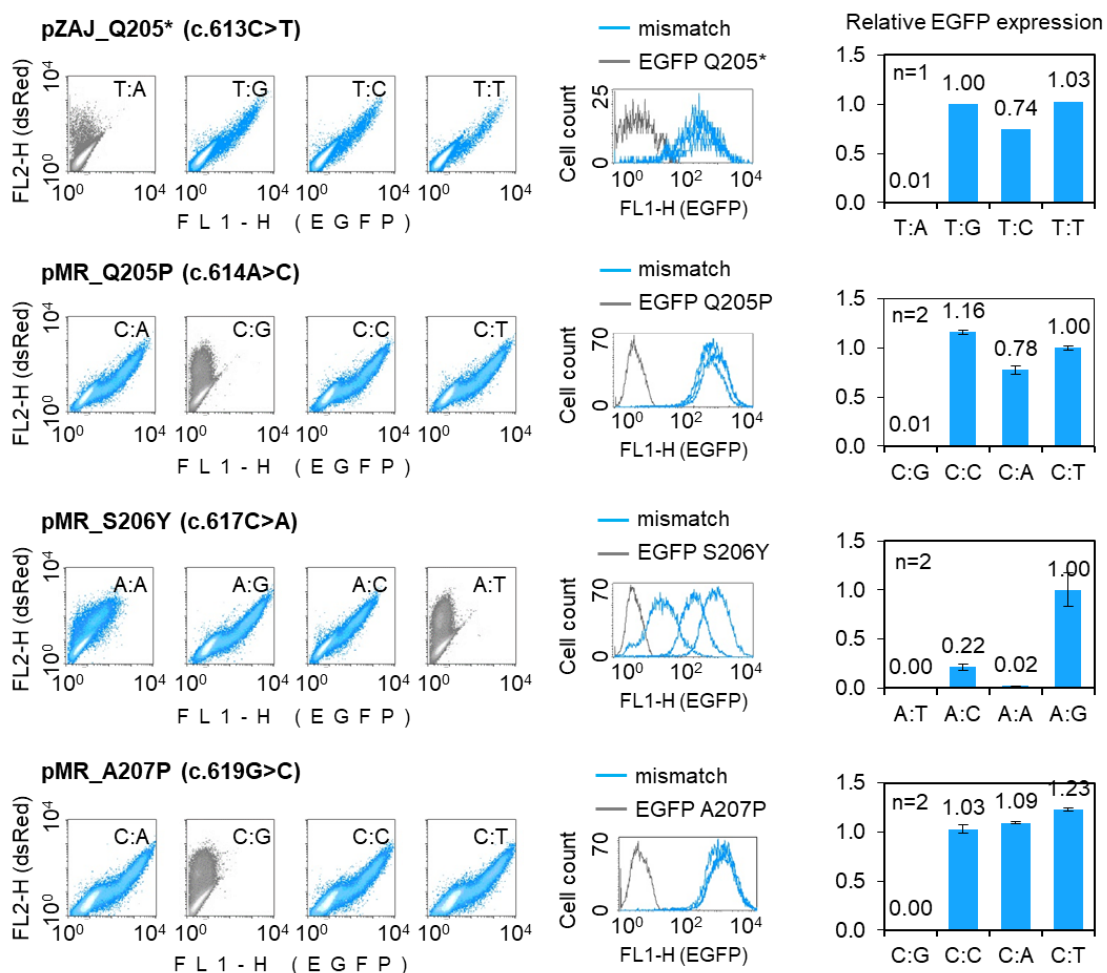


Figure 3-9: Characterisation of all possible nucleotide substitutions at the mutated position of pZAJ_Q205*, pMR_Q205P, pMR_S206Y and pMR_A207P. Relative EGFP expression of mismatched constructs with nucleotide substitutions in the TS is indicated in blue color and compared to the original base pair (grey), 24h after transfection of mismatch repair deficient cell line (DLD1). Upper letter on the scatter plots indicates the nucleotide in NTS opposite to: the original or mispaired nucleotide place in TS (NTS:TS). Figure adapted from Rodriguez-Alvarez et al., 2020.

Results

3.1.6 Mutant pMR_S206* as an alternative reporter to study guanine modifications in the non-template strand

The only candidate with a mutation involving guanine in the non-template strand (NTS) was pMR_S206A; however, this mutant was discarded as it displayed a residual EGFP fluorescence (**Figure 3-7**). Therefore, we attempted to design an alternative mutant reviewing again the coding sequence of the EGFP gene between the Bpu10I nicking sites. Unfortunately, we did not find any single guanine mutation that would theoretically affect the amino acid sequence. With this limitation, we decided to create a stop codon within the target sequence selecting not one, but two consecutive positions: 617 and 618. Thus, we generated c.617C>G and c.618C>A substitutions creating a nucleotide triplet TGA (UGA codon in transcribed mRNA) where the guanine would be the candidate position for further analysis of DNA lesions. The obvious limitation of this mutant is that two of the three possible alternative substitutions encoded for a stop codon (**Figure 3-10A**). Consequently, this plasmid cannot be used to detect adenine misincorporation during transcriptional bypass or thymine misincorporation during replicative bypass occurring at the target position.

We generated two hybrid EGFP constructs each containing the innate non-template strand (NTS) and two specific nucleotide substitutions in the TS creating two mismatches: the designed nonsense mutation (c.617C>G-c.618C>A) and the remaining uncharacterised missense mutation (c.617C>T-c.618C>A) to assess their fluorescent phenotype. Following the screening procedure described in section 3.1.3 (**Figure 3-5**), we transfected the constructs into DLD1 cell line and analyzed the EGFP fluorescence 24h after transfection. Both nucleotide substitutions tremendously affected the protein fluorescence resulting in less than 11% relative fluorescence (**Figure 3-10B**). This result indicated that EGFP fluorescence can only be restored by a G>C transversion mutation (TCA) at position c.617. Consequently, this mutant could still be used to detect mutagenic events that would insert either guanine during transcription or cytosine during DNA synthesis at position 617. Therefore, we decided to clone the newly designed vector and verify the nucleotides substitution via Sanger sequencing, which confirmed the ts.617G>C; ts.618G>T substitutions (**Figure 3-10C**). Following the same procedure as for the other four mutants, we also transfected HeLa cells to verify that the absence of any residual fluorescent phenotype. As indicated by the FACS analysis, we confirmed that the new reporter lacks EGFP fluorescence (**Figure 3-10D**) and thus would be suitable for the detection of G:C to C:G transversions.

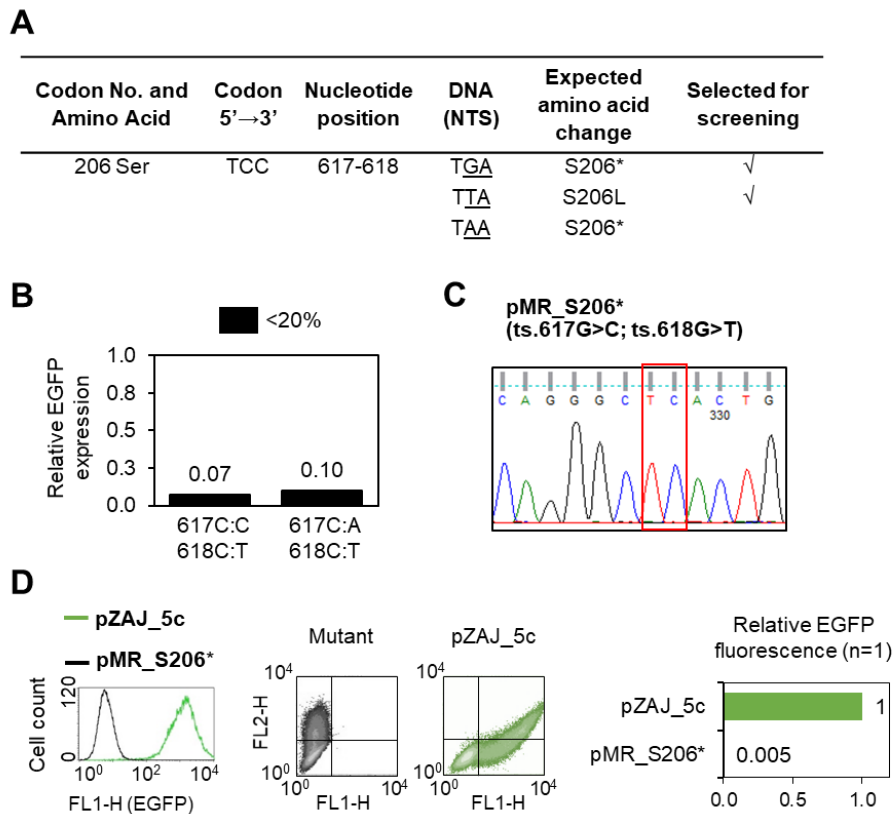


Figure 3-10: Generation and characterisation of pMR_S206* mutant reporter. (A) Analysis of the codon 206 on the EGFP coding sequence to identify mutations at a G:C pair that will lead to a change in the amino acid sequence of the EGFP protein. (*) Termination codon. (B) Relative EGFP expression was measured in the DLD1 cells transfected with 2 different constructs containing the specified nucleotide mismatches (NTS:TS). (C) Chromatogram of the sequencing result of the newly generated pMR_S206* mutant reporter, the sequence in the template strand is indicated above. (D) Flow cytometry analyses of the mutant candidate 24h after transfection in HeLa cell-line. EGFP fluorescence distribution plots show expression data of the mutant (black) overlaid with a fluorescent EGFP plasmid (pZAJ_5c) as a positive control (green).

Altogether, the results compiled in this chapter described the design and characterisation of five different EGFP mutants; pZAJ_Q205* (c.613T), pMR_Q205P (c.614C), pMR_S206Y (c.617A), pMR_A207P (c.619C), and pMR_S206* (c.617G). They are potentially suitable as reporters for sensitive detection of nucleotide misincorporations at the affected positions (**Table 3-2**). In the first four reporters, any alternative mutation (with exception of A and C in position ts.617) at the specified base pairs can be detected with high sensitivity by reactivation of EGFP fluorescence. The pMR_S206* reporter would allow detection of mutagenic bypass over DNA lesions that result in G misincorporation into the template strand. In the following sections, we will demonstrate the usefulness of these EGFP reporters for direct and sensitive detection of mutagenic bypass of DNA lesions in human cells.

Results

Table 3-2: Summary of mutant plasmids lacking EGFP fluorescence and the result of possible nucleotide substitutions at the target position during DNA or RNA synthesis based on data of figure 3-9.

Vector	DNA	Base pair (NTS:TS)	DNA (TS) 3'→5'	RNA 5'→3'	Amino acid	Brightness (% EGFP)
pZAJ_Q205*	c.613C>T	T:A	<u>A</u> TC	<u>U</u> AG	Stop	0
			<u>G</u> TC	<u>C</u> AG	Gln	~100
			<u>C</u> TC	<u>G</u> AG	Glu	~74
			<u>I</u> TC	<u>A</u> AG	Lys	~100
pMR_Q205P	c.614A>C	C:G	<u>G</u> GC	<u>C</u> CG	Pro	0
			<u>G</u> TC	<u>C</u> AG	Gln	~100
			<u>G</u> AC	<u>C</u> UG	Leu	~80
			<u>G</u> CC	<u>C</u> GG	Arg	>100
pMR_S206Y	c.617C>A	A:T	<u>A</u> TG	<u>U</u> AC	Tyr	0
			<u>A</u> GG	<u>U</u> CC	Ser	~100
			<u>A</u> AG	<u>U</u> UC	Phe	~3
			<u>A</u> CG	<u>U</u> GC	Cys	~20
pMR_S206*	c.617C>G c.618C>A	GA:CT	<u>A</u> CT	<u>U</u> GA	Stop	0
			<u>A</u> GT	<u>U</u> CA	Ser	~100
			<u>A</u> AT	<u>U</u> UA	Leu	~10
			<u>A</u> TT	<u>U</u> AA	Stop	~0
pMR_A207P	c.619G>C	C:G	<u>G</u> GG	<u>C</u> CC	Pro	0
			<u>C</u> GG	<u>G</u> CC	Ala	~100
			<u>A</u> GG	<u>U</u> CC	Ser	>100
			<u>I</u> GG	<u>A</u> CC	Thr	>100

3.2 Mutagenic bypass of abasic DNA lesions during transcription

3.2.1 Generation of constructs accommodating BER-resistant AP lesion for detection of transcriptional mutagenesis in pZAJ_Q205*

For direct detection of transcriptional mutagenesis over AP lesions in human cells, two EGFP mutants Q205* and A207P that allow the substitution of adenine or guanine respectively by an abasic DNA lesion were chosen. (**Figure 3-11**).

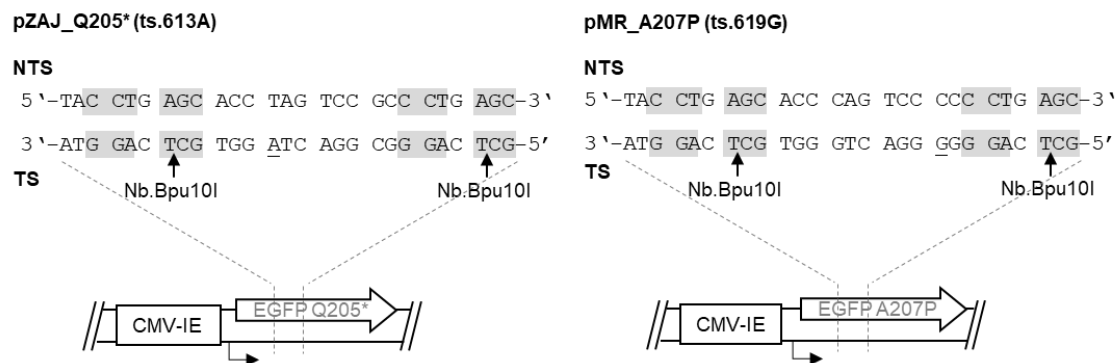


Figure 3-11: Schematic representation of pZAJ_Q205* and pMR_A207P reporters encoding for the non-fluorescent EGFP mutants. Recognition sites of the Nb.Bpu10I nicking endonuclease (grey shadow in TS), the arrows indicate the specific cutting position. Broken arrows show the transcription start position. The rectangle indicates the CMV-IE promoter under whose control the reporter EGFP gene is located. The position where the synthetic lesion will be placed is underlined.

To prepare the constructs for transcriptional mutagenesis (TM) assay, we selectively depleted the 18-nucleotide fragment from the vector enclosed by Nb. Bpu10I nicking sites and instead, integrated synthetic oligonucleotides containing a single AP lesion at the mutated position in the transcribed strand (TS) (**Figure 3-12**).

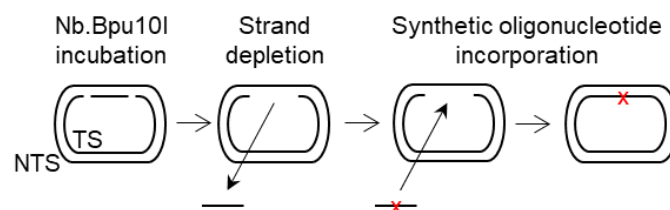


Figure 3-12: Scheme of preparation of constructs for TM assay. Double nicking of the template strand via Nb.Bpu10I incubation, followed by strand depletion and incorporation of synthetic oligonucleotides in the template strand (TS). The X represents the DNA lesion embedded into the synthetic oligonucleotide.

Because the repair of AP lesions by the BER pathway is extremely efficient, it was necessary to exclude the AP site lyase and endonucleases activities associated with the first step of this pathway. Since APE1 is an essential gene in human cells (Fung & Demple, 2005; Xanthoudakis et al., 1996) and five mammalian glycosylases exist with AP-lyase activities, it was more straightforward to use a chemical rather than a genetic approach to render AP lesions resistant to BER. Thus, we used the synthetic analog of

Results

the ring-closed form of natural AP sites known as tetrahydrofuran (THF) type AP site (**Figure 3-13A**) which is inherently resistant to β -elimination due to the lack of the hydroxyl group positioned in C1 (Takeshita et al., 1987; Wilson III David M et al., 1995); however, this THF AP site is still a good substrate of AP endonuclease 1 (APE1). Therefore, we substituted the phosphodiester bond by a phosphorothioate linkage 5' to the THF AP site (S-THF) which has been proven to confer resistance to APE1 incision (Allgayer et al., 2016; Wilson III David M et al., 1995), (**Figure 3-13B**).

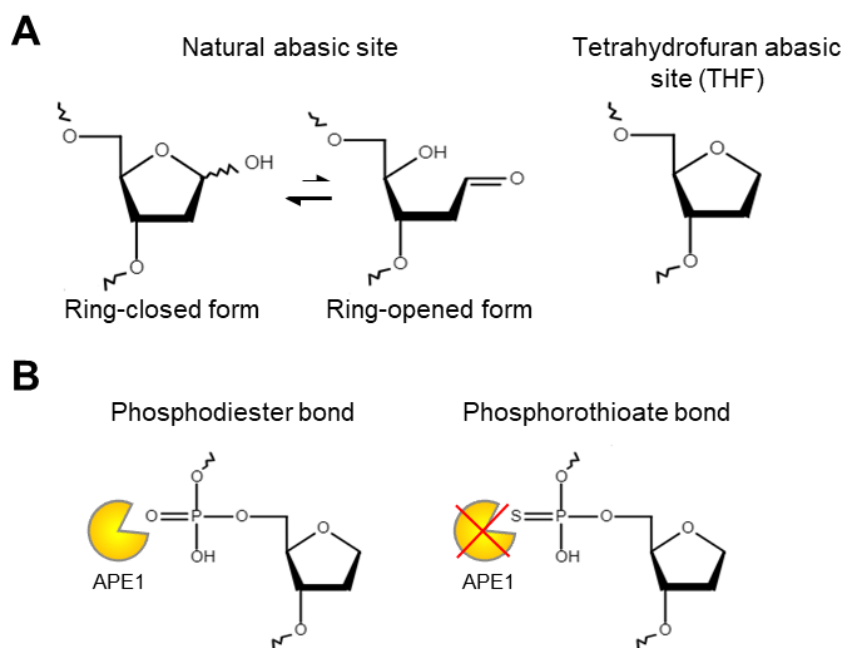


Figure 3-13: Abasic DNA lesions and linkages used to analyze TM. (A) Chemical structure of natural AP sites showing the equilibrium between the ring-closed, furanose, and ring-open, aldehyde form (left). Chemical structure of tetrahydrofuran (THF) AP site (right). (B) Chemical structure of an APE1-sensitive phosphodiester linkage (left) and an APE1-resistant phosphorothioate linkage (right).

The constructs for TM analysis were prepared starting from pZAJ_Q205* mutant, which contains a point mutation at position ts.613 (G>A) encoding for an early termination codon that generates a truncated EGFP Q205* protein. Incorporation of a synthetic oligonucleotide in the transcribed resulted in the substitution of the adenine in position 613 for a synthetic abasic site (ts.613A>THF). If transcriptional mutagenesis through the abasic lesion occurs, the resulting mRNA would encode for a fluorescent EGFP protein. Thus, the misincorporation of cytosine, guanine, or adenine opposite to the lesion will be immediately detected by fluorescence analysis (**Figure 3-14**).

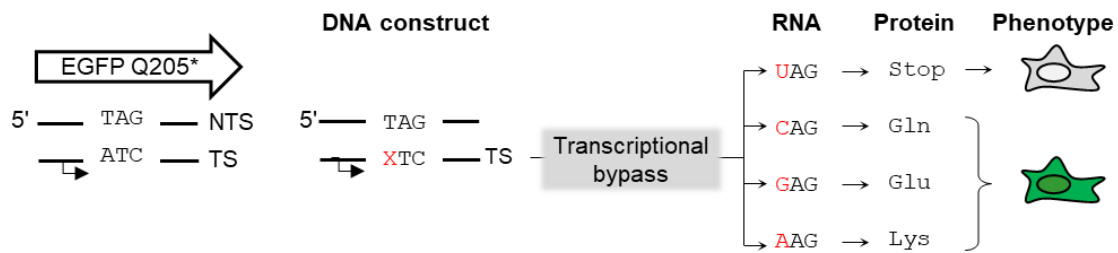


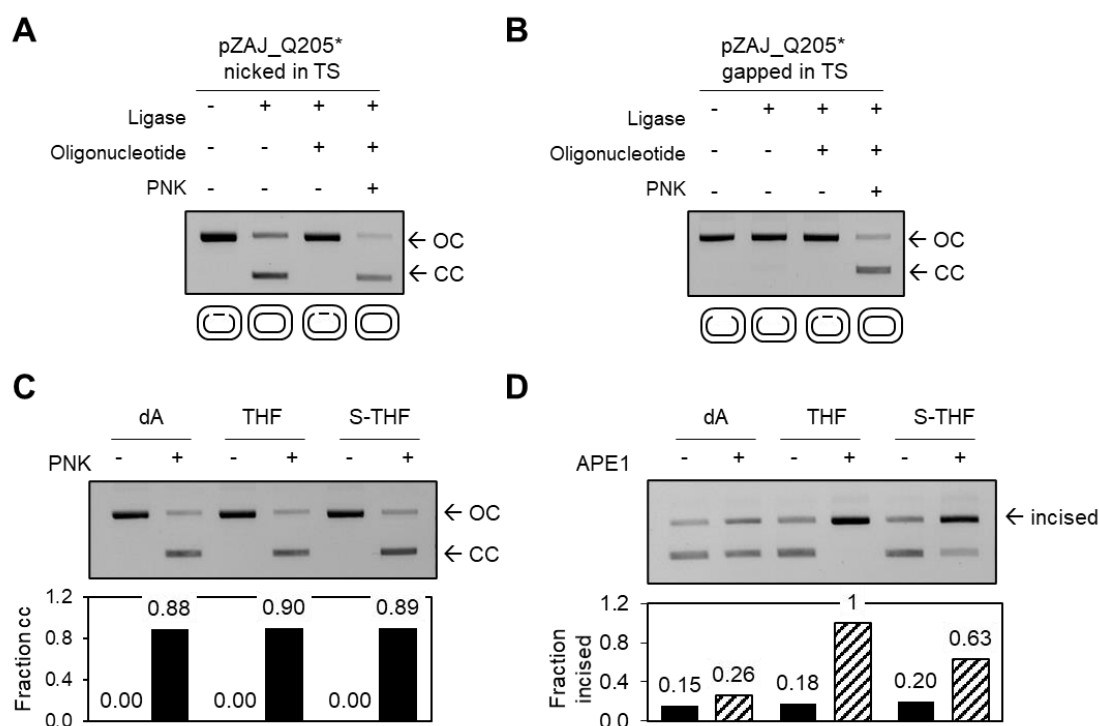
Figure 3-14: Scheme of the pZAJ_Q205* vector used for transcriptional mutagenesis (TM) assays. The red X indicates the position of the abasic lesion. After transcriptional bypass of the lesion, all possible outcomes are indicated at RNA, protein and phenotypic levels.

To generate constructs harboring site-specific THF and S-THF AP sites, it was necessary to incubate the pZAJ_Q205* vector with the strand-specific Nb.Bpu10I nicking endonuclease that generated two tandem nicks in the template strand (TS) as described in 2.2.7. Then, we verified the presence of double nicked plasmid via analytical ligation [2.2.8] and agarose gel electrophoresis (**Figure 3-15A**). After incubation with T4 DNA ligase, the double-nicked plasmid religated with the DNA strand and ran as a covalently closed (cc) form in the agarose gel. In a parallel reaction, the nicked plasmid was incubated with ligase and 180-fold excess of 5'-unphosphorylated synthetic oligonucleotides. Even though these oligonucleotides were annealed with their complementary sequence within the nicked plasmid, the absence of a phosphate group at the 5'-end prevented their ligation. Thus, the sample ran as an open circular (oc) form into the agarose gel. The addition of polynucleotide kinase (PNK) to the last sample, led to the phosphorylation of the 5'-OH group of the synthetic oligonucleotides. This allowed further ligation into the nicked vector restoring the closed circular (cc) form of the plasmid. As no cc form of the plasmid was detected in absence of PNK, it was deduced that the plasmid has been nicked at both Nb.Bpu10I sites. To eliminate the original 18-mers DNA fragment hanging between the nicking sites, we performed a strand depletion reaction by incubation with 920-fold excess of competitor oligonucleotide [2.2.10]. The remaining hybrid duplexes and the complementary oligonucleotides were removed by Amicon filtration [2.2.2], thus obtaining an 18-nucleotide gap in the template strand of the vector. The presence of the gap was verified by performing an additional analytical ligation under the same conditions as previously although this time using a much lower excess (45-fold) of synthetic oligonucleotides. Under these conditions, the absence of a covalently closed (cc) form of the plasmid in the presence of ligase indicated that the original 18-mers DNA fragments have been removed and that the plasmid is properly gapped (**Figure 3-15B**). Once the gapped pZAJ_Q205* vector was obtained, we proceeded with the insertion of synthetic oligonucleotides containing the DNA modifications into the gap as described under 2.2.11. By this means, three different constructs were generated,

Results

two containing synthetic AP lesions (THF and S-THF) and one containing dA at position ts.613 as a control for a non-fluorescent protein. Incorporation of the synthetic oligonucleotides into the gapped plasmid was shown as cc form of the plasmid with >88% ligation efficiency only after incubation with PNK (**Figure 3-15C**).

To confirm the presence of THF as well as the resistance of the S-THF to BER, we incubated the constructs with an excess of purified human AP endonuclease 1 (APE1) as described in 2.2.17. Due to APE1 incision, 100% of the constructs containing THF acquired an open circular form that confirmed the presence of THF AP lesions in all plasmids (**Figure 3-15D**). On the contrary, the S-THF constructs showed only an increase of 37% of cleavage compared to the control dA. The 11% of unspecific cleavage detected in constructs harboring unmodified oligonucleotide (dA) was attributable to the presence of abasic sites preexisting in the plasmid DNA isolated from bacteria. These results clearly indicated that THF and S-THF modifications were efficiently incorporated into the gapped vector and that the 5'-phosphorothioate linkage turned the abasic site resistant to APE1 activity.



To further characterise the APE1 activity towards THF and S-THF substrates, we performed digestion assays titrating APE1 concentration in reactions containing 100 ng of each construct (**Figure 3-16**). Here, conversion of the covalently closed form into the open circular form of the plasmid is proportional to the APE1 activity. Unlike the complete cleavage obtained on the THF constructs incubated with 1U of APE1, S-THF only showed an increase of 33% compared to the dA construct under the same conditions. Below saturation levels, 0.2 units of APE1 were sufficient to cleave more than 92% of the THF substrate whereas S-THF constructs remained completely resistant to APE1 when compared to the basal cleavage showed by the control sample dA. This digestion assay indicated that 5-fold more concentration of APE1 is necessary to cleave S-THF substrates at the same level as THF.

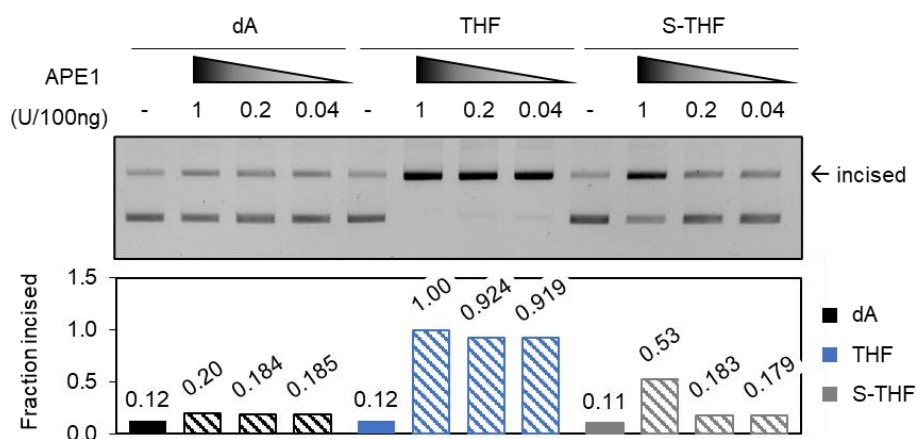


Figure 3-16: APE1 digestion assays of *ts.613A*, THF and S-THF constructs derived from *pZAJ_Q205 reporter.** Specified units of the enzyme for 100ng of DNA are indicated above the gel image; the table indicates quantification in the percentage of the incised fraction.

3.2.2 Transcriptional bypass of repair resistant AP sites results in mutant mRNA

To detect transcriptional mutagenesis occurring at AP sites in human cells, it was important to evaluate the EGFP expression in the repair-proficient HeLa cell-line transfected with the previously characterised constructs harboring THF or BER-resistant S-THF abasic sites. Cells were co-transfected with equal copy numbers of the reporter constructs in combination with the pDsRed vector as a transfection marker and analyzed 24 hours later by microscopy and flow cytometry as described in 2.2.16. Cells transfected with THF and the unmodified control dA showed a complete lack of EGFP signal in the microscopy images as well as in the quantitative FACS analysis. In contrast, cells transfected with S-THF constructs displayed a large proportion of reversal to a fluorescent EGFP phenotype (35%) (**Figure 3-17**). These results suggested that the efficient repair of THF restored the A:T base pair encoding for a truncated protein prior to its transcription. In this context, BER resistance of the S-THF substrate extended the

Results

half-life of the lesion within the cell allowing the direct transcription through the abasic site. Regain of EGFP fluorescence indicated that transcriptional bypass occurs at BER-resistant AP sites with a base substitution rather than a base deletion or insertion. As the incorporation of cytosine, guanine or adenine into the nascent mRNA opposite to the lesion would encode for an amino acid that reactivates protein fluorescence (**Figure 3-14**), 35% of EGFP fluorescence indicated that at least one-third of the transcripts were synthesized with a ribonucleotide misincorporation. However, the nature of the nucleotide incorporated is uncertain. These results demonstrated that transcriptional bypass of S-THF results in the synthesis of a high rate of mutant transcripts and that these TM events can be easily detected using constructs originated from the pZAJ_Q205* mutant.

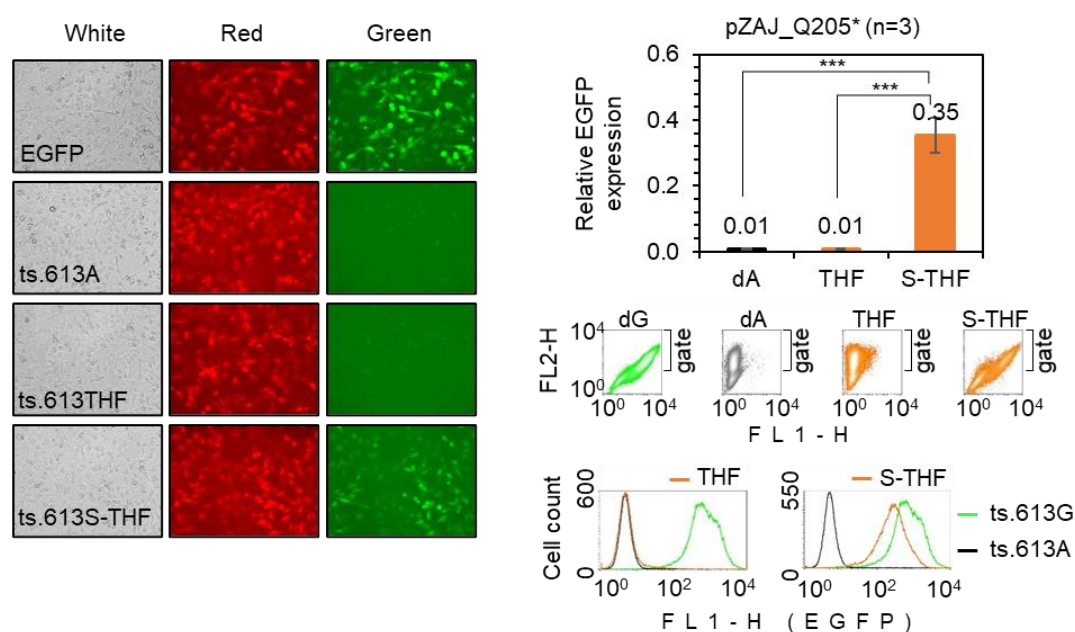


Figure 3-17: Transcriptional mutagenesis at BER-Resistant AP sites (S-THF) reactivates EGFP fluorescence in position ts.613 in HeLa cells. Left: Microscope images of HeLa 24 hours after transfection with pZAJ_5c and pZAJ_Q205* (dA, THF and S-THF). Right: Quantification of relative EGFP fluorescence of three different independent experiments in HeLa cells; ***: $p\text{-value}=3.8 \times 10^{-4}$ (top). Scatter plots and overlaid fluorescence distribution plots from a representative experiment (bottom). Figure adapted from Rodriguez-Alvarez et al., 2020.

After demonstrating the capacity of the pZAJ_Q205* mutant reporter to detect AP site-derived transcriptional mutagenesis in human cells, we aimed to try an alternative mutant for the same purpose but placing the AP site at a different nucleotide. The pMR_A207P reporter contains a missense mutation in position ts.619C>G that eliminates the protein fluorescence. Since depurination of guanine is one of the main sources of AP sites formation under physiological conditions (Loeb & Preston, 1986), this would allow the study of the transcriptional bypass of the THF lesion located at a guanine position. In this mutant, detection of TM via EGFP fluorescence analysis is only possible if it occurs

through guanine, uracil, and/or adenine incorporation opposite to the abasic DNA lesion (**Figure 3-18**).

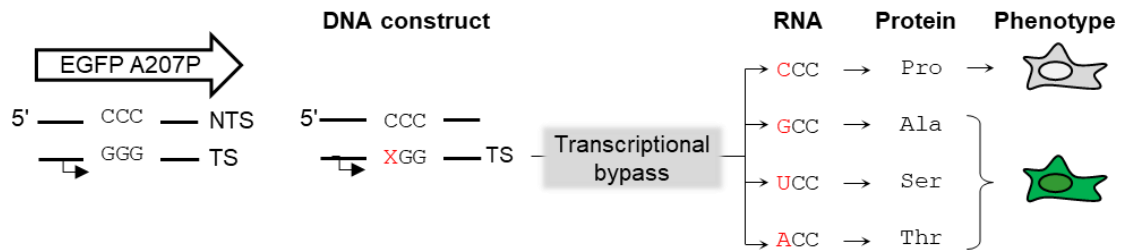


Figure 3-18: Scheme of the mutant pMR_A207P used for transcriptional mutagenesis (TM) assays. The red X indicates the position of the abasic lesion. After transcriptional bypass of the lesion, all possible outcomes are indicated at RNA, protein, and phenotypic levels.

To perform the TM assay using this mutant, it was necessary to create a gap into the TS of pMR_A207 as previously described for the pZAJ_Q205* mutant. The analytical ligation experiment verified the presence of the gap by showing a complete absence of cc form of the plasmid in the presence of ligase (**Figure 3-19A**); thus indicating the successful elimination of the original 18-mers DNA fragment placed between the Nb.Bpu10I nicking sites. The further incorporation of the synthetic oligonucleotides carrying THF and S-THF AP lesions at the position ts.619 was performed with a ligation efficiency of >90% illustrated as cc form of the plasmid only after incubation with PNK (**Figure 3-19B**). Then, the APE1 digestion assay confirmed the presence of THF and S-THF AP lesions and the resistance of S-THF substrate against APE1 incision. While complete cleavage of THF constructs was detected as oc form of the plasmid, S-THF constructs remained mostly resistant to APE1 showing an increase of only 28% of the incised fraction compared to dG (**Figure 3-19C**).

Results

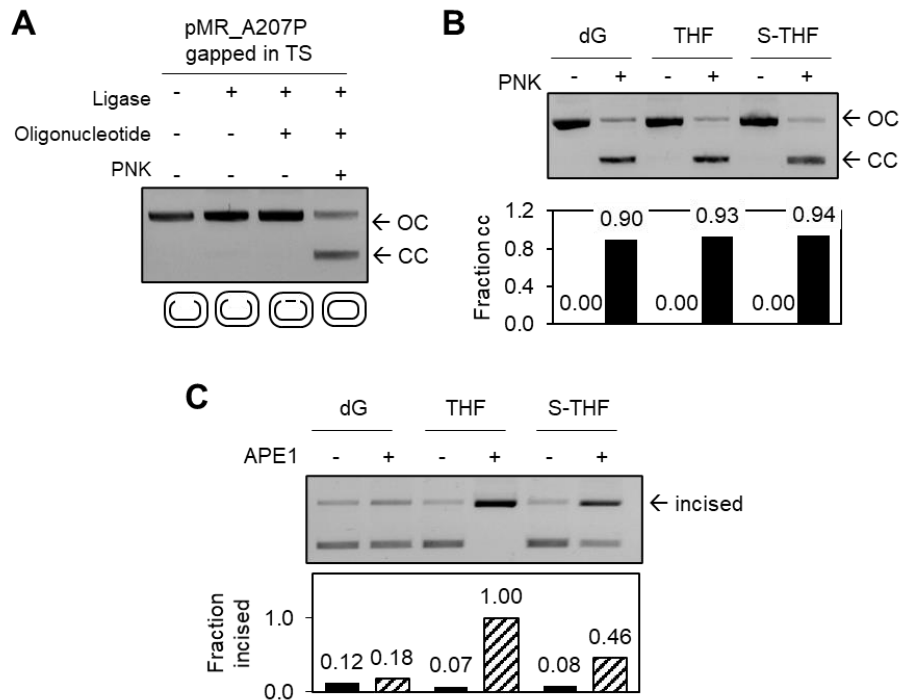


Figure 3-19: Preparation of constructs with THF and S-THS in position ts.619 of the pMR_A207P reporter. (A) Analytical ligation of the gapped pMR_A207P showing the open circular (oc) and covalently closed (cc) form of the plasmid. (B) The efficient incorporation of three synthetic oligonucleotides (dG, THF and S-THF) in position ts.619 is shown by the covalently closed (cc) form of the plasmid in presence of PNK. (C) APE1 digestion assay of the indicated constructs using 1U APE1/100 ng of DNA and quantification of the incised fraction of the plasmid.

Last, we transfected HeLa cells with the newly generated constructs and analyzed the resulting EGFP signal by microscopy and flow cytometry. The microscopy images already showed a re-gain of EGFP fluorescence only in cells transfected with the wild-type EGFP construct (c.6129C) and the S-THF construct. Similar to the Q205* mutant analyzed previously, the BER-resistant S-THF lesion caused a very strong gain of EGFP fluorescence (33%). In contrast, the control dG and BER-sensitive THF constructs showed no detectable fluorescence (**Figure 3-20**). This result indicated that reactivation of the fluorescent phenotype in cells transfected with S-THF constructs emerged as a result of the TM occurring at the lesion site with the incorporation of guanine, uracil and/or adenine in the nascent mRNA.

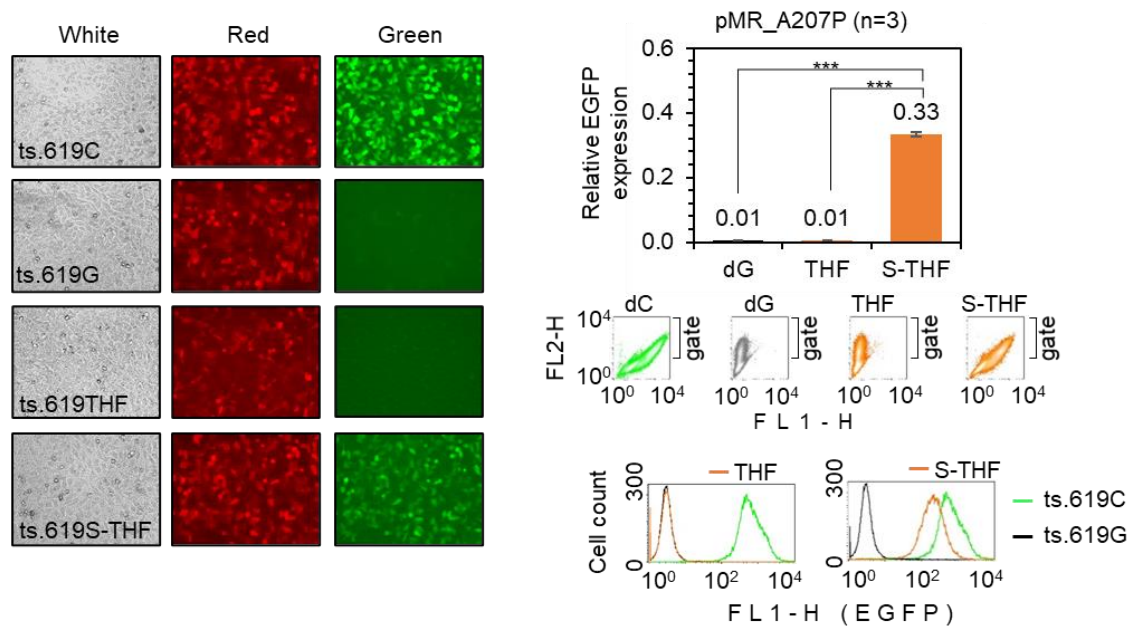


Figure 3-20: Transcriptional mutagenesis at BER-Resistant AP sites (S-THF) reactivates EGFP fluorescence in position ts.619 in HeLa cells. Left: Microscope images of HeLa 24h after transfection with pZAJ_5c and pMR_A207P (dG, THF and S-THF). Right: Quantification of relative EGFP fluorescence of three different independent experiments in HeLa cells; ***: $p\text{-value} = 4.8 \times 10^{-7}$ (top). Scatter plots and overlaid fluorescence distribution plots from a representative experiment (bottom). Figure adapted from Rodriguez-Alvarez et al., 2020.

In both reporter systems (pZAJ_Q205* and pMR_A207P) transfected in HeLa cells, we observed 33-35% of the EGFP fluorescence intensity produced by a fully functional EGFP protein. These results indicate that at least one-third of all transcripts contained ribonucleotide substitutions at the lesion site and illustrate the high miscoding potential of AP lesions during transcriptional bypass events. As a demonstration of the remarkable efficiency of repair of AP lesions in human cells, the EGFP signal was completely absent when cells were transfected with BER-sensitive THF lesions as documented by direct comparison with control constructs at the analyzed positions, dA for Q205* and dG for A207P. In summary, both reporter systems proved to be extremely sensitive for TM detection over AP lesions in human cells.

3.2.3 Transcriptional mutagenesis at AP sites is enhanced in the absence of nucleotide excision repair (NER)

The reporter reactivation assay established in section 3.2.2, offers a broad dynamic range for the detection of transcriptional mutagenesis (TM). This, together with the inverse correlation existing between repair of the lesion and the level of TM detected, provides a potential tool to assess the repair capacity of the cells towards abasic sites. It has been previously reported that AP sites are substrates of nucleotide excision repair (NER) proteins in yeast (Guillet Marie & Boiteux Serge, 2002; Kim & Jinks-Robertson,

Results

2010); however, the implication of NER in the repair of abasic DNA lesions in mammalian cells remains uncertain.

To study the role of NER proteins in the repair of AP sites, we transfected the THF and S-THF lesions derived from the mutant pZAJ_Q205* into a broad panel of NER-deficient cell lines. Human GM04312 cells derived from an XP-A patient, lack a functional XPA protein that is essential for the proper functioning of the NER pathway (Levy et al., 1995), thus XP-A cells are completely deficient in NER. Expression analysis of the constructs containing S-THF revealed an 11% regain of fluorescence in repair proficient cells (MRC-5) relative to a fully functional EGFP construct (ts.613G). This result indicated that TM occurred in 11% of the transcripts synthesized from an S-THF template in the control cell line. An effect that was extremely enhanced in NER-deficient XP-A cells depicting a 79% of relative EGFP expression. Hence, corresponding to at least a 60-fold increase of the median EGFP fluorescence intensity compared to the control dA construct (**Figure 3-21**). Interestingly, this huge level of reversal to a fluorescent EGFP protein observed in XP-A cells was partially reversed (by 40%) after complementation of the cells with XPA protein (XP-A/+XPA). In contrast, BER-sensitive THF AP sites did not lead to any reactivation of the EGFP fluorescence, which indicates that BER of AP sites efficiently prevents transcriptional mutagenesis in all cell lines tested. This result confirmed the specific involvement of XPA protein, and therefore NER pathway, in the repair of synthetic AP lesions in human cells.

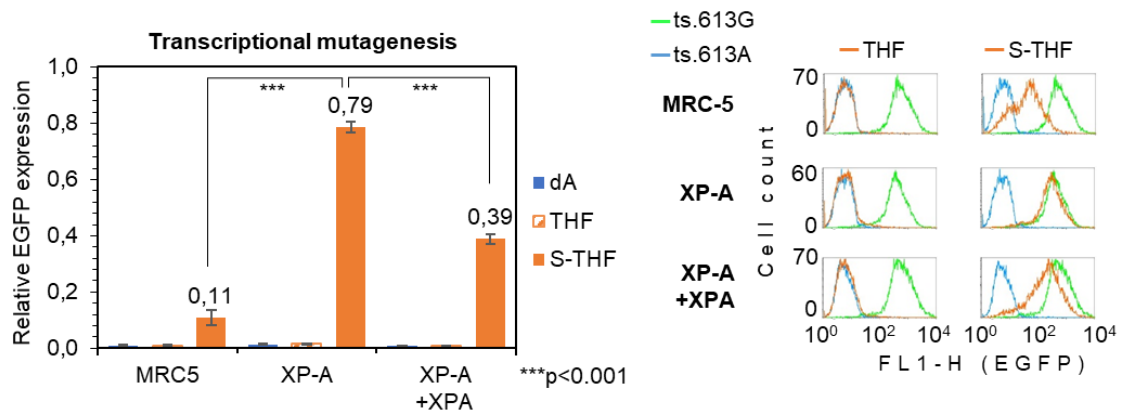


Figure 3-21: Transcriptional mutagenesis triggered by BER-resistant AP sites in NER deficient cell lines. Flow cytometry assay for detection of single nucleotide substitutions induced by the specified AP lesions (THF, S-THF) during transcriptional bypass. Left: Quantification of relative EGFP expression in MRC-5, XP-A and XP-A+XPA cell-lines ($n=3$) (mean \pm SD); Right: Overlaid fluorescence distribution plots from a representative experiment, ts.613G (green), ts.613A (blue) and constructs carrying synthetic AP sites ts.613THF/S-THF (orange). Figure adapted from Kitsera et al., 2019.

NER is divided into two different subpathways that differ in the damage recognition step: global genomic NER (GG-NER) and transcription-coupled NER (TC-NER). Therefore, it was interesting to examine whether repair of AP sites via NER mechanism occurred via GG-NER or TC-NER. Thus, we transfected THF and S-THF constructs into host cells with hindered TC-NER (CS-A and CS-B) and GG-NER (XP-C) subpathways (**Figure 3-22**). Predictably, BER-sensitive AP sites (THF) did not show any significant gain of fluorescence as a result of its efficient repair regardless of the cell line used. However, S-THF lesion produced an increase of EGFP fluorescence in all cell lines tested compared to the reference-line construct (dA) indicating that TM was occurring at the lesion. Interestingly, the relative EGFP expression levels increased in all NER deficient cell lines as compared to NER-proficient MRC-5 cells. As illustrated by the median EGFP fluorescence intensity, a defect in the TC-NER pathway doubled the frequency of reversion to functional EGFP (21-22%); no matter whether CSA or CSB was the affected gene. Inactivation of the GG-NER pathway (XP-C cells) caused an even stronger increase of the transcriptional mutagenesis rate (41%). Still, the gain of fluorescence in either GG-NER (4-fold) or TC-NER (2-fold) deficient cells did not concur with the effect observed in XP-A cells (8-fold) where both subpathways are impaired. In summary, these results indicated that NER pathway contributes to the repair of BER-resistant abasic sites and that GG-NER as well as TC-NER, are involved in the repair of this lesion to some extent.

Results

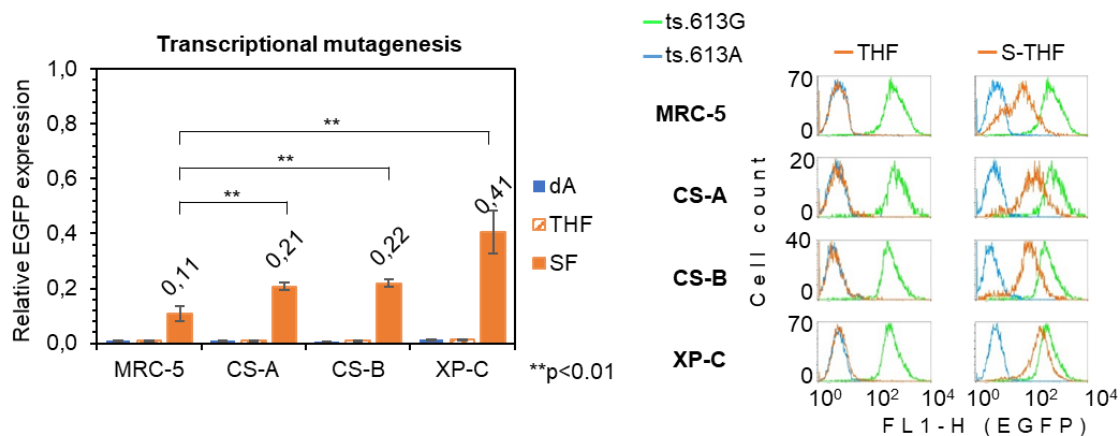


Figure 3-22: Transcriptional mutagenesis triggered by BER-resistant AP sites in NER deficient cell lines. Flow cytometry assay for detection of single nucleotide substitutions induced by the specified AP lesions (THF, S-THF) during transcriptional bypass. Left: Quantification of relative EGFP expression in MRC-5, CS-A, CS-B and XP-C cell-lines (n=4) (mean ±SD); Right: Overlaid fluorescence distribution plots from a representative experiment (right), green represents ts.613G, blue represents ts.613A and orange represents the constructs carrying synthetic AP sites ts.613THF/S-THF. Figure adapted from Kitsera et al., 2019.

3.2.4 Lack of TM induced by physiological AP sites indicates that this lesion may be repaired within the cells

To continue with the characterisation of transcriptional bypass of abasic DNA lesions, we decided to apply the TM assay using constructs containing natural abasic sites (**Figure 3-13**). In order to integrate natural AP lesions into the gapped pZAJ_Q205* mutant, insertion of three different synthetic oligonucleotides containing uracil in position ts.613 of the EGFP gene was performed. The constructs included unprotected uracil, a phosphorothioate linkage 5' to uracil (S-dU) and a uracil nucleotide with 5' and 3' linkage protection by a phosphorothioate bond (S-dU-S). An additional construct containing dA as a positive control for the non-mutagenic readout was used. The reason for using a protective linkage 3' to the uracil nucleotide was that, unlike the inherent resistance of THF AP sites to cleavage by β -lyases, natural AP sites are still good substrates of bifunctional glycosylases. Once the oligonucleotides were integrated into the vector, excision of the uracil by incubation with UDG would create a natural AP site. The ligation efficiencies of the synthetic oligonucleotides into the vector were >74% (**Figure 3-23A**).

To verify the quality of the constructs as well as their sensitivity to endonuclease activity, it was necessary to perform a digestion assay combining uracil DNA glycosylase (UDG) and Endonuclease IV (Endo IV) enzymes as described in method 2.2.18. In the course of the reaction, UDG catalyzes the uracil removal creating a physiological AP site, that further serves as a substrate for Endo IV which cleaves the phosphodiester bond 5' to the AP site. Results of the digestion assay showed a basal level of Endo IV activity towards preexisting AP sites of 5-7% in the control construct (dA) and 9-12% in the constructs containing uracil (**Figure 3-23B**). Treatment of Endo IV in combination with

UDG showed 100% conversion to open circular form of the plasmid in dU constructs. This result indicated first, that all plasmids contained dU modification and second, that uracils have been efficiently removed by UDG creating an AP site which is cleavable by Endo IV. Constructs harboring phosphorothioate linkages (S-dU and S-dU-S) only showed 15-20% incision upon incubation with UDG and Endo IV. This indicated that the sulfurization of the 5'-phosphodiester bond prevents the cleavage by Endo IV.

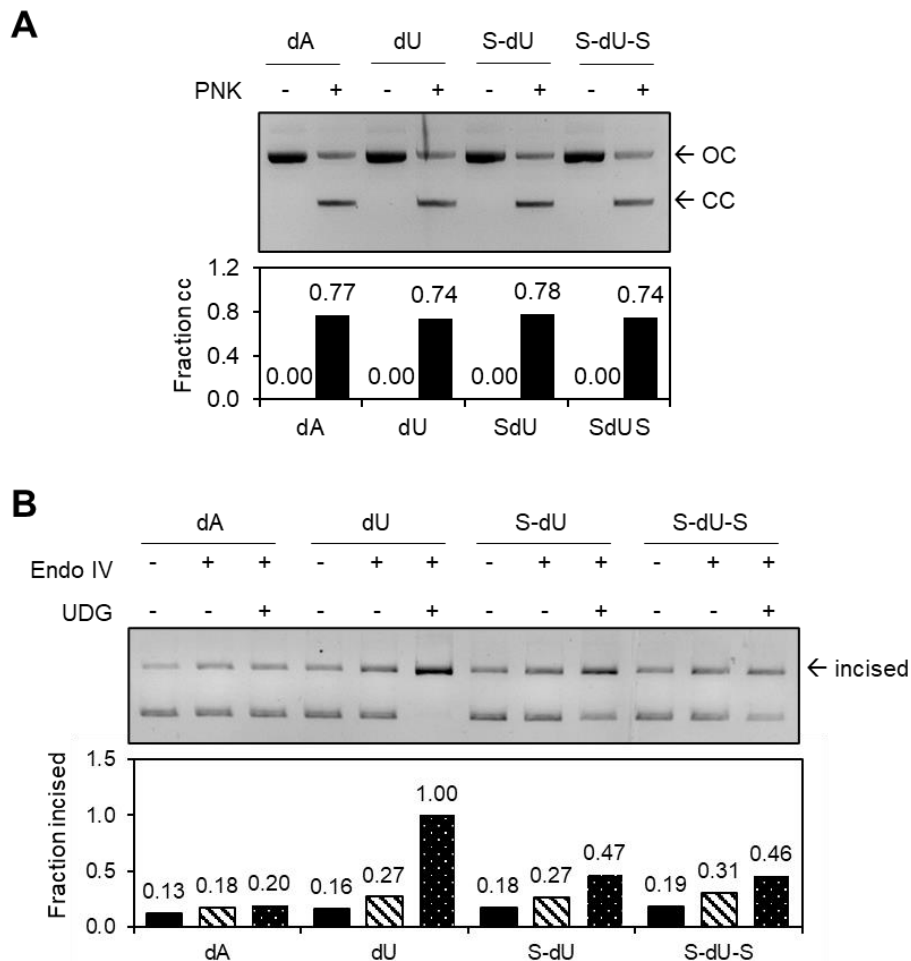


Figure 3-23: Preparation and characterisation of constructs with variations of uracil (U) modification in position ts.613 of the pZAJ_Q205* mutant reporter. (A) The efficient incorporation of four different synthetic oligonucleotides (dA, dU, S-dU, and S-dU-S) in position ts.613 is shown by the covalently closed (cc) form of the plasmid in presence of PNK. **(B)** Endo IV-UDG (2U/100ng) digestion assay in double-stranded DNA after preparative ligation. Quantification of the incised form of the plasmids depicted in the table underneath the agarose gel image.

Since this was the first experiment performed with physiological AP sites, we studied their sensitivity towards different endonucleases compared to THF AP sites. Endo IV was used as the AP endonuclease that catalyzes the cleavage of the phosphodiester bond 5' to the AP sites. Endonuclease III (Endo III) or Formamidopyrimidine DNA Glycosylase (Fpg) are bifunctional DNA glycosylases with AP lyase activities [2.2.19]. Even though THF AP sites were 100% digested by Endo IV, they showed complete

Results

resistance towards EndoIII or Fpg (**Figure 3-24**). The same percentage of open circular form of the plasmid (23-29%) as the control dA construct (23-28%) confirmed the inherent protection of synthetic AP sites (THF) from β -lyases activities. Sulfurization of the phosphodiester bond 5' to the THF AP site conferred resistance to Endo IV (the major AP endonuclease in bacteria), thus preserving almost completely the covalently closed form of the plasmid. In contrast, natural AP sites created by incubation of the dU construct with UDG were good substrates for all three endonucleases. Conversion of the covalently closed form into the open form of the plasmid directly illustrates EndoIII, Endo IV, and Fpg activities towards this lesion. Once again, sulfurization of the phosphodiester bond 5' to natural AP site (S-dU) protected the constructs towards Endo IV activity; however, these constructs remained good substrates for Endo III and Fpg. Here, the AP-lyase activities of Endo III and Fpg that catalyze cleavage reactions 3' to the lesion were not affected by modifications of the 5' linkage.

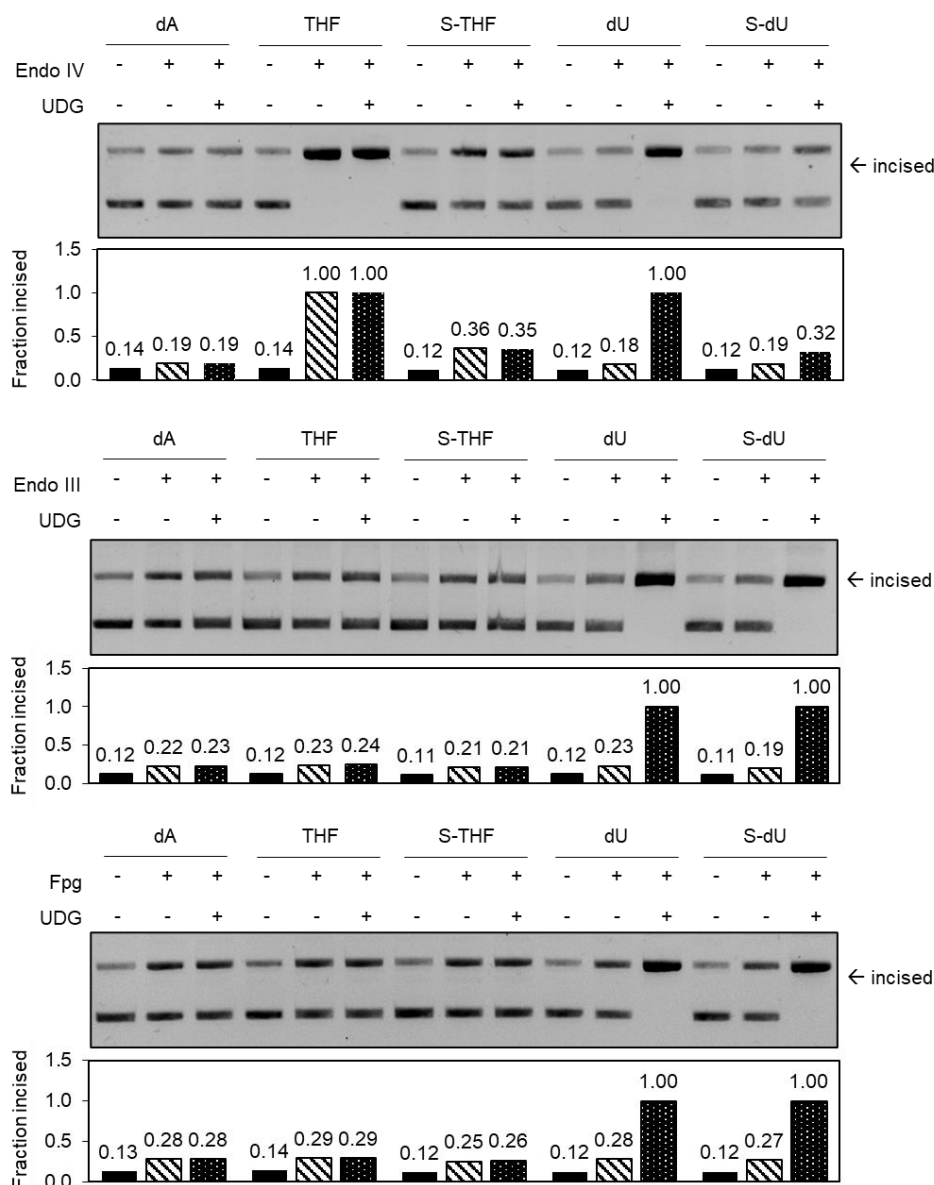


Figure 3-24: Analysis of the sensitivity of various types of AP lesions towards Endo IV, Endo III and Fpg activities. Digestion of *ts.613A*, *ts.613THF*, *ts.613S-THF*, *ts.613U*, *ts.613sU* constructs derived from *pZAJ_Q205** reporter using different glycosylases such as endonuclease IV (Endo IV); endonuclease III (Endo III), Formamidopyrimidine DNA Glycosylase (Fpg) in combination with Uracil DNA glycosylase (UDG). All enzymes listed have been used at a concentration of 2U/100ng DNA.

As mentioned at the beginning of the section, the aim was to create physiological AP sites by an *in vitro* UDG treatment [2.2.20] of the constructs accommodating uracil modifications (**Figure 3-25A**). We, therefore, incubated dU, S-dU and S-dU-S constructs with UDG followed by a standard phenol-chloroform DNA purification protocol [2.2.21]. To verify the presence of the newly created abasic DNA lesions, it was necessary to perform a digestion assay using Endo IV and Endo III enzymes. In the presence of unprotected natural AP sites (AP), both Endo IV and Endo III completely digested the constructs as illustrated by 100% of the open circular form of the plasmid (**Figure 3-25B**). Natural AP sites protected with a 5'-phosphorothioate bond (S-AP) showed a 50%

Results

cleavage rate, thus indicating that these AP sites were relatively resistant to Endo IV endonuclease activity. Conversely, the same construct showed complete digestion by Endo III suggesting that S-AP is still a good substrate for enzymes with β -lyase activities. Sulfurization of both 5' and 3' phosphodiester bonds (S-AP-S) protected natural AP sites from endonuclease, as well as β -lyase activities. Thus, mild conversion of only 54% of the plasmid into the incised form in presence of an excess of EndoIII or Endo IV indicated that this substrate is mostly resistant to incision by endonucleases.

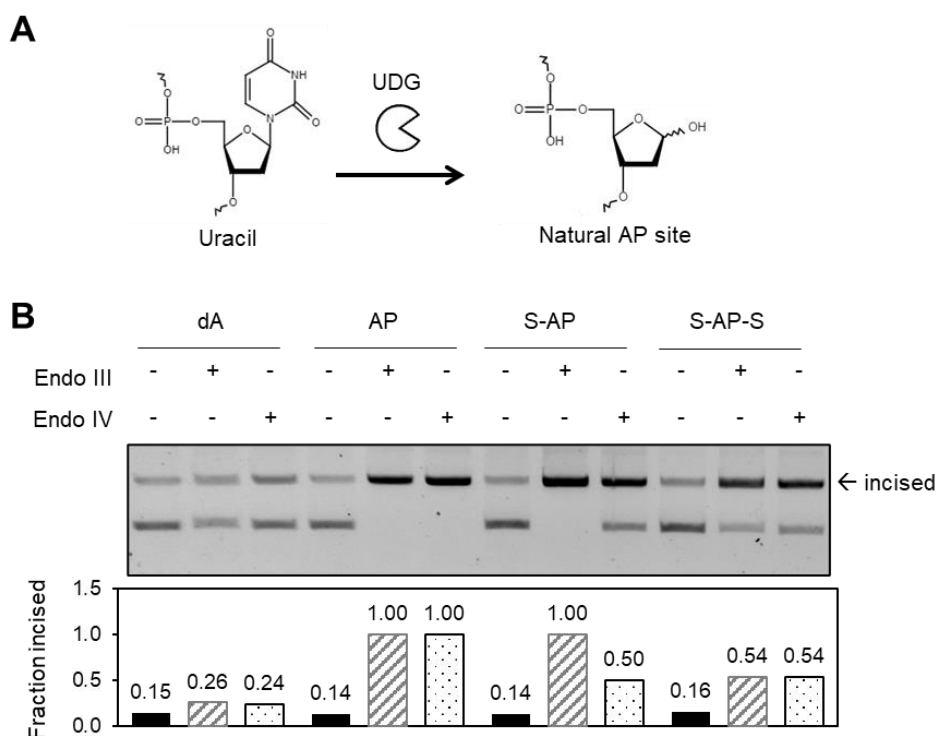


Figure 3-25: Preparation of constructs with natural AP sites in position *ts.613* of the pZAJ_Q205* mutant vector. (A) Schematic representation of the UDG digestion performed in the *ts.613A*, dU, S-dU, and S-dU-S constructs to create a natural AP site. (B) EndoIII and Endo IV (2U/100ng) digestion assays in double-stranded DNA after UDG digestion and phenol-chloroform purification. Quantification of the incised form of the plasmids depicted in the table underneath the agarose gel image.

Like the tetrahydrofuran (THF) placed in the transcribed DNA strand of the pZAJ_Q205* mutant, natural abasic sites (AP) were also expected to induce transcriptional mutagenesis. To investigate the mutagenic outcome induced by erroneous transcription of this type of abasic lesion, we transfected the AP, S-AP, and S-AP-S constructs into NER-proficient (MRC5 and XP-A+XPA) and NER-deficient (XP-A) cell lines. Surprisingly, none of the constructs induced a detectable fraction of mutant EGFP proteins that resulted in the re-gain of EGFP fluorescence (**Figure 3-26**).

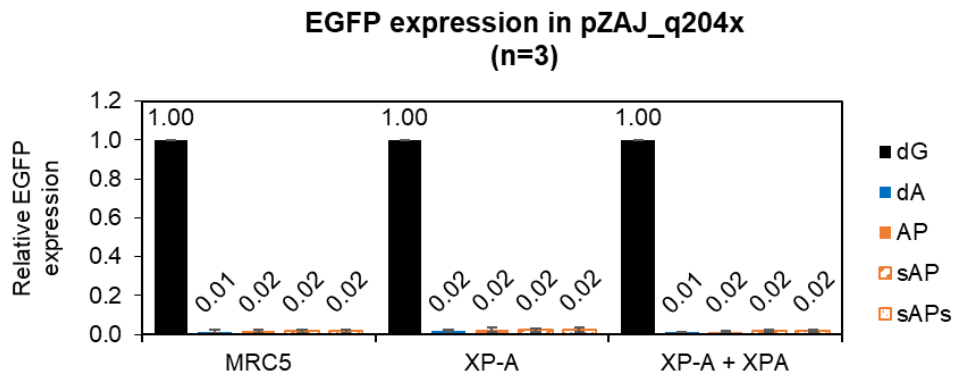


Figure 3-26: Relative EGFP expression triggered by BER-resistant natural AP sites constructs in NER proficient (MRC5 and XP-A+XPA) and NER deficient (XP-A) cells. Flow cytometry assay for detection of the mRNA single nucleotide substitutions induced by the specified AP. Quantification of relative EGFP expression in all cell lines of three independent experiments (mean \pm SD).

In order to explain this result, we considered four different hypotheses. First, there could be a complete repair of natural BER-resistant AP sites within the NER deficient cell line (XP-A) by an unknown mechanism. Second, natural AP sites might be strong blocks for RNA pol II as has been previously reported *in vitro* (Tornaletti et al., 2006). Third, RNA pol II could incorporate uracil by default opposite to physiological AP sites during transcriptional bypass restoring the STOP codon and therefore, resulting in a non-fluorescent protein. Fourth, since natural AP sites are a more reactive form of the lesion, the hydroxyl group in C1 might react during transfection or within the cell disrupting the construct. To address the possibility that physiological AP sites are repaired by an alternative repair pathway, we transfected a broad range of human cellular backgrounds with AP, S-AP, and S-AP-S constructs. We analyzed the EGFP expression in two mismatch repair (MMR) deficient cell lines: DLD1 (MSH6 deficient) and LoVo (MSH2 deficient). In addition, we transfected two cell lines deficient in the ERCC1-XPF complex that, apart from NER, also works as a 5' structure-specific endonuclease during interstrand cross-links (ICL) repair. However, transfection of the natural AP site constructs into none of these DNA repair-deficient cell lines displayed any trace of EGFP fluorescence (**Figure 3-27**).

Results

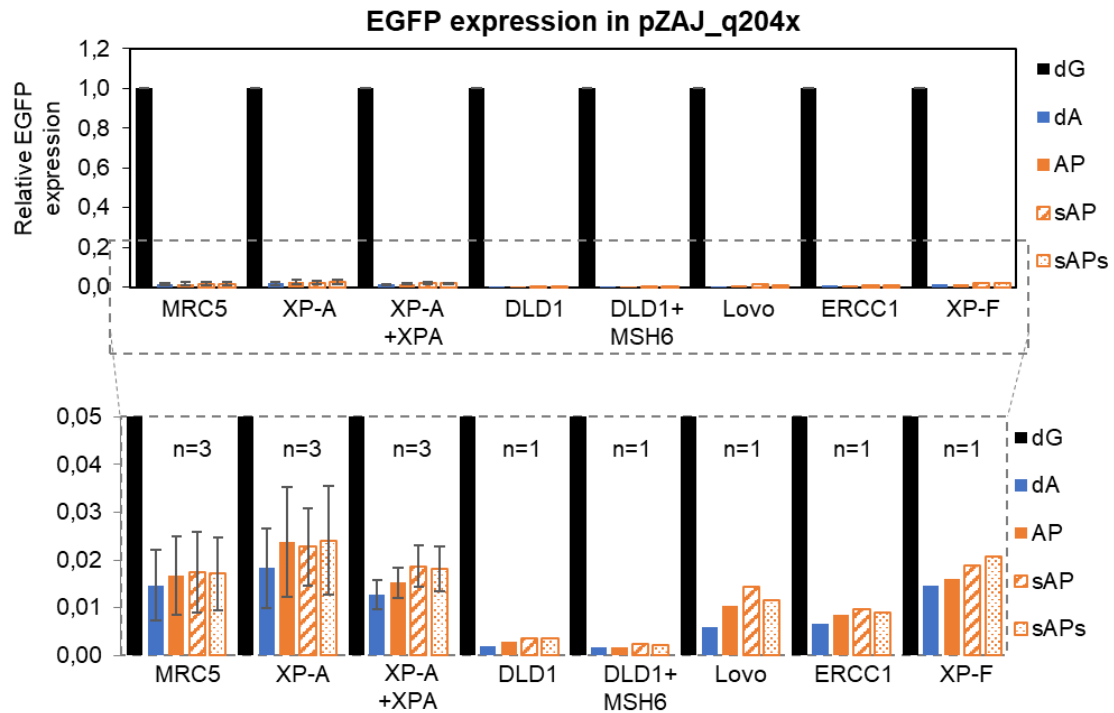


Figure 3-27: Transcriptional mutagenesis triggered by BER-resistant natural AP sites is undetectable in all cell lines tested. Flow cytometry assay for detection of the mRNA single nucleotide substitutions induced by the specified AP lesions in repair-deficient cell lines. Quantification of relative EGFP expression in all cell-lines (mean \pm SD).

To test the remaining hypotheses, we selected a fully functional wild-type EGFP coding vector (pZAJ_5c) that only differs from the pZAJ_Q205* reporter in one nucleotide residue c.613T>C. Here, insertion of dU, S-dU and S-dU-S modifications into the template strand occurred opposite to a cytosine rather than a thymine (**Figure 3-28**).

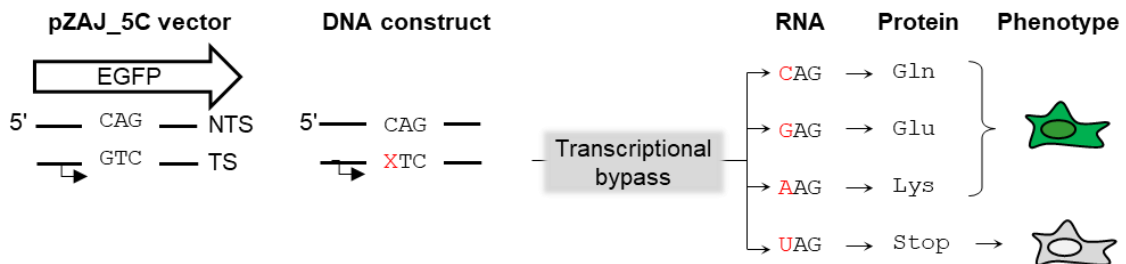


Figure 3-28: Scheme of the pZAJ_5c reporter vector. The red X indicates the position of the nucleotide 613 in the template strand (TS) that has been substituted by an abasic lesion. After transcriptional bypass of the lesion, all possible outcomes are indicated at RNA, protein and phenotypic level.

As illustrated by the analytical ligation, all four synthetic oligonucleotides were successfully inserted in the TS of a gapped pZAJ_5c plasmid (**Figure 3-29A**). Like in constructs derived from pZAJ_Q205* backbone, only dU showed complete digestion by Endo IV after uracil excision (**Figure 3-29B**). Samples containing phosphorothioate linkages in either, 5' or 5' and 3', showed only partial incision demonstrating their resistance towards the AP endonuclease activity of Endo IV.

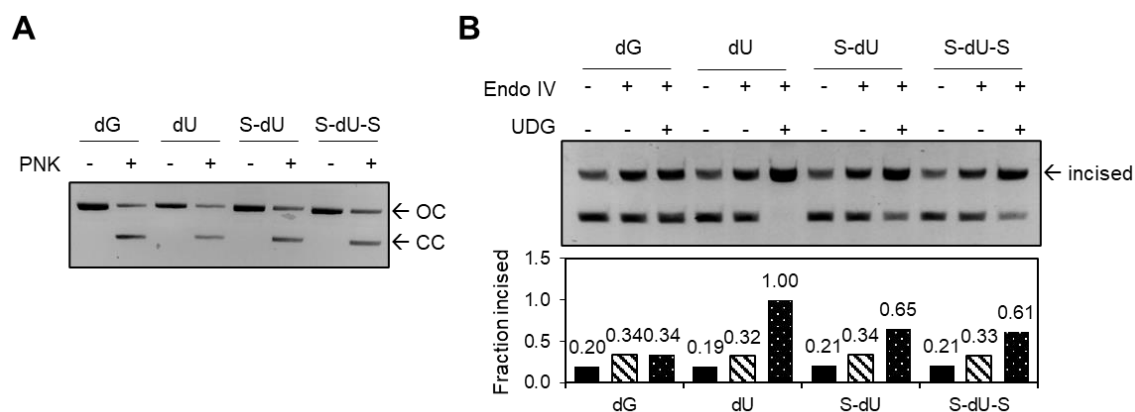


Figure 3-29: Preparation of constructs with variations of U in position ts.613 of the pZAJ_5c vector. (A) The efficient incorporation of four different synthetic oligonucleotides (dG, dU, S-dU, and S-dU-S) in position ts.613 is shown by the covalently closed (cc) form of the plasmid in presence of PNK. (B) Endo IV-UDG (2U/100ng) digestion assay in double-stranded DNA after preparative ligation. Quantification of the incised form of the plasmids depicted in the table underneath the agarose gel image

Subsequently, treatment of the new templates with UDG led to the excision of uracil and the creation of physiological AP sites. APE1 digestion assay verified the successful formation of natural AP sites, since 100% of the constructs surrounded by phosphodiester bonds acquired an open circular form (Figure 3-30). In addition, this experiment showed that sulfurization of the linkage 5' to the lesion made the S-AP and S-AP-S constructs resistant to APE1 endonuclease activity.

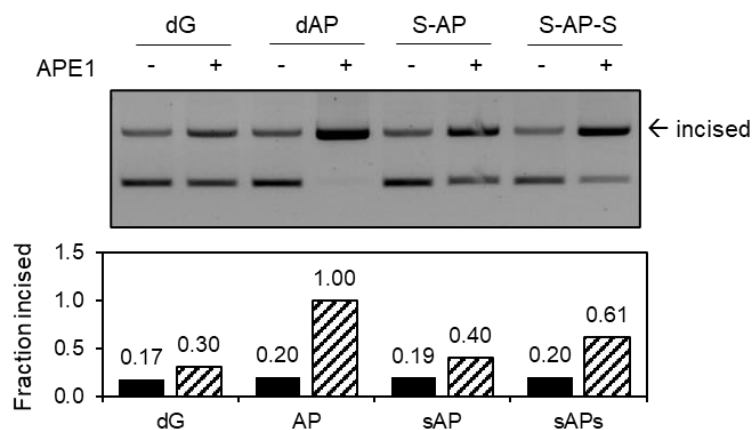


Figure 3-30: Verification of the presence of natural AP sites in position ts.613 of the pZAJ_5c reporter. APE1 (5U/100ng) digestion assay in double-stranded DNA after UDG digestion. Quantification of the incised form of the plasmids depicted in the table underneath the agarose gel image.

All the templates carrying natural AP sites opposite to a cytosine transfected into repair-proficient and deficient cell lines generated a very strong EGFP fluorescent signal regardless of the presence of phosphorothioate linkages (Figure 3-31). The slight decrease in EGFP fluorescence detected by BER-sensitive AP constructs, correlated with previous findings in the group demonstrating that BER incision leads to transcriptional decline due to the ss-break generated in the template (Lühnsdorf et al.,

Results

2014). Detection of EGFP fluorescence in all cell lines transfected with s-AP and s-AP-s constructs suggested that either repair or transcriptional bypass occurred. If transcriptional bypass of the abasic lesion in these templates took place, it was certainly not solely through uracil incorporation since that would be the only nucleotide that would encode for a non-fluorescent EGFP (**Figure 3-28**). Furthermore, as S-AP and S-AP-S templates revealed high EGFP fluorescent values in all cell lines, we concluded that the constructs have entered the cell and the nucleus without disrupting the structure of the plasmid.

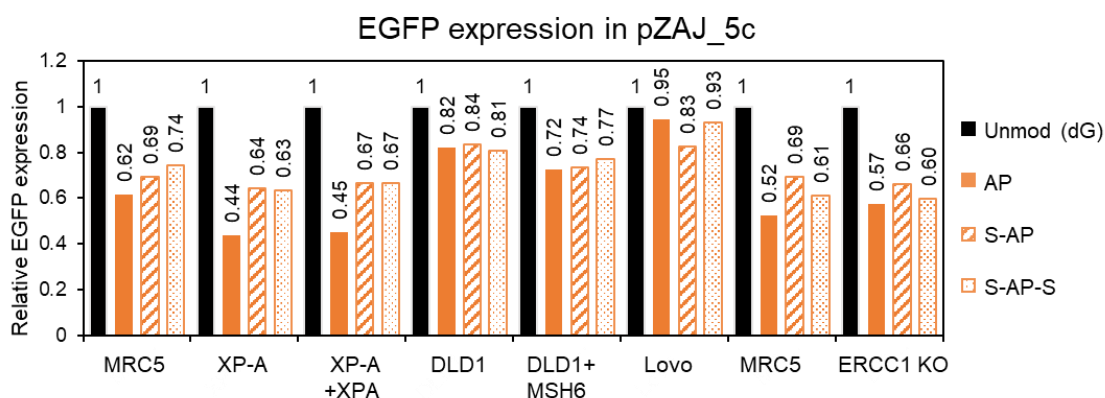


Figure 3-31: EGFP expression of human cells transfected with ts.613dG/AP/sAP/sAPs opposite to C constructs. Flow cytometry assay for detection of relative EGFP expression induced by the specified AP lesions in repair-deficient cell lines ($n=1$); MRC5 in position 7 corresponds to the parental cell line used to create the ERCC1 KO.

In summary, **Figure 3-27** indicated that natural AP-sites derived from pZAJ_Q205* mutant did not induce any EGFP reactivation by TM even when protected from the BER pathway by chemical modification of the DNA backbone. This suggested that the stop codon has been accurately restored in DNA. Based on the obtained results, we deduce that natural abasic sites in transcribed DNA can undergo repair in a BER- and NER-independent manner, which would imply the presence of a hitherto unknown repair mechanism in human cells.

3.2.5 Transcriptional bypass analysis of uracil and abasic ribonucleotide lesions

Since deamination of cytosine results in uracil formation and uracil removal accounts for a great proportion of AP site formation under physiological conditions (Krokan et al., 2002), we aimed to investigate transcriptional bypass of uracil placed in ds-DNA. Thus, we transfected templates for TM detection harboring uracil in position ts613 of the pZAJ_Q205* vector which did not undergo *in vitro* UDG treatment (**Figure 3-32A**). As expected, we did not detect any reversal to a fluorescent EGFP in cells transfected with dA or dU templates since uracil undergoes an efficient repair in cells (**Figure 3-32B**). However, we detected about 5-11% EGFP expression in cells transfected with sU and

sUs constructs indicating that efficient BER of these lesions mostly prevents transcriptional mutagenesis.

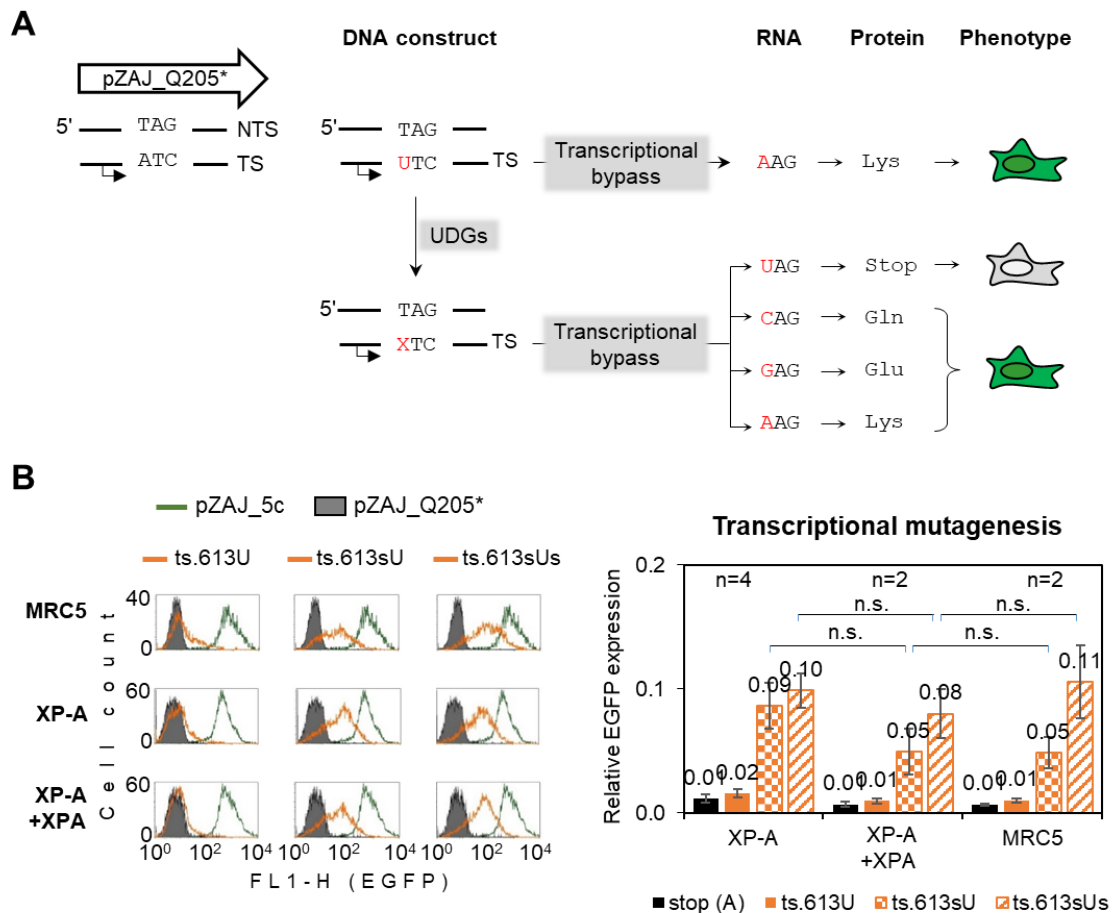


Figure 3-32: Transcriptional mutagenesis triggered by uracil (ts.613U) in NER deficient (XP-A) and proficient (XP-A+XPA) cell lines. (A) Scheme of the construct used in this experiment. All possible outcomes that emerge from the transcriptional bypass of the uracil or the resulting AP site (X) are indicated at RNA, protein and phenotypic levels. **(B)** Left: Overlaid fluorescence distribution plots from a representative experiment (left), green represents ts.613G, grey represents ts.613A and orange represents the constructs carrying uracil with different backbone modifications in pZAJ_Q205* vector. Right: Quantification of the relative EGFP expression of the ts.613A, ts.613U, ts.613sU and ts.613sUs constructs relative to the functional EGFP construct (c.613G); XP-A (n=4); XP-A+XPA (n=2) and MRC5 (n=2) mean \pm SD.

Since natural AP sites in the same position did not induce any fluorescence (**Figure 3-27**), we are assuming that a small fraction of residual fluorescence appeared from the transcriptional bypass of still unrepaired uracil. Bypass of dU should be accompanied by adenine incorporation into the nascent mRNA creating AAG codon (**Figure 3-32A**). The resulting lysine residue would reactivate the protein fluorescence as shown previously (pZAJ_Q205* T:T scatter plot of **Figure 3-9**). Surprisingly, sU and sUs constructs did not lead to higher EGFP expression in NER deficient (XP-A) compared to NER proficient cell line (MRC5) suggesting that the NER pathway does not significantly contribute to repair of either uracil or natural AP sites in this context.

Results

Because our system has been proved to be suited for the detection of TM by THF AP sites and a recent report proposed that abasic ribonucleotides are processed by human APE1 *in vitro* (Malfatti et al., 2017), we also performed the transcriptional mutagenesis assay with constructs containing ribo-THF AP sites (**Figure 3-33A**). Thus, we generated constructs originated from the pMR_A207P mutant and accommodated ribo guanosine (rG), 5'phosphorothioate protected rG (SrG), ribo tetrahydrofuran AP site (rTHF) and rTHF with 5'phosphorothioate linkage (SrTHF) into the vector. *In vitro* APE1 digestion assay showed that, unlike THF constructs that showed 100% conversion to the incised form, rTHF showed only 56% cleavage by APE1. This result indicates that ribo-THF is not a good substrate for APE1 (**Figure 3-33B**). Surprisingly, 5'-phosphorothioate THF (S-THF) and rTHF (S-rTHF) did not confer any additional protection towards APE1 cleavage suggesting that the phosphorothioate linkages have not been properly incorporated into these synthetic oligonucleotides. In addition, the FACS analysis showed a complete absence of fluorescence in XP-A and XP-A/+XPA cell lines (**Figure 3-33C**). Lack of EGFP signal indicated that either these lesions were efficiently repaired within the cells or that the nature of the transcriptional bypass was non-mutagenic though cytosine incorporation opposite to the lesion. Since the phosphorothioate linkages did not protect from APE1 incision in these constructs and this enzyme partially cleaved r-THF in the *in vitro* assay, we cannot eliminate the possibility that APE1 could have repaired the lesion within the cells.

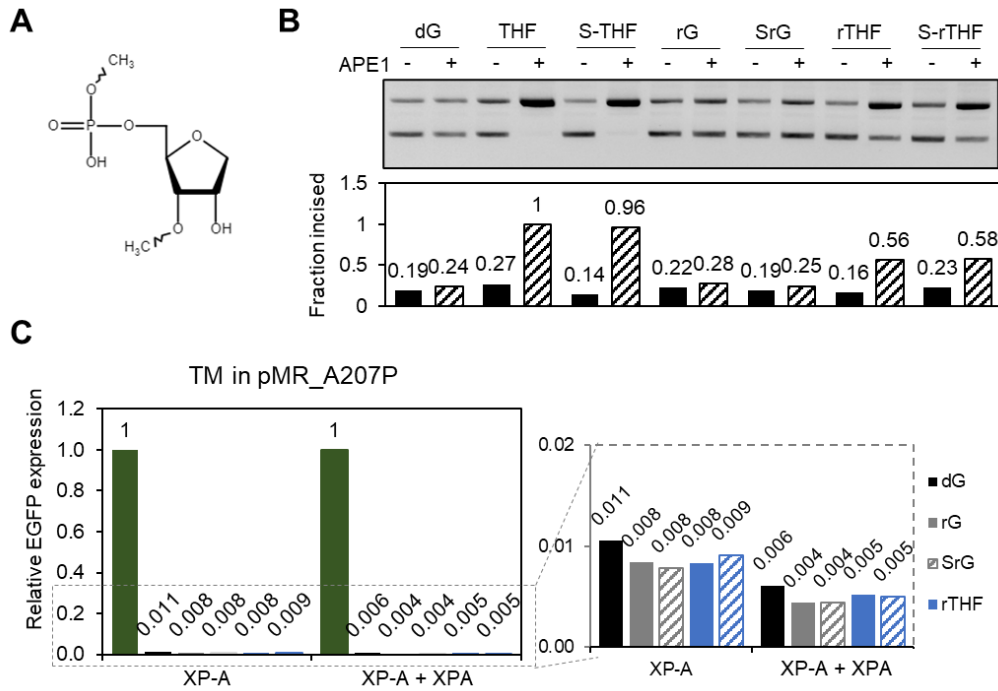


Figure 3-33: Transcriptional mutagenesis triggered by ribonucleotide abasic DNA lesions in NER deficient (XP-A) and proficient (XP-A/+XPA) cell lines. (A) Chemical structure of a ribo-tetrahydrofuran modification (B) APE1 digestion assay (0-5U APE1/100 ng) in double-stranded DNA after containing guanine (dG), tetrahydrofuran (THF), 5' phosphorothioate THF (s-THF), ribo-guanine (rG), 5'phosphorothioate protected ribo-G (SrG), ribo-tetrahydrofuran (rTHF) and 5'phosphorothioate protected ribo-tetrahydrofuran (S-rTHF) in position ts.619 of pMR_A207P vector. Quantification of the incised form of the plasmids depicted in the table underneath the agarose gel image. (C) Quantification of the relative EGFP expression of the ts.619G, ts.619rG, ts.619S-rG, ts.619rTHF, ts.619S-rTHF constructs relative to the functional EGFP construct (c.619C) in XP-A and XP-A/+XPA.

In summary, in constructs harboring different types of BER-resistant AP lesions, we detected transcriptional mutagenesis events occurring during RNA synthesis over BER-resistant synthetic (S-THF) abasic lesions. Furthermore, with the help of S-THF lesion, we disclosed the role of nucleotide excision repair (NER) as a backup repair mechanism for AP lesions; based on the finding that mutation rates at the affected position are greatly increased in NER-deficient cell lines. In contrast, expression analysis of constructs containing BER-resistant natural abasic sites did not reveal any gain fluorescence in NER-proficient nor NER-deficient cell lines. Thus, it is very likely that this lesion was efficiently repaired in human cells.

Results

3.3 Mutagenic bypass of abasic DNA lesions during DNA synthesis

3.3.1 Generation of reporter constructs to detect mutagenic TLS over an abasic lesion

The mutant EGFP reporters generated in the results section 3.1 can also be used to detect errors of translesion DNA synthesis (TLS) over AP lesions. In order to study DNA synthesis directly through the damaged template, it was necessary to place abasic lesions into the coding strand of the vector opposite to a gap. In transfected cells, the gap must be filled by DNA synthesis for the transcription to occur. The synthesis of the missing strand, only if accompanied by nucleotide misincorporation opposite to the lesion, would encode for a fluorescent EGFP. Once a nucleotide is misincorporated opposite to the lesion, the restored ds-DNA would serve as a substrate for the subsequent repair of the lesion fixing the mutation into the plasmid DNA.

To prepare constructs for detection of mutagenic translesion synthesis, it was required to generate an 18-nucleotide gap in the non-template strand (NTS) of one of the mutant vectors. The resulting gap was used to insert synthetic oligonucleotides carrying an abasic site at the desired position followed by the depletion of the 18-nucleotide fragment from the template strand (TS) opposite to the lesion (**Figure 3-34**).

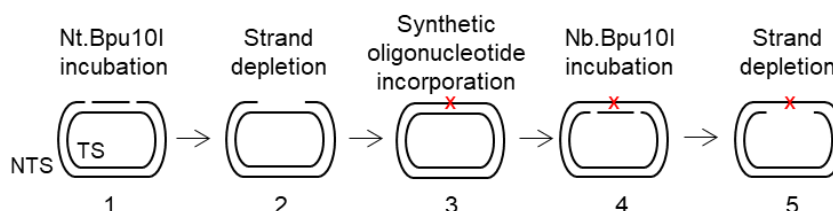


Figure 3-34: Scheme depicting the preparation of constructs for detection of mutagenic TLS. (1) Incubation with *Nt.Bpu10I* nicking enzyme to create tandem nicks on the non-template strand (NTS) of the vector. (2) After incubation with competitor oligonucleotide and AMICON filtration, the strand positioned in between the nicks in the NTS is eliminated. (3) Annealing and ligation of the synthetic oligonucleotide without or with desired modification in the NTS. (4) Incubation with *Nb.Bpu10I* nicking enzyme to create tandem nicks on the template strand (TS) of the vector. (5) Incubation with competitor oligo for TS followed by Amicon filtration, creation of the gap in TS opposite to the modification

Validation of the TLS assay by using the mutant EGFP reporters started by generating gapped constructs from the pZAJ_Q205* mutant harboring a synthetic AP lesion (THF). In this plasmid, incorporation of guanine, cytosine or thymine opposite to the lesion during DNA synthesis would lead to the formation of a fluorescent EGFP (**Figure 3-35A**). Incubation of plasmid with strand-specific *Nt.Bpu10I* nicking endonuclease generated two nicks placed 18-nucleotides apart at the NTS of the EGFP gene [2.2.7]. After depletion of the DNA stretch located between the nicking sites [2.2.10], the gapped plasmid was analyzed by analytical ligation as described in the Methods section [2.2.8].

In this experiment, we incubated the previously gapped plasmid in presence of T4 DNA ligase, ligase + synthetic oligonucleotides, and ligase + synthetic oligonucleotides + PNK in four different reactions. Even though the stretch of 18-nucleotides placed between the nicking sites had been eliminated by strand depletion, incubation of the vector with ligase led to the formation of a covalently closed (cc) form of the plasmid (**Figure 3-35B**). By nicking any plasmid originated from pZAJ_5c backbone in the NTS with Nt.Bpu10I enzyme, the ss-DNA stretch originated after depletion of the hanging 18-nucleotides forms a hairpin-like secondary structure (Lühnsdorf et al., 2012). Thus, the ligation of the 5'-phosphate end with the 3'-hydroxyl end generated a plasmid that ran as a covalently closed form. In parallel, the sample containing 45x excess of synthetic oligonucleotide exclusively showed an open circular (oc) form of the plasmid. That is because the synthetic oligonucleotides annealed with the ss-DNA stretch preventing the formation of the hairpin structure. Only by addition of PNK, the nicked vector run as closed circular (cc) form due to the phosphorylation of the synthetic oligonucleotides at the 5'-OH that allowed their ligation into of the plasmid. Lack of cc form of the plasmid in the absence of PNK indicated that only unphosphorylated oligonucleotides annealed into the vector and therefore, we concluded that the plasmid has been successfully gapped.

As illustrated by the preparative ligation in the absence and presence of PNK, three different synthetic oligonucleotides were successfully incorporated into the NTS gap with >74% ligation efficiency (**Figure 3-35C**). One construct carrying thymine in position c.613 (T) as a control for regular DNA synthesis through an undamaged template and two constructs carrying synthetic abasic DNA lesions unprotected (THF) and protected (S-THF) from APE1 activity. To assess the presence of synthetic abasic sites (THF) and the resistance of S-THF to endonuclease excision, Endonuclease IV (Endo IV) digestion assay was performed [2.2.22]. Complete digestion of the c.613THF construct (100% incised) in presence of Endo IV indicated that synthetic abasic sites were present in all plasmid copies (**Figure 3-35D**). In addition, incomplete excision of the S-THF construct indicated that 5'-phosphorothioate linkage protected THF abasic sites from Endo IV cleavage.

Results

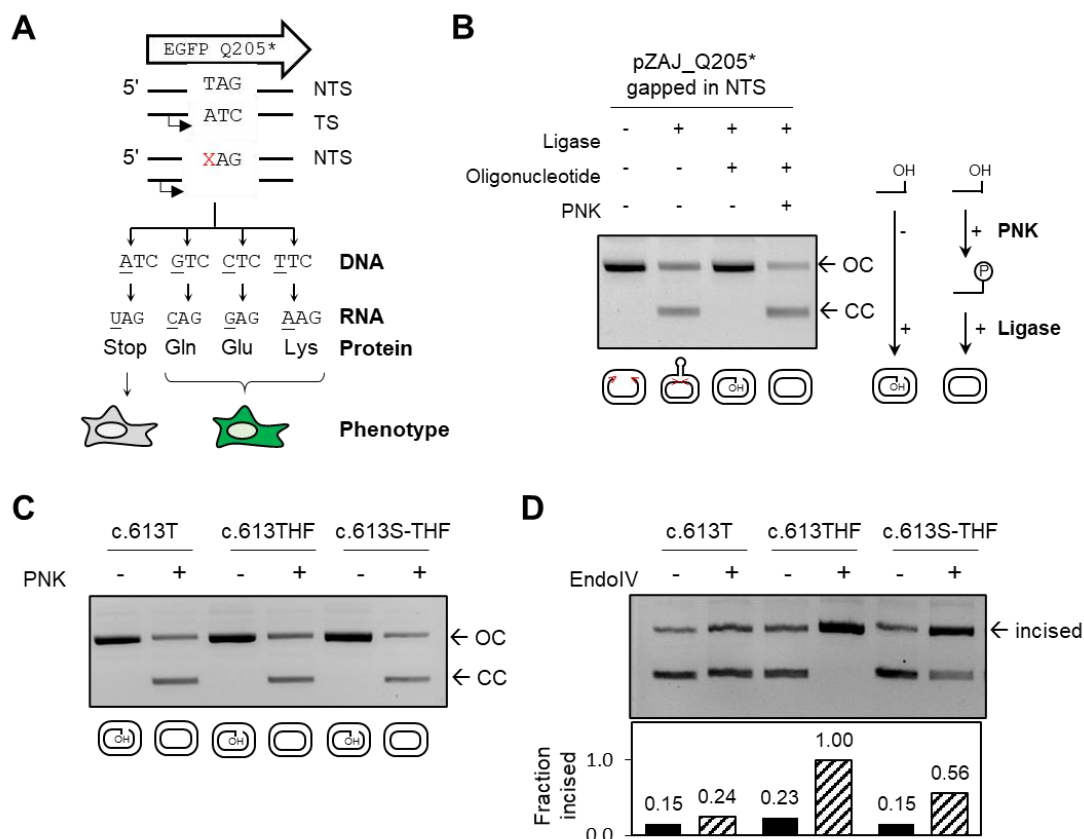


Figure 3-35: Generation of constructs with site-specific THF AP sites in the coding strand of the pZAJ_Q205* mutant vector. (A) Scheme of the ligation reaction in the absence and presence of PNK **(B)** Representative experiment of an analytical ligation of gapped pZAJ_Q205* vector in the coding strand; open circular (oc) and covalently closed (cc) forms of the plasmid are indicated with arrows, pictograms below the gel show schematic representation of reaction products **(C)** Efficient incorporation of the three different synthetic oligonucleotides (dT, THF and S-THF) in position c.613 is shown in the absence and presence of PNK generating a covalently closed (cc) form of the plasmid. **(D)** Endo IV digestion assay of the constructs using 5U Endo IV/100 ng of DNA and quantification of the open circular (oc) fraction of the plasmid.

The following step involved the generation of an 18-nucleotide gap opposite to the lesion. Thus, we incubated the newly generated constructs with strand-specific Nb.Bpu10I nicking endonuclease as described in the methods section 2.2.7. Generation of two tandem nicks in the template strand (TS) of the constructs, was followed by strand depletion of the excised DNA fragments. Subsequent analytical ligation confirmed the presence of the gap in the TS of each construct. The absence of cc form of the plasmid in the samples incubated with only T4 DNA ligase indicated a complete elimination of the 18-mers DNA stretch placed between the nicking sites (**Figure 3-36**). This result demonstrates that the constructs containing T, THF and S-THF in position c.613 also contain a gap in the opposite strand.

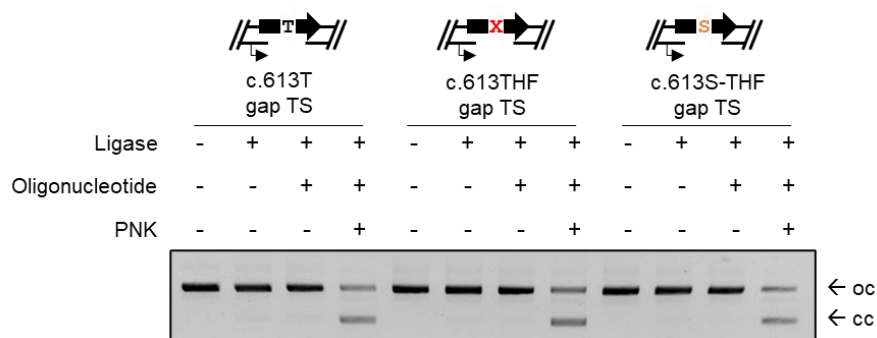


Figure 3-36: Analytical ligation of constructs with modifications in position 613 of the non-template strand (NTS) opposite to an 18-nucleotide gap into the template strand. Verification of gapped plasmids is shown by the absence of cc form of the plasmid in presence of ligase

It has been shown that APE1 is less efficient cleaving abasic substrates in single-stranded DNA than in double-stranded DNA (Fan et al., 2006; Marenstein et al., 2004; Wilson, 2003). Hence, to study the sensitivity of the new constructs towards APE1, plasmids containing THF in ds-DNA as well as ss-DNA were analyzed. In the control constructs containing thymine in ss- and ds-DNA no effect was observed upon incubation with the APE1 enzyme. The total conversion (100%) of the cc form of the plasmid containing THF in ds-DNA into oc form correlated with a regular APE1 incision rate of this substrate (**Figure 3-37**). In contrast, in ss-DNA, APE1 showed only 6% cleavage of THF illustrated by the conversion of the gapped plasmid (without APE1) into the linear form of the plasmid (lin). This result indicates that despite the high efficiency of APE1 cleaving AP sites in ds-DNA, this enzyme is far less efficient towards a substrate placed on single-stranded DNA.

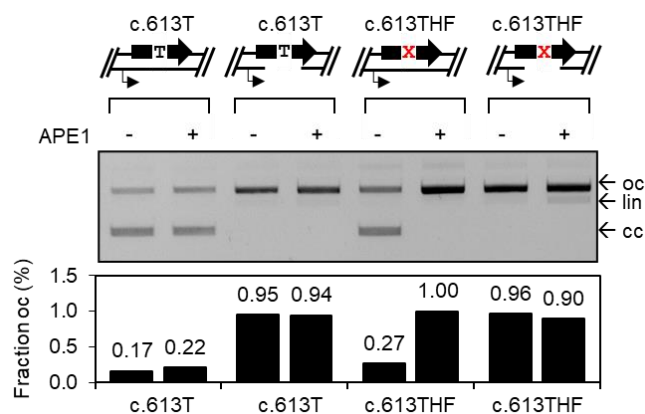


Figure 3-37: Analyses of APE1 activity towards AP lesion (THF) in ss-DNA and ds-DNA. The constructs containing the specified AP lesion were incubated with 1U/100ng of vector DNA. For ds-DNA, APE1 activity was measured by conversion of the covalently closed DNA (cc) into the open circular form (oc); graphical bars show % of constructs in oc form.

Results

3.3.2 Translesion synthesis over tetrahydrofuran abasic lesion is mutagenic

To detect mutagenic TLS over THF abasic lesions, we transfected the pZAJ_Q205* derived constructs containing T, THF or S-THF in position c.613 opposite to a gap into HeLa cells. Thereby, the transfected cells are compelled to seal the single-stranded gap in the TS using as a template the NTS strand where the abasic lesion has been placed. Since the coding sequence for the essential EGFP 203-208 fragment is only available in the damaged strand, translesion DNA synthesis (TLS) must take place in order to generate double-stranded plasmids. Only the constructs that gained G, T, or C nucleotides opposite to the lesion would subsequently yield a fluorescent EGFP protein (**Figure 3-38**).

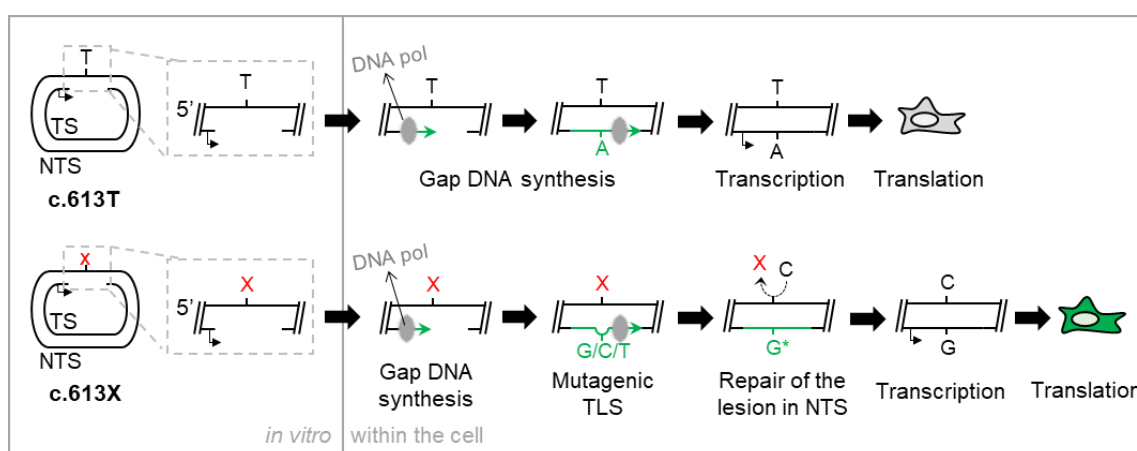


Figure 3-38: Reporter reactivation assay for direct detection of mutagenic TLS using pZAJ_Q205* as a model vector. Outcomes of transfection with the control construct (c.613T) and AP site (c.613X) construct for mutagenic TLS assay, being G* one of the three possible nucleotide misincorporations (G/C/T) occurring during mutagenic TLS which would reactivate the fluorescence of the EGFP protein.

To follow the repair efficiency of the gap occurring within the cells, we additionally generated a wild-type EGFP construct (c.613C) harboring an 18-nucleotide gap in the TS at the same position as the other TLS constructs. As illustrated by the FACS analysis performed 24 hours after transfection, 83-84% of cells transfected with gapped EGFP c.613C construct showed EGFP fluorescence (**Figure 3-39**). In this context, a broad distribution of cells in the upper right (UR) quadrant indicated that the repair efficiencies were heterogeneous among the cell population. To calculate the repair rate of the gap for the c.613C construct and the mutation frequencies for the rest of the templates derived from pZAJ_Q205*, the percentages of cells showing green fluorescence in the upper right (UR) quadrant were divided by the total percentage of transfected cells located in the UR and UL (upper left) quadrant.

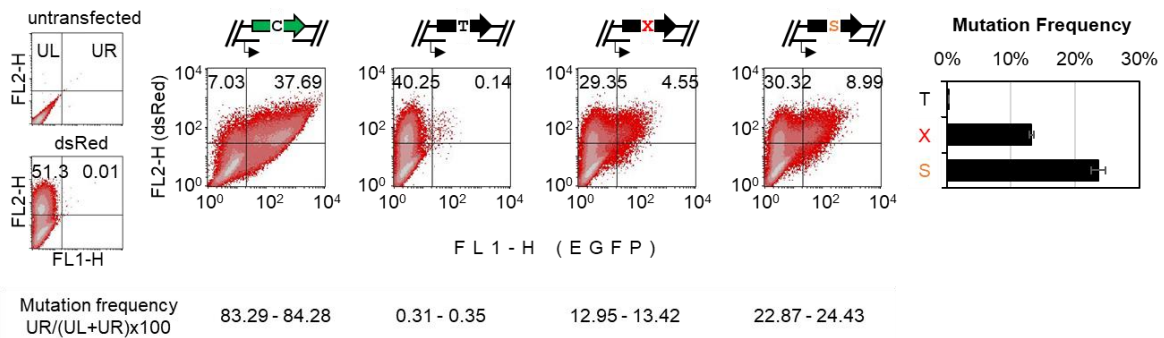


Figure 3-39: Detection of mutagenic TLS on the gapped pZAJ_Q205* template containing THF and S-THF at position c.613 transfected to HeLa cells. Cells expressing fluorescent EGFP were analyzed by FACS 24 hours post-transfection without and with a single THF AP site (X) and S-THF (S) opposite to an 18-nt gap. The repair rate (for the wild-type EGFP) and the mutation frequency (for c.613T, THF, and S-THF) were calculated by dividing the percentage of green cells (UR quadrant) by the sum of the percentage of transfected cells (UR+UL) and are indicated underneath the fluorescence distribution plots (range, n=2 of two independent experiments). Data adapted from (Marta Rodriguez-Alvarez et al., 2020).

Spontaneous mutations arising at a non-damaged template occur very rarely during DNA synthesis (Schmitt et al., 2009), therefore, cells transfected with the control construct (c.613T) showed a negligible fraction of EGFP fluorescent cells (0.31-0.35%) (**Figure 3-39**). In contrast, cells transfected with constructs carrying THF abasic sites (X) displayed 13% higher fluorescence as compared to the control (dT) construct. This indicates that TLS opposite THF was accompanied by a high rate of nucleotide misincorporation. Interestingly, sulfurization of the 5' bond to the THF (S-THF) enhanced the mutation frequency observed in THF by approximately 10%. Even though the APE1 digestion assay of THF placed in ss-DNA revealed low activity of this enzyme towards the lesion, this result suggests that 5' cleavage at synthetic AP lesions partially prevents mutagenic TLS in HeLa cells. Since dA incorporation would restore the stop codon responsible for the absence of fluorescence, regain of EGFP signal in THF and S-THF indicated that incorporation of T, G, and/or C occurs opposite THF during TLS. It is important to note that samples transfected with constructs that led to a regain of EGFP fluorescence, showed two populations of cells: one displayed only red fluorescence (UL quadrant) whereas another showed both red and green fluorescence (UR quadrant). This variety of fluorescent intensities might be explained by the high heterogeneity in the expression of DNA repair genes reported in HeLa cells, thus affecting the damage tolerance capacities of the cells.

Since results in section 3.2.3 have demonstrated that NER pathway works as a backup mechanism for THF repair in human cells, we investigated whether an impairment in this pathway would lead to an increase of mutagenic TLS over THF abasic sites. Therefore, we transfected XP-A (NER-deficient), XP-A/+XPA (XPA corrected) and MRC5 cell lines with the constructs suited for TLS detection.

Results

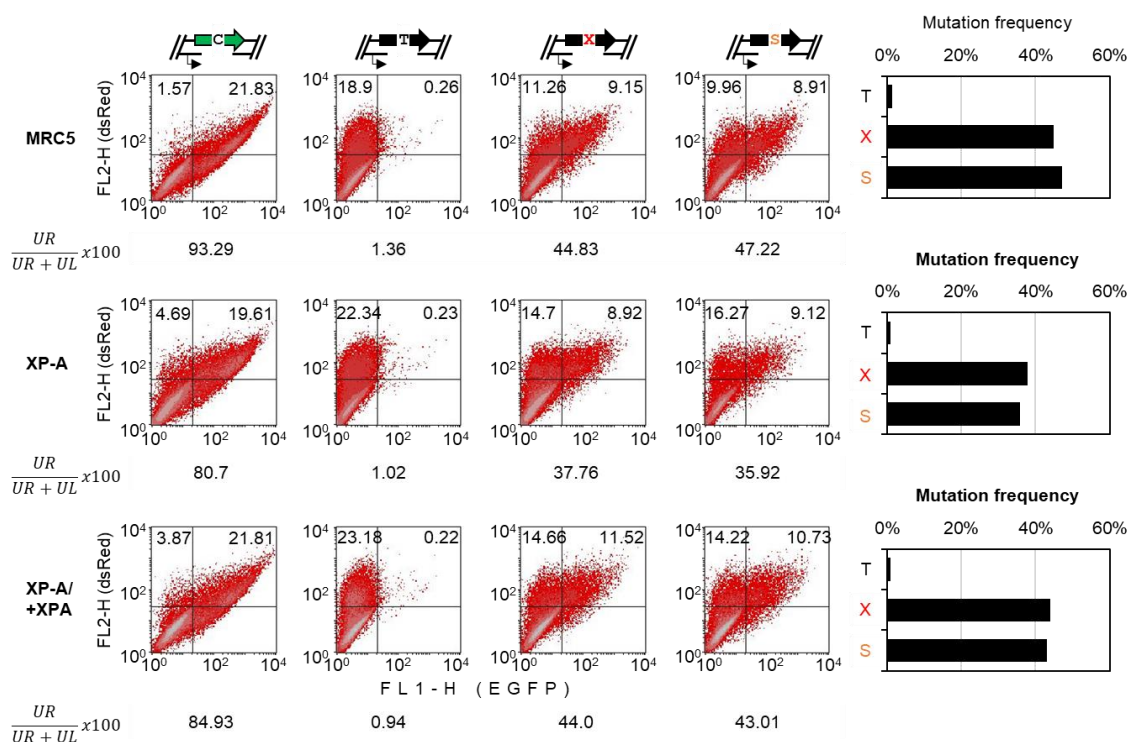


Figure 3-40: Detection of mutagenic TLS on the DNA template containing THF AP lesion in MRC5, XP-A, and XP-A/+XPA cell lines. Cells expressing the reverted fluorescent EGFP were analyzed by FACS 24 hours post-transfection in the absence and presence of a single THF (X) and S-THF (S) placed in the position c.613 opposite to an 18-nt gap in three different cell lines. The repair rate (for the wild-type EGFP) and the rate of reversal to fluorescent EGFP (for c.613T, THF, and S-THF) were calculated as previously described and are indicated underneath the fluorescence distribution plots and in the bar graphs ($n=1$).

By looking at the cells transfected with gapped EGFP c.613C construct, we estimated a repair rate of the gap between 80-93% in all three cell lines (**Figure 3-40**). It was interesting to note that the population of transfected cells appeared to be more homogeneous in these cell lines than in the previously analyzed HeLa cells. XP-A cells transfected with THF constructs did not show higher mutation frequency (37.36%) in comparison to MRC5 (44.8%) as it would be expected if a deficiency in NER would increase the half-life of the lesion increasing the signal of the mutagenic bypass. The further complementation of these cells with XPA protein showed an increase of fluorescence of 6% compared to XP-A cells. However, the interpretation of these small differences between cell lines would have been more precise if we would have increased the statistical power by repeating the experiment several times. Still, these results suggested that the NER pathway did not play an essential role in THF repair in ss-DNA. Surprisingly, unlike HeLa cells, none of the cell lines showed a significant increase of green fluorescence when transfected with S-THF constructs compared to THF. Thus, indicating that THF lesion placed in ss-DNA is a poor substrate for BER proteins in MRC5 and XP-A cells as confirmed by the *in vitro* digestion assay in **Figure 3-37**. It is also possible that, unlike MRC5 and XP-A, HeLa cells contain a nuclease activity that

degrades ss-DNA and the phosphorothioate bond protects the template from degradation. As a consequence, we pursued the investigation of TLS over abasic sites using only THF constructs lacking phosphorothioate linkage in the 5'.

3.3.3 Translesion synthesis over natural abasic lesion results in high mutation frequency

Because THF is a broadly used analogue of abasic sites but does not precisely mimic the common form of abasic sites produced under physiological conditions, we decided to extend our approach to study mutagenic TLS over natural AP sites. Natural AP sites possess less stability than their THF analog due to their reactive hydroxyl group in the C1 of the sugar backbone (**Figure 3-13**). Instead of using constructs directly carrying a natural AP site, we used constructs carrying uracil to avoid potential AP site reactivity during the transfection process. In addition, the skeptical character of the APE1 excising the resulting AP site in ss-DNA, would increase the half-life of the lesion favoring TLS. Thereby, we relied on the efficient activity of uracil DNA glycosylases (UDGs) present within the cells to cleave uracil in ss-DNA and create a natural AP site that will not be effectively excised by APE1.

Starting from the pZAJ_Q205* mutant, we inserted synthetic oligonucleotides containing uracil into the NTS in position c.613 by strand depletion [2.2.7-2.2.10]. Endo IV-UDG digestion performed in ds-DNA prior to the generation of the gap in the opposite strand served for the characterisation of the new constructs. After UDG incubation, the cleavage of the N-glycosyl bond formed a natural AP site that was further digested by Endo IV 5' to the lesion. Therefore, detection of 100% incision of the plasmids harboring uracil only upon incubation with both enzymes confirmed the proper insertion of the synthetic oligonucleotides in all c.613U constructs (**Figure 3-41A**). Since the control construct (c.613T) only showed 9% of unspecific Endo IV cleavage that did not vary in presence or absence of UDG, we attribute this 9% to the presence of AP sites preexisting in the plasmid from the amplification procedure in bacteria. Following the insertion of the synthetic oligonucleotides, it was necessary to generate a gap on the TS opposite to the lesion by strand depletion protocol after incubation with Nb.Bpu10I. To confirm the presence of the gap, we performed an analytical ligation. Only upon incubation of the gapped constructs with ligase, synthetic oligonucleotides, and PNK the covalently closed form of the plasmid was shown. The lack of a covalently closed form of the plasmid in presence of T4 DNA ligase confirmed the absence of the original DNA stretch previously located between the nicking sites and therefore, the presence of the gap within the template strand of the constructs (**Figure 3-41B**).

Results

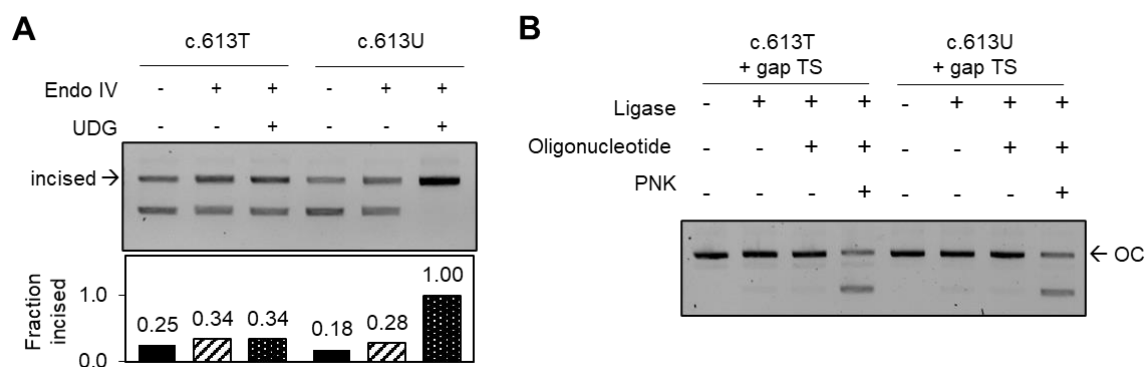


Figure 3-41: Preparation of constructs with uracil in position c.613 of the pZAJ_Q205* reporter opposite to an 18-nt gap (A) Endo IV-UDG digestion (2U/100ng) in double-stranded DNA after ligation of synthetic oligonucleotides containing T or U in position c.613. The lower graph shows quantification of the fraction of the open circular (incised) form of the plasmids. **(B)** A representative experiment of an analytical ligation of gapped vector in the NTS.

To monitor the TLS over natural or physiological AP sites, we transfected uracil together with THF and dT constructs into HeLa and UNGsh cell lines. As previously seen in **Figure 3-39**, TLS over THF AP sites led to a 14% increase of mutation frequency in HeLa pEps compared to the control dT construct (**Figure 3-42**). Similarly, higher mutation frequency (16%) was detected in UNGsh cells transfected with THF lesion compared with the background levels of dT (0.58%). The observed mutation frequency was even higher in cells transfected with uracil constructs reaching 19.9% of transfected cells showing green fluorescence in both cell lines. However, no significant differences were observed in the mutation frequency when transfecting the uracil construct between HeLa and UNGsh cells. Since not only UNG but also SMUG1, has been reported to effectively excise uracil from ss-DNA (Kavli et al., 2002), the depletion of UNG protein alone did not significantly affect the excision of uracil and subsequent formation of natural AP sites. Thus, we conclude that mutagenic TLS occurs opposite synthetic (THF) as well as natural AP sites (U), leading to the synthesis of a mutated template that encodes for a fluorescent EGFP in human cells.

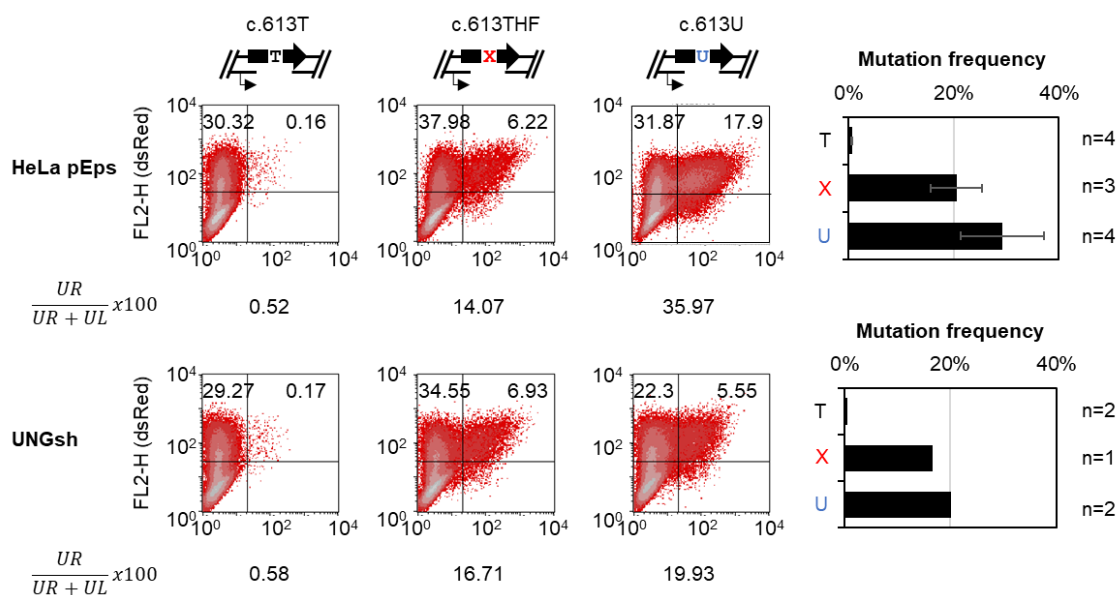


Figure 3-42: Detection of mutagenic TLS on DNA templates containing THF and U in HeLa and UNGsh cell-lines by FACS analysis. Scatter plots from a representative experiment of cells 24 hours post transfection with constructs harboring a single U and THF in position c.613 opposite to an 18-nt gap. The rate of reversal to fluorescent EGFP showed underneath was calculated by dividing the percentage of green cells (UR quadrant) by the sum of the percentage of transfected cells (UR+UL).

Since we cannot exclude that the mutagenicity detected in cells transfected with uracil arises from the sum of the erroneous bypass of uracil and abasic site, we used a 2'-(R)-fluorinated derivative uracil (FU) which is inherently resistant to glycosylase cleavage (Schröder et al., 2016) (**Figure 3-43A**). To test the resistance of FU towards glycosylase activity, we performed an Endo IV-UDG digestion assay using ds-DNA constructs. Here, we observed 6-9% of unspecific cleavage by Endo IV in the control construct (c.613T) as well as in c.613U and FU constructs corresponding to the cleavage of preexisting AP sites within the vectors. Incubation of c.613THF with Endo IV alone sufficed to acquire 100% of the open circular form of the plasmid regardless of the presence of UDG (**Figure 3-43B**). Because UDG releases the uracil and creates an AP site that acts as a good substrate for endonuclease activity, the combination of UDG and Endo IV showed almost complete digestion (89%) of the uracil construct (c.613U). As expected, the combination of both enzymes did not result in a significant increase of the open circular form of the plasmid in the c.613FU construct. This result confirmed that the 2' fluorine atom protects the uracil base from UDG glycosylase activity. To complete the preparation of the TLS constructs, it was necessary to nick and gap the template strand opposite to the modification. The absence of cc form of the plasmid in presence of T4 DNA ligase illustrated the success of the gapping procedure (**Figure 3-44A**).

Results

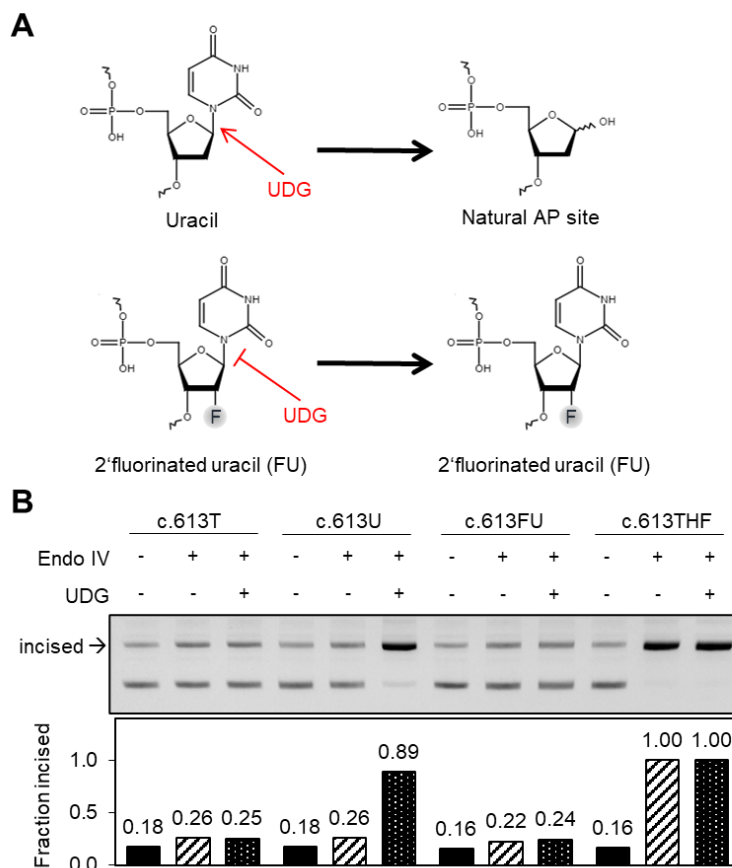


Figure 3-43: Preparation of constructs with U and FU opposite to dA in double-stranded DNA. (A) Chemical structure of U, 2'-FU and natural AP site modifications used in this experiment **(B)** Endo IV-UDG digestion (2.5U/100ng) in double-stranded DNA after ligation of synthetic oligonucleotides containing T, U or FU in position c.613. Quantification of the % open circular (oc) form of the plasmids appears below the gel image.

Analysis of the EGFP fluorescence in HeLa cells transfected with uracyl construct (c.613U) showed an enormous regain of EGFP fluorescence (37%) compared to the template containing thymine in the same position (0.36%) (**Figure 3-44B**). Interestingly, this regain of fluorescence was reverted almost completely (1.3%) when cells were transfected with the glycosylase resistant construct c.613FU. This result indicated that only a negligible percentage of the mutation frequency observed appeared after TLS over uracyl residues. Thus, nearly all the fluorescence observed in cells transfected with uracyl template arose from mutagenic bypass of the abasic sites derived from uracyl excision.

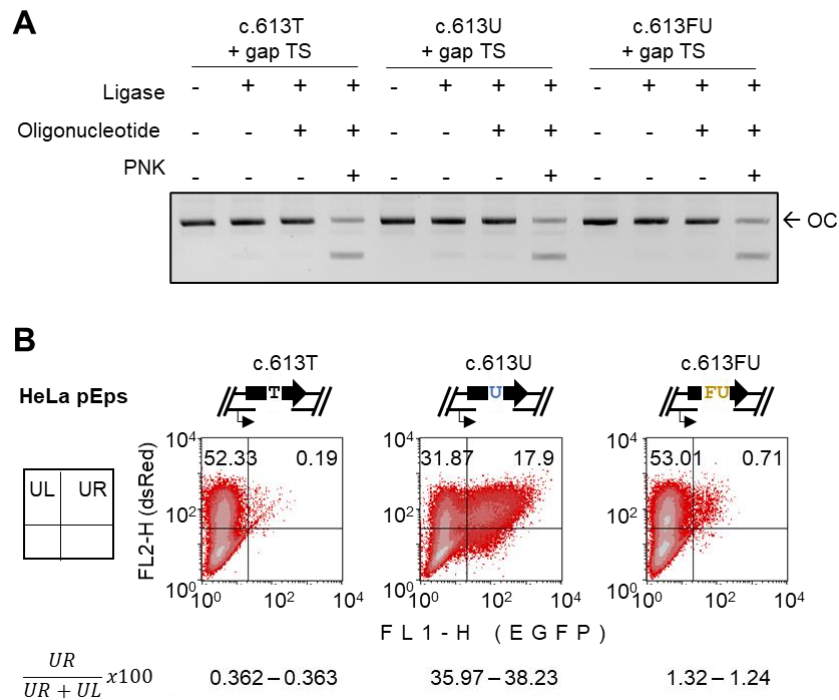


Figure 3-44: Detection of mutagenic TLS on the DNA template containing U and FU lesions in the HeLa cell line. (A) Analytical ligation of constructs gapped in the NTS. **(B)** Cells expressing the reverted fluorescent EGFP were analyzed by FACS 24 hours post-transfection with 3 different constructs: T (control), U (uracil), and FU (2-fluorinated uracil) in the position c.613 opposite to an 18-nt gap. Scatter plots of a representative experiment are shown. The percentage of reversal to fluorescent EGFP is indicated underneath the fluorescence distribution plots (range, n=2)

3.3.4 Although highly mutagenic, synthetic and natural abasic sites are bypassed in different manners during TLS

Regain of EGFP signal in cells transfected with c.613THF and c.613U templates indicated that not only adenine but also guanine, cytosine and/or thymine were incorporated opposite abasic lesions during TLS. To study whether the sequence context influences the mutation pattern followed by the TLS, it was interesting to analyze mutation rates of the same type of lesion in different mutant reporters. Because for TLS constructs generation modification of both DNA strands is necessary, any mutant EGFP plasmid would be suitable as a starting backbone. Thus, we used the previously gapped pZAJ_Q205* vector to insert five different oligonucleotides in the NTS of the EGFP coding sequence. Each oligonucleotide accommodates a THF that substitutes the mutated nucleotide in one of the mutant reporters generated in section 3.1. A summary of the potential phenotypical outcomes derived from TLS events that occurred at the replaced nucleotide is shown in **Figure 3-45**.

Results

pZAJ_Q205*	pMR_Q205P	pMR_S206Y	pMR_S206*	pMR_A207P	Mutant vector
5'—CTA— └─▶	5'—CCG— └─▶	5'—TAC— └─▶	5'—TGA— └─▶	5'—CCC— └─▶	
c.613THF	c.614THF	c.617THF	c.617*THF	c.619THF	Construct
5'—CXA— └─▶	5'—CXG— └─▶	5'—TXC— └─▶	5'—TXA— └─▶	5'—CXC— └─▶	
A C G T	G A C T	T A C G	G A C T	G A C T	Nucleotide opposite to the lesion
— ✓ ✓ ✓	— ✓ ✓ ✓	— ✓ ✓ ✓	✓ — —	— ✓ ✓ ✓	Florescence

Figure 3-45: Overview scheme of all mutant vectors used for translesion synthesis (TLS) assays. X shows positions of THF abasic DNA lesions and the expected fluorescence phenotypes after gap repair opposite to the lesion are indicated at the lower row called fluorescence.

After insertion of the five synthetic oligonucleotides each carrying a tetrahydrofuran abasic site (THF) in a different position of the non-template strand, it was important to confirm the presence of AP sites in all constructs via APE1 digestion. Incubation of THF constructs with 1U APE1 showed a total conversion of the plasmids (100%) containing THF into an open circular (incised) form (**Figure 3-46A**). The control construct dT only depicted 8% of cleavage in presence of APE1 that illustrated the basal percentage of plasmids containing pre-existing AP sites. An analytical ligation validated the subsequent nicking and gapping of the opposite strand (TS). The absence of cc form of the plasmid in presence of ligase indicated that all constructs lack the 18-mers stretch of nucleotides placed between the Nb.Bpu10I nicking sites. Hence, the six constructs (five containing THF plus the dT control) have been properly gapped in the template strand opposite to the modified bases (**Figure 3-46B**).

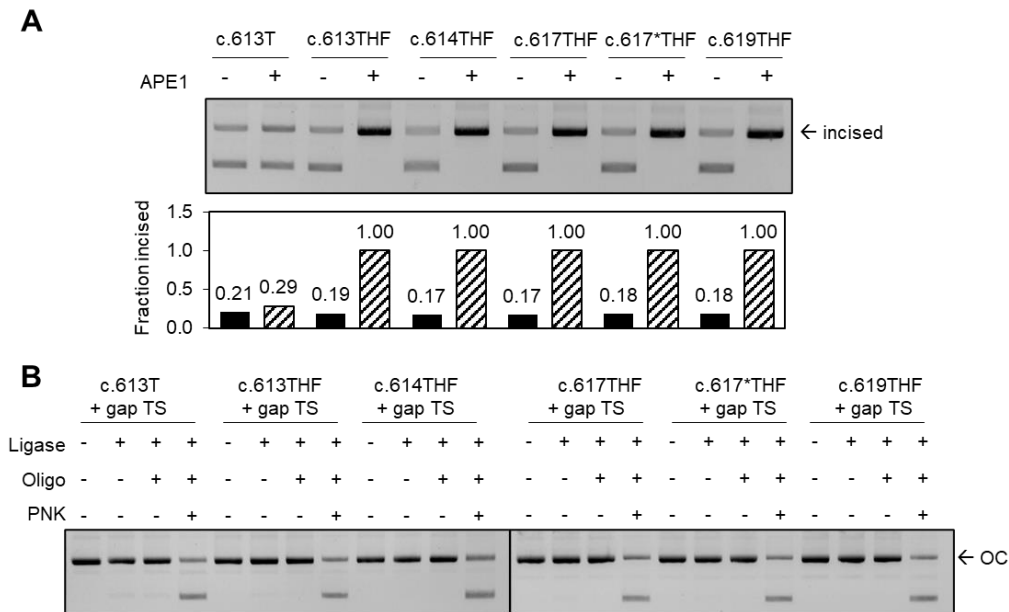


Figure 3-46: Generation of constructs with site-specific THF AP sites in the coding strand of all mutant vectors. (A) APE1 digestion of ds-DNA constructs harboring a THF AP lesion in five different positions of the coding strand. (B) Analytical ligation to validate the presence of the gap in the TS created opposite to the lesion.

Rates of reversion to a fluorescent EGFP were analyzed in HeLa cells transfected with constructs harboring THF lesion in five different positions opposite to a gap. Repair of the gap via DNA synthesis opposite to THF led to the reactivation of the EGFP fluorescence in all different positions and none of the constructs showed a complete lack of fluorescence (**Figure 3-47**). Based on this result, incorporation of all four nucleotides (A, C, T, or G) was theoretically possible opposite to synthetic abasic lesions. It is interesting to note that THF in positions c.614 (C>THF) and c.619 (C>THF) produced similar levels of green fluorescence (50.6% and 54.2% respectively) as well as the highest mutant frequencies compared to other positions. In both constructs, incorporation of A, C and/or T opposite to the lesion leads to the reactivation of the protein fluorescence. Thus, these results suggest that erroneous bypass of an AP site arising from cytosine depyrimidination leads to the highest mutation rates.

Results

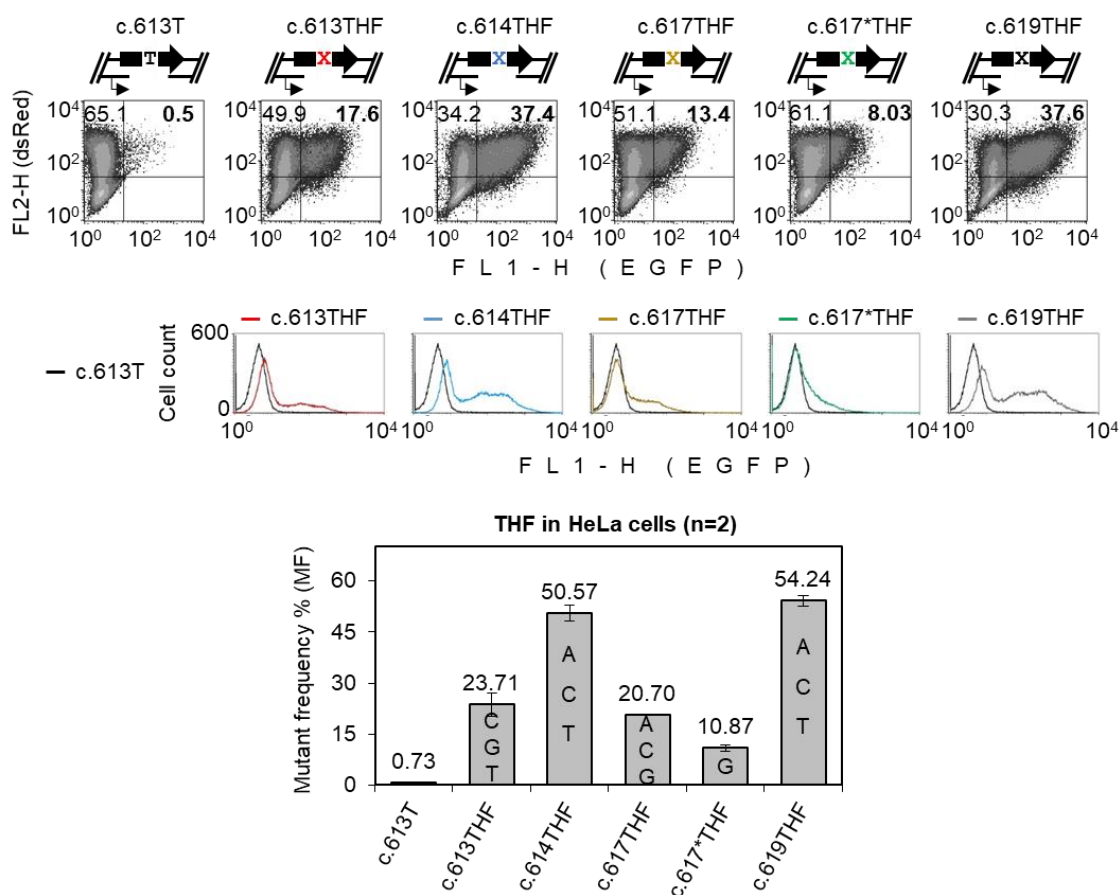


Figure 3-47: Detection of mutagenic TLS on the DNA template containing THF in 5 different positions. Cells expressing the reverted fluorescent EGFP were analyzed by FACS 24 hours post-transfection with 5 different constructs carrying X (THF) in positions c.613, 614, 617, 617, and 619 opposite to 18-nt gap. Scatter plots of a representative experiment are shown with correspondent overlays. The mutant frequency was calculated by dividing the percentage of green cells (UR quadrant) by the sum of the percentage of transfected cells (UR+UL) and are indicated in the table (mean \pm SD, n=2), letters within the bars indicate the nucleotide that would reactivate the fluorescence in each position.

Although very informative, the calculation of the mutant frequency only provides an estimation of the mutagenicity of the lesion. Therefore, we aimed to find a cell line that shows only one population of cells after transfection with EGFP plasmids and with a more homogeneous background than HeLa. We decided to switch to the HAP1 cell line that is commonly used for genetic research as they are rapid in doubling time. Additionally, HAP1 cells only carry a single copy of each gene while HeLa cells are well-known for having abnormal chromosome numbers that affect several DNA repair genes (Landry et al., 2013; Macville et al., 1999). The haploid condition of HAP cells also facilitates the creation of KO cell lines since the disruption of one copy of the gene is sufficient to generate a phenotypical change. Indeed, we detected a homogeneous population when HAP cells were transfected with abasic site constructs for TLS assay (**Figure 3-48**). This allowed the study of the mutagenic TLS over abasic sites in a quantitative manner by extracting the median EGFP fluorescence of a unimodal distribution.

Expression analysis of the constructs containing THF revealed an overall regain of fluorescence in HAP cells (**Figure 3-48**). This fluorescence was especially intense in constructs accommodating THF in positions c.614 and c.619 as previously seen in HeLa cells. These two constructs harbor a THF lesion that substitutes a cytosine residue thus confirming that TLS over an AP site generated by cytosine depyrimidination is extremely mutagenic. Interestingly, the median EGFP fluorescence differed significantly between THF placed in positions 614 and 619. Because adenine, cytosine and thymine reactivate the EGFP fluorescence in both positions, this result also suggested that the incorporation of a nucleotide opposite to the lesion might be influenced by its sequence context. Indeed, c.613THF, c.614THF, and c.619THF constructs depicted always higher mutagenicity regardless of the cell line used. This might be related to the fact that all these constructs contain cytosine residue 5' to the THF, while in constructs generated from S206Y and S206* mutants, thymine is placed 5' to the THF residue (**Figure 3-45**). However, each mutant has a particular pattern of mutations that would reactivate the protein fluorescence and not all amino acid changes lead to the same fluorescence intensity (**Table 3-2**). As a consequence, these results only allow speculation about the influence of the sequence context in the mutation profile of THF AP sites. The double mutant c.617*THF (derived from S206*) only leads to the reactivation of the EGFP signal if guanine gets incorporated opposite to the lesion. Showing an EGFP median of only 17,86, this result suggested that incorporation of guanine opposite THF rarely occurred. THF placed in position c.617 showed a median fluorescence of 64,3 where thymine incorporation leads to a non-fluorescent EGFP; however, THF placed in position c.613 where thymine incorporation reactivates the protein fluorescence showed a much higher median (115,48). Taken together, these results suggested that the incorporation of thymine opposite THF might also occur during TLS. To confirm previous theories about the mutation profile of synthetic AP sites during TLS, analysis of the fluorescent pattern in every mutant is not sufficient. That is the reason why section 3.4 describes an optimized protocol for RNA sequencing of the transcripts generated after gap-filling process accompanied by mutagenic TLS in these cells.

Results

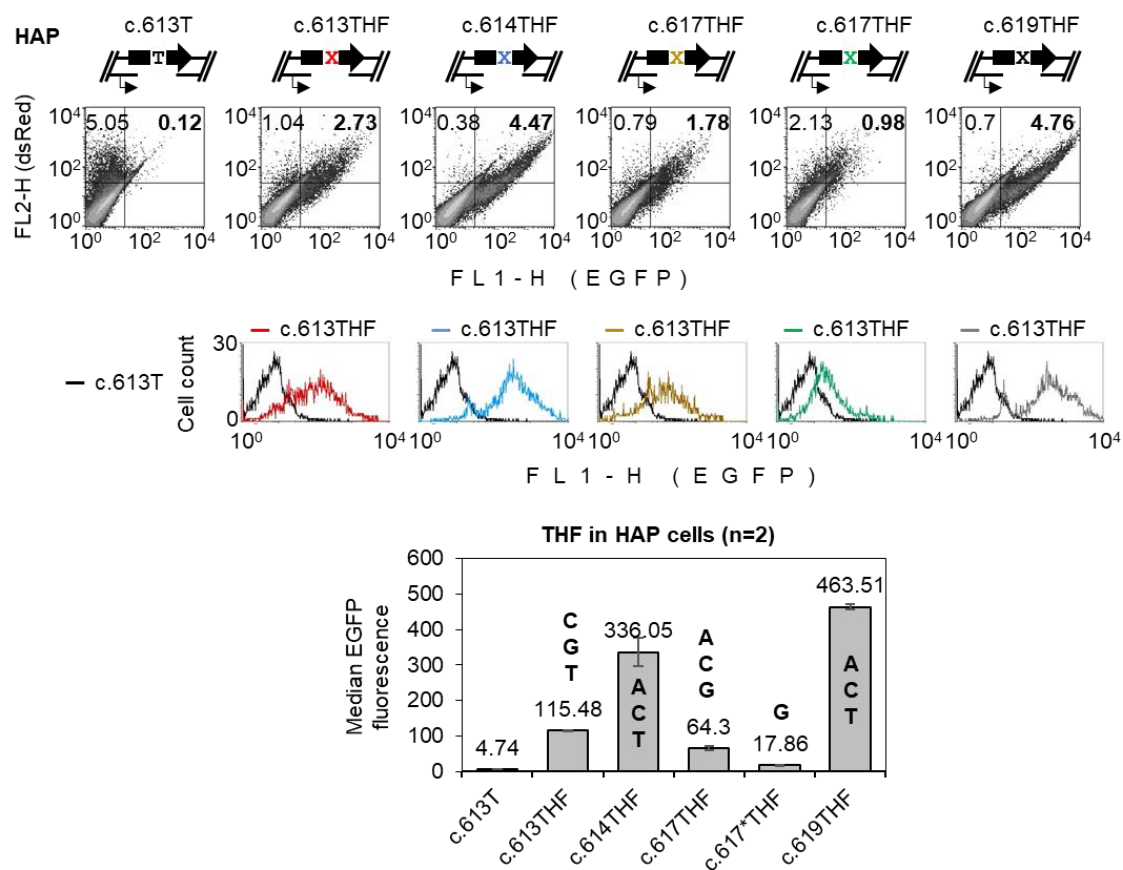


Figure 3-48: Detection of mutagenic TLS on the DNA template containing THF in 5 different positions in HAP cells. Cells expressing the reverted fluorescent EGFP were analyzed by FACS 24 hours post-transfection with 5 different constructs carrying X (THF) in the position c.613, 614, 617, 617, and 619 opposite to an 18-nt gap. Scatter plots of a representative experiment are shown with correspondent overlays. The graph represents the median EGFP fluorescence of two independent experiments (mean \pm SD, n=2). Letters within or above the bars indicate the nucleotide that would reactivate the fluorescence in each position.

To study whether the mutagenic pattern followed by synthetic abasic sites was similar to the one depicted by physiological AP sites, we generated constructs carrying uracil (U) at different positions of the EGFP coding sequence. Ligation of synthetic oligonucleotides carrying uracil and the dT control into the gapped NTS pZAJ_Q205* plasmid worked with an efficiency higher than 92% (**Figure 3-49A**). To detect the presence of uracil in plasmid DNA, we incubated our still double-stranded constructs with a combination of Endo IV and UDG enzymes. Incubation with Endo IV alone indicated unspecific cleavage of <9% in all constructs compared to their control sample (-/-) (**Figure 3-49B**). Uracil in positions c.613, 617, and 619 showed >93% cleavage by Endo IV after UDG generated an abasic site hydrolyzing the N-glycosyl bond placed between the nitrogen base and the sugar. Surprisingly, Endo IV cleavage was not that efficient in constructs containing uracil in positions c.614U (58%) and c.617*U (68%).

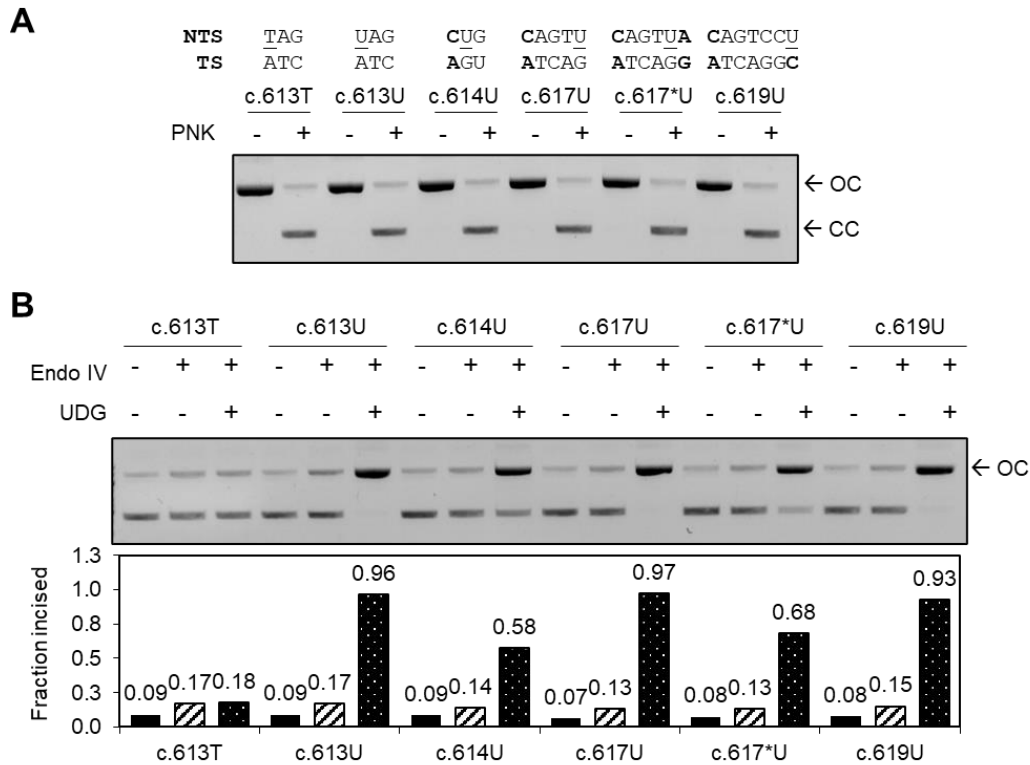


Figure 3-49: Generation of constructs with site-specific uracil sites in the coding strand of all mutant vectors. (A) Preparative ligation of synthetic oligonucleotides insertion into NTS gapped pZAJ_Q205* vector. Above the labeling the resulting sequence of the construct is indicated; bold letters indicate mismatches and the uracil is underlined. **(B)** Endo IV-UDG digestion of double-stranded constructs harboring uracil in the indicated positions of the coding strand in ds-DNA.

Since we generated the constructs starting from a pZAJ_Q205* (c.613T) plasmid gapped in the NTS, the insertion of the synthetic oligonucleotide with uracil in position 614 additionally creates a mismatch located 3' to the uracil in position 613 (C:A, NTS:TS) as shown in the sequences placed above the gel image in **Figure 3-49A**. Likewise, insertion of the synthetic oligonucleotide with uracil in position 617 for the double mutant pMR_S206* creates not only the 613C:A mismatch but also a second mismatch in position 618A:G placed 5' to the uracil. As it has been previously shown that mismatches placed immediately 5' or 3' to an AP site reduce the rate of APE1 cleavage *in vitro* (Wilson III David M et al., 1995), we deduced that an additional mismatch close to the uracil affected the excision by the Endo IV in these constructs as well.

FACS analysis of HAP cells transfected with templates carrying uracil showed an overall re-gain of the EGFP signal (**Figure 3-50**). Thus, synthesis of the gap through TLS over natural abasic sites was accompanied by nucleotide misincorporations that after being transcribed, reactivated the protein fluorescence. In this experiment, the addition of the c.613C construct encoding for a functional EGFP harboring an 18-nucleotide gap in the TS allowed the normalization of the median EGFP fluorescence. Uracil in positions c.614 and c.619, where incorporation of dA, dC, or dT opposite to the lesion completely

Results

restores the protein fluorescence, depicted the highest EGFP fluorescence. This result indicates once again that TLS over an AP site generated by cytosine depyrimidination is extremely mutagenic. Surprisingly, the EGFP fluorescence shown by c.619U constructs was much higher than c.614 and exceeded the standard fluorescence detected by the original EGFP gapped construct (c.613C + gap). This might be explained by the enhanced EGFP fluorescence showed by this reporter when thymine or adenine are incorporated in position ts.619 (**Figure 3-9**). Unlike results obtained with synthetic AP sites where fluorescence of the c.617THF construct was always higher than c.617*THF (**Figure 3-48 and Figure 3-48**), natural AP sites in both positions showed similar EGFP fluorescence (**Figure 3-50**). Because c.617*X constructs only show reactivation of the EGFP fluorescence if guanine gets incorporated opposite to the lesion, these results suggest higher incorporation of guanine in templates containing uracil than in templates containing THF.

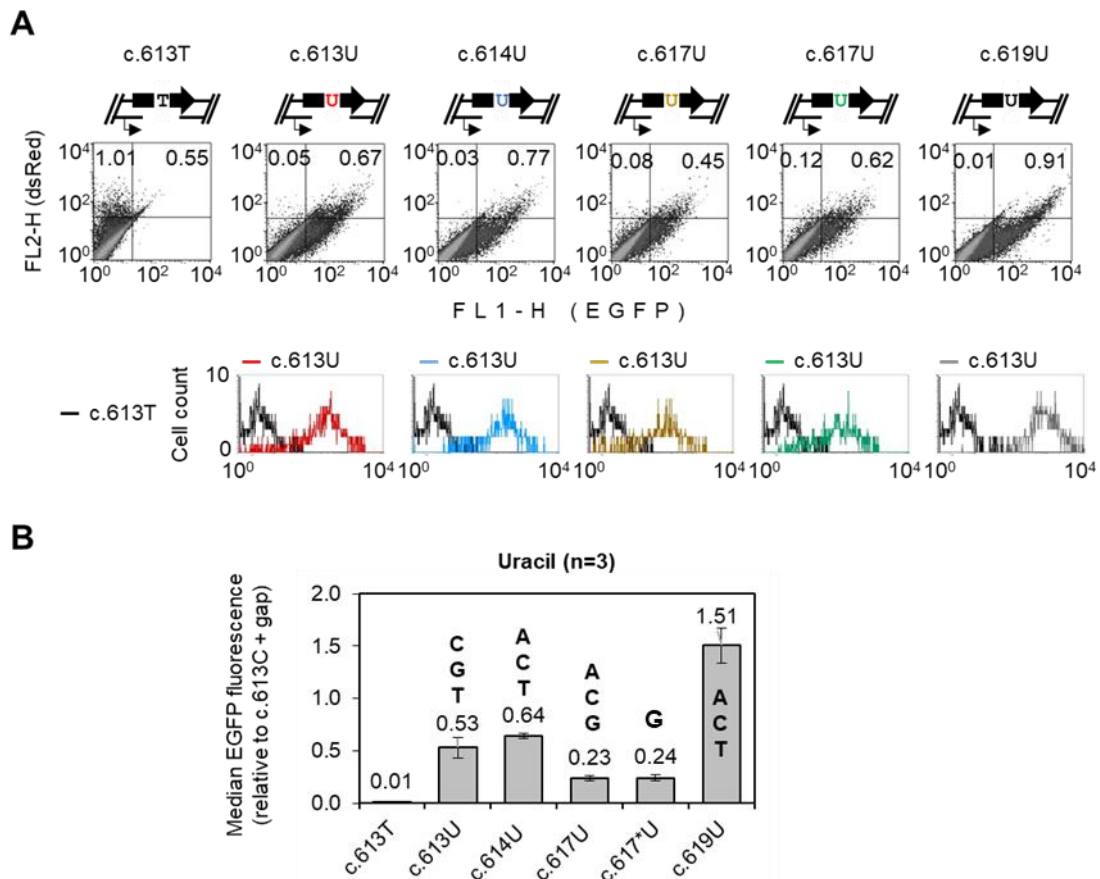


Figure 3-50: Detection of mutagenic TLS on the DNA template containing U in 5 different positions in HAP cells. Cells expressing the reverted fluorescent EGFP were analyzed by FACS 24 hours post-transfection with 5 different constructs carrying uracil in positions c.613, 614, 617, 617 and 619 opposite to an 18-nt gap. Scatter plots of a representative experiment are shown with correspondent overlays. The graph represents the mean EGFP fluorescence of two independent experiments relative to the gapped c.613C construct (mean \pm SD, n=2). Letters within or above the bars indicate the nucleotide that would reactivate the fluorescence in each position

By using the median EGFP fluorescence obtained in HAP cells together with a c.613C + gap construct that allowed the normalization of the obtained data, we aimed to directly compare the mutagenic outcome of synthetic (THF) versus physiological (U) abasic sites. After transfection of the wild-type EGFP construct (c.613C), we observed 98% of transfected cells that efficiently synthesized the missing strand and moved to the upper right quadrant of the scatter plots showing green fluorescence (**Figure 3-51**). As expected, synthesis of the missing strand in c.613T constructs showed a negligible percentage of fluorescent cells (0.7%) as the DNA synthesis through the thymine residue would restore the TA pair that encodes for the stop codon that characterizes the Q205* mutant. Templates carrying both uracil and tetrahydrofuran in position c.613 showed a high percentage of green fluorescent cells confirming that TLS over these two abasic lesions occurred with a high rate of nucleotide misincorporation (**Figure 3-51**). In addition, the lack of fluorescent phenotype in cells transfected with FU template confirmed that uracil bypass was not responsible for the reactivation of the fluorescence in U constructs; instead, the effect arose due to the bypass of natural AP sites created after uracil excision. Interestingly, our results showed that although both types of AP sites undergo mutagenic TLS, natural AP sites displayed higher mutation frequency (89.7%) than its synthetic analogue THF (78.4%). Also confirmed by the median EGFP fluorescence normalized to c.613C where natural AP sites (U) showed a 2.5-fold increase in EGFP fluorescence compared to THF. Taken together these results indicated that both synthetic (THF) and natural (U) are erroneously bypassed by TLS machinery and that natural AP sites are far more mutagenic than their broadly used synthetic analogue tetrahydrofuran (THF) in this context.

Results

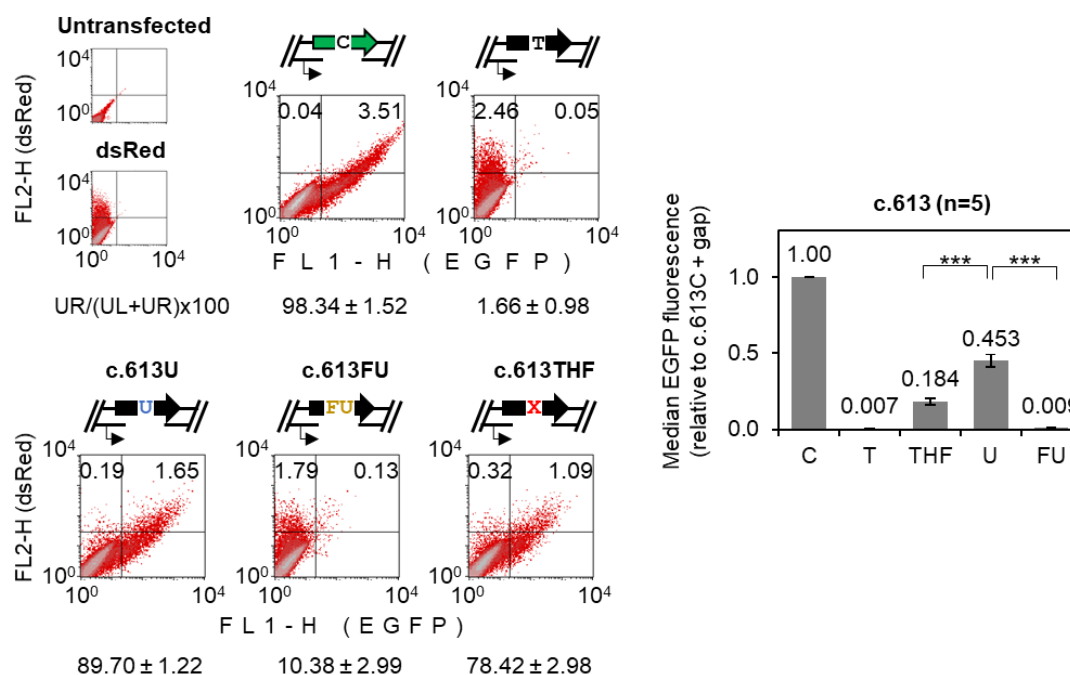


Figure 3-51: Detection of mutagenic TLS on the DNA template containing synthetic and physiological AP lesions in HAP cell line. FACS analysis 24 hours post-transfection using C, T, U, FU, THF constructs in position c.613 opposite to an 18-nt gap. **Left:** Scatter plots of a representative experiment with the corresponding mutation frequency showed below the plots (mean \pm SD). **Right:** Median EGFP values relative to wild-type EGFP construct (c.613C + gap) of 5 independent experiments (mean \pm SD, n=5), *** p-value < $2.47 \cdot 10^{-5}$.

3.3.5 Mutagenic TLS over synthetic and natural abasic lesions shows higher adenine incorporation in absence of Rad18

Abasic sites are known to be mutagenic and as they do not generally disturb the B-conformation of the DNA helix, also considered non-bulky lesions. Thereby, it has been reported that both replicative and TLS polymerases are capable of their bypass during DNA synthesis. To study whether inactivation of TLS polymerases recruitment would eliminate the mutagenic TLS observed as regain of EGFP signal, we used Rad18 KO cells. RAD18 gene encodes for the E3-ubiquitin-ligase responsible for the mono-ubiquitylation of PCNA during replication, which activates the error-prone translesion synthesis pathway by TLS polymerases recruitment (**Figure 1-3**) (Parker & Ulrich, 2009; Watanabe et al., 2004). We, therefore, hoped that in the absence of Rad18, only replicative polymerases would perform the AP site bypass via its well-defined pattern of adenine incorporation opposite to the lesion (McCulloch & Kunkel, 2008; Obeid et al., 2010). For this reason, we used the pZAJ_Q205* mutant where incorporation of dA opposite AP site in position 613 will maintain the non-fluorescent EGFP status of the construct and only mutagenic bypass by TLS polymerases will be detected.

Rad 18 KO cell line had a longer doubling time than its parental HAP1 cell line (**Table 2-7**). After transfection experiments, the viability of these cells decreased and resulted in

very low transfection efficiency. Therefore, we tested whether the absence or presence of serum during the transfection process could play a role in the outcome of the experiment. Indeed, FBS-free transfection was better suited, not only for Rad18 KO cells but also for all HAP-derived KO cells used in further experiments (**Supplementary figure 1**).

FACS analysis of Rad 18 KO cells transfected with TLS constructs harboring synthetic and natural AP sites showed a significant decrease of 11% in EGFP fluorescence compared to their parental cell-line when transfected with THF constructs (**Figure 3-52**). Excitingly, the decrease in EGFP fluorescence was more pronounced (up to 23%) when transfecting constructs carrying U at the same position (c.613U). This result served as an indication that either the bypass of the lesion was generally hindered or the rate of incorporation of potentially mutagenic nucleotides (dG, dT or dC) opposite to AP sites decreased in absence of Rad18. It is important to note, that even in Rad18 KO we still detected 33.5% of EGFP fluorescence, which is possibly appointed to the retention of some mutagenic TLS activity independent of Rad18 that has been previously reported (Hendel et al., 2011; Simpson et al., 2006).

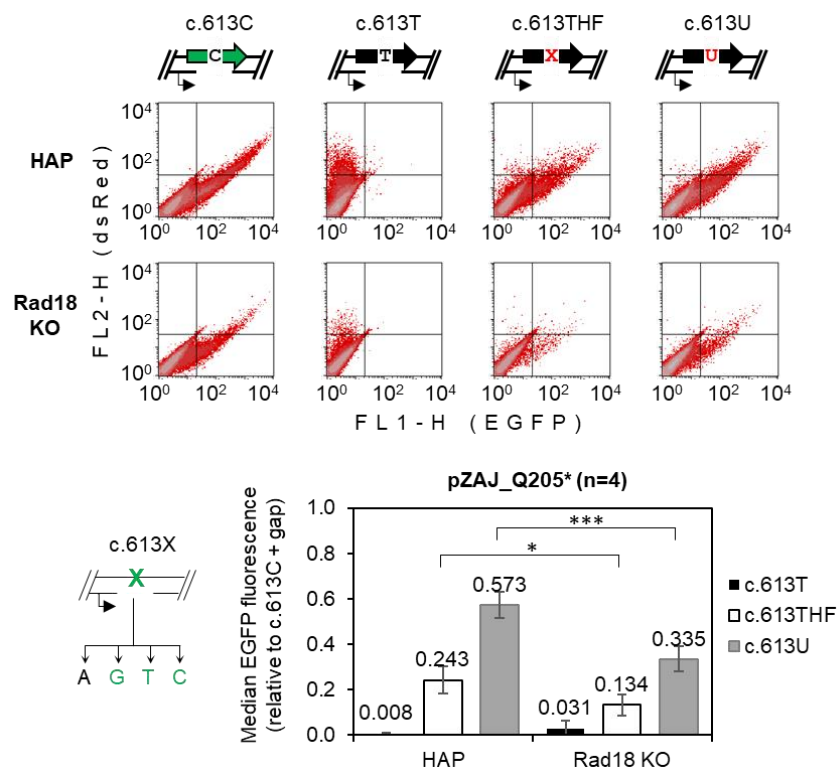


Figure 3-52: Detection of mutagenic TLS over synthetic and natural AP sites using pZAJ_Q205* mutant in HAP and Rad18KO cells. Cells expressing the reverted fluorescent EGFP were analyzed by FACS 24 hours post-transfection with 3 different constructs: T (control), THF (X), and U (uracil) in position c.613 opposite to an 18-nucleotide gap. Scatter plots of a representative experiment are shown. Below: Scheme of fluorescent phenotypes as a consequence of TLS occurring at position 613 and quantification of the median EGFP fluorescence relative to c.613C opposite to an 18-nucleotide gap of four independent experiments (mean \pm SD), *** p-value: 0.00095/* p-value: 0.04926.

Results

To study whether the decrease of EGFP signal in c.613THF and U constructs in Rad18 KO cells was a result of a generally impaired TLS or due to an increase of dA incorporation opposite to the lesion, we used an alternative mutant. In pMR_A207P the incorporation of dA opposite to the lesion leads to the synthesis of a fluorescent EGFP (Figure 3-45). Interestingly, we did not observe any significant decrease of EGFP fluorescence in Rad18 KO cells transfected with AP sites templates compared to HAP (Figure 3-53). This result indicated that the overall DNA synthesis through abasic sites is not impaired in absence of Rad18. The maintenance of the fluorescent status of the c.619THF and c.619U constructs in the absence and presence of Rad18 suggested that the decrease of bypass by TLS polymerases observed in position 613 was only compensated by dA incorporation opposite to the lesion possibly by replicative polymerases. This would result in a lower median fluorescence in position 613 where incorporation of adenine opposite to the lesion would lead to the synthesis of a non-fluorescent protein while maintaining a steady fluorescence in position 619. Taken together, results in the two different reporters (c.613 and c.619) indicated that AP sites are efficiently bypassed by both replicative and TLS DNA polymerases with major incorporation of adenine opposite to the lesion.

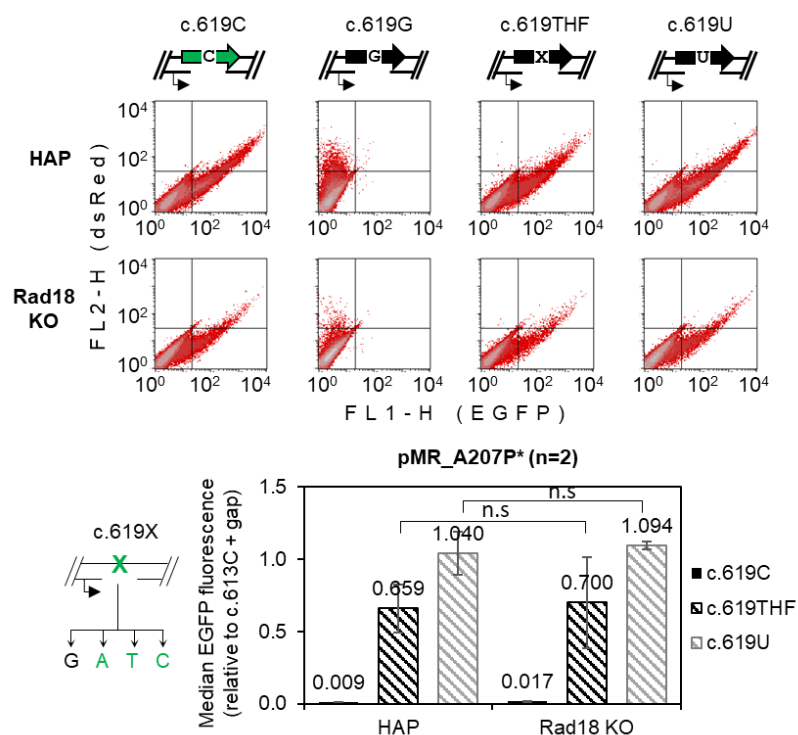


Figure 3-53: Detection of mutagenic TLS over synthetic and natural AP sites using pMR_A207P mutant in HAP and Rad18KO cells. Cells expressing the reverted fluorescent EGFP were analyzed by FACS 24 hours post-transfection with 3 different constructs: T (control), THF (X) and U (uracil) in position c.619 opposite to an 18-nucleotide gap. Scatter plots of a representative experiment are shown. Below: Scheme of fluorescent phenotypes as a consequence of TLS occurring at position 619 and quantification of the median EGFP fluorescence relative to c.619C opposite to an 18-nucleotide gap of two independent experiments (mean \pm SD, n.s.: non-significant).

3.3.6 Absence of Pol ι and Rev1 TLS polymerases reduces the mutagenic bypass of synthetic abasic sites in human cells

Y-family polymerases are specialised in the bypass of DNA lesions that block the normal progression of DNA synthesis. Pol ι , pol η , pol κ , and Rev1 belong to this family and each of them has unique features as well as different preferences regarding the nucleotide insertion opposite to a specific DNA lesion. Therefore, in addition to the mutation signature of AP sites during translesion DNA synthesis bypass, we investigated whether a specific TLS polymerase was accountable for the mutagenic bypass of THF AP sites during DNA synthesis. For that purpose, c.613THF constructs were transfected into four different Y-family polymerase knockout (KO) cell lines (Pol ι , pol η , pol κ , or Rev1) purchased from *Horizon Discovery*. All knockouts (including the previously used Rad18 KO) were created using the CRISPR-Cas method, followed by the isolation of the clones with frame-shift mutations and subsequent validation using PCR amplification and Sanger sequencing.

EGFP analysis showed a 24% EGFP fluorescence detected in HAP cells transfected with THF substrate, a percentage that did not suffer any significant alteration in absence of pol η and pol κ polymerases showing a median fluorescence of 22% and 21.6% respectively (**Figure 3-54**). Interestingly, Pol ι and Rev1 KO transfected cells showed a decrease of EGFP fluorescence by half compared to the parental HAP cell line. Thus, indicating that the absence of both polymerases decreases the mutagenicity of the THF bypass. It is important to note that the reduction in EGFP fluorescence in these two cell lines does not eliminate the fluorescent signal completely. This might be explained by the redundancy of TLS polymerases within the cells, where usually bypass of a DNA lesion is reported by more than one polymerase. Thus, the knockout of a single TLS polymerase might affect the overall bypass of the lesion but does not lead to its complete elimination. The results presented here indicate that mutagenic bypass of synthetic abasic sites is impaired in absence of Pol ι and Rev1 suggesting their potential role in THF AP sites derived mutagenicity.

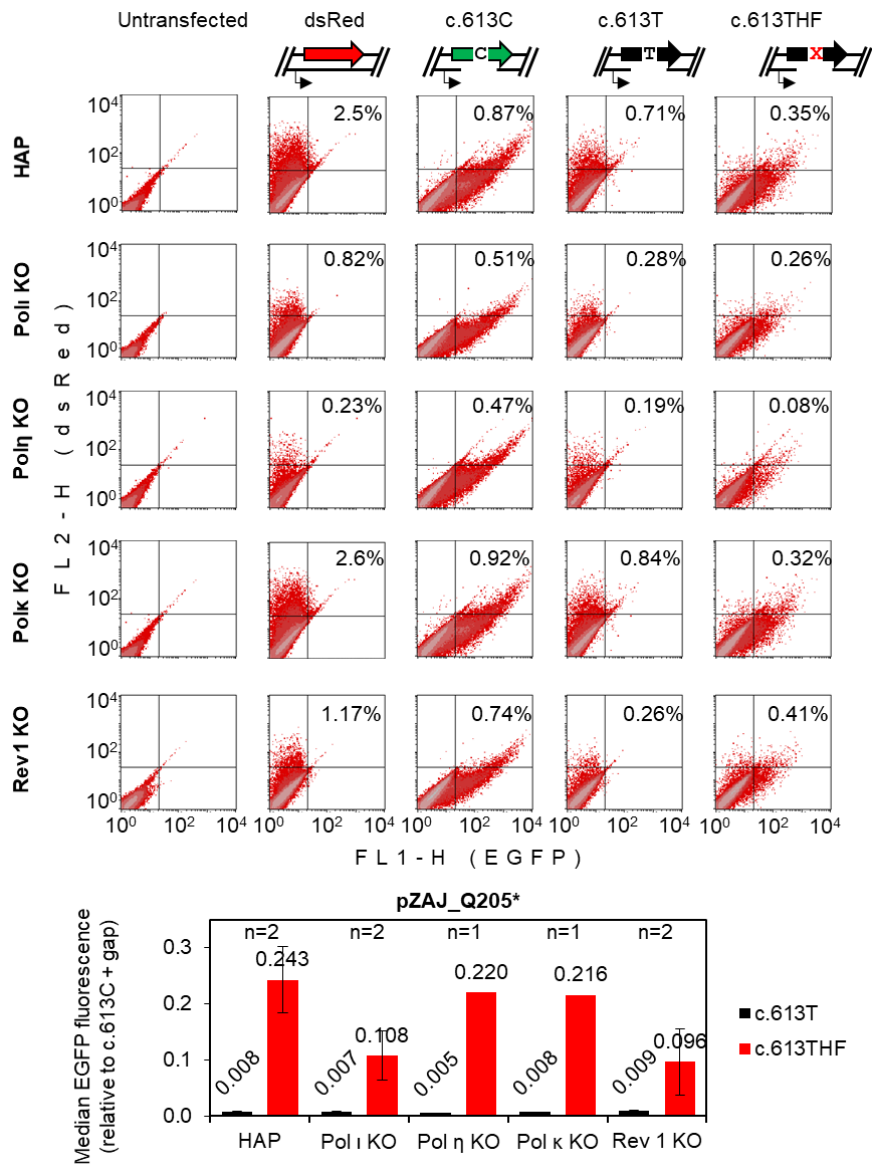


Figure 3-54: Detection of mutagenic TLS over synthetic AP sites using pMR_Q205* mutant in HAP and Y-family polymerases KOs cells. FACS analysis 24 hours post transfection with 3 different constructs: C (control), T (control) and THF (X) in position c.613 opposite to an 18-nucleotide gap. Above: Scatter plots of a representative experiment are shown. Below: Quantification of the median EGFP fluorescence relative to c.613C opposite to an 18-nucleotide gap.

3.4 Mutation profile of abasic DNA lesions during replicational and transcriptional bypass

3.4.1 RNA sequencing library preparation

To characterise the mutagenic footprint generated by erroneous bypass of AP sites during replication and transcription, we created RNA sequencing libraries using different cell lines and constructs for TM as well as TLS assays. In both cases, sample preparation was identical and involved eight consecutive steps: transfection, FACS analysis, RNA extraction, cDNA synthesis, analytical PCR, first PCR using primers containing adaptors, second PCR using indexes to label specific samples and finally the RNA sequencing run (**Figure 3-55**) [2.2.23-2.2.32].

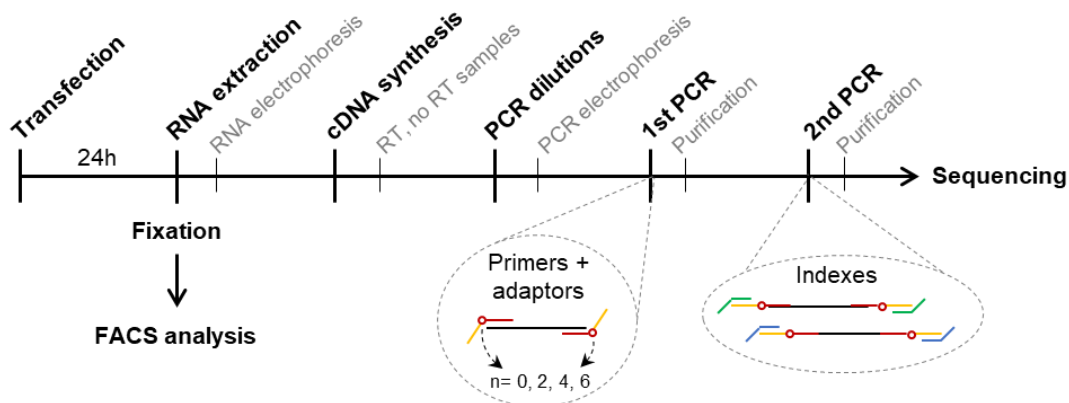


Figure 3-55: Experimental approach to create RNA sequencing library. The main experimental steps are depicted in bold letters, checkpoint experiments are depicted in grey.

To start the library preparation, we transfected the constructs in duplicates into the desired cell-line. All constructs carried a specific AP lesion either for TM or mutagenic TLS detection and their respective controls as described in previous sections 3.2 and 3.3 (**Supplementary table 1** and **Supplementary table 2**). One part of the transfected cells was fixed for FACS analysis and the other part was used for RNA extraction [2.2.23]. The FACS analysis allowed us to monitor a good transfection efficiency and the percentage of reversal to a fluorescent EGFP detected served as a control for further sequencing analysis. After RNA isolation, we divided the samples into three different parts: the first set of samples was stored at -80°C because of their low stability in solution, thus serving as a backup in case of RNA degradation; the second part was analyzed by RNA electrophoresis [2.2.24] and the third part was used for reverse transcription (cDNA synthesis) as described in 2.2.25. The RNA electrophoresis gel showed in **Figure 3-56** depicted two intense bands that corresponded to 28S and 18S rRNA in all four samples. As these bands were clearly defined, this result illustrated a good quality and integrity of the extracted RNA that can be further used as a template for the cDNA synthesis.

Results

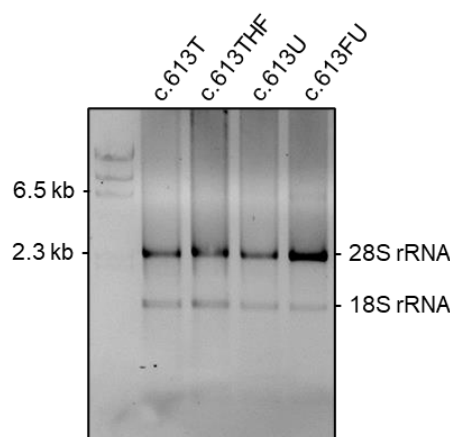


Figure 3-56: 1% agarose RNA electrophoresis gel of RNA extracted from HAP cell-line transfected with c.613T/THF/U/FU constructs. The upper band corresponds to 28S ribosomal RNA (rRNA) and the lower band to 18S ribosomal RNA. This gel illustrates a representative experiment that was repeated for all samples prepared in libraries 1 and 2.

Once the quality of the samples was verified, the RNA served as a template for cDNA synthesis. In these experiments, we used a gene-specific primer (GSP) that targets the mRNA encoding for the EGFP. In the optimization steps, GSP showed a more intense signal corresponding to a better amplification of the target sequence than the one obtained using random hexamers (**Supplementary figure 2**). The cDNA synthesis included incubation of the RNA samples in the presence (*RT* samples) and absence (*noRT* samples) of reverse transcriptase. Thus, if we detected any amplification of *noRT* samples after PCR, we could estimate the amount of DNA contamination that had been carried during the RNA extraction procedure. For this purpose, we performed an analytical PCR using serial dilutions of the cDNA samples (1:30, 1:300, 1:3000 for *RT* samples and 1:3, 1:30, 1:300 for *noRT* samples) as described in methods section 2.2.26 and run the resulting PCR products into a 1.5% agarose gel. Lack of DNA detection in the negative control, using water (-) instead of cDNA, indicated that there was no DNA contamination among the reagents used for the experiment (**Figure 3-57**). To validate the correct amplification of the fragment of interest, 3 pg of pZAJ_5c vector were used as a positive control (+). Here an intense signal of the expected size (174 bp) indicated that the PCR reaction worked properly. A faint band amplified in all 1:3 *noRT* samples of c.613THF/U/FU and further disappeared using 1:30 dilution factor. Samples that were subjected to reverse transcription (RT), showed strong band amplification in 1:30 dilution that was also visible at 1:300, thus indicating a low level (<1%) of DNA contamination. Surprisingly, DNA contamination seems to be higher in c.613T samples showing amplification of the template in *noRT* 1:30 dilution. This result suggested that the contamination level in this sample is below 10% but higher than in other cDNA samples. Since all samples had been treated in parallel, different amplification patterns pointed to a stochastic contamination arising from an unknown source. Still, as amplification of the

fragment of interest was positive in all *RT* samples, both RNA extraction and reverse transcription were successful.

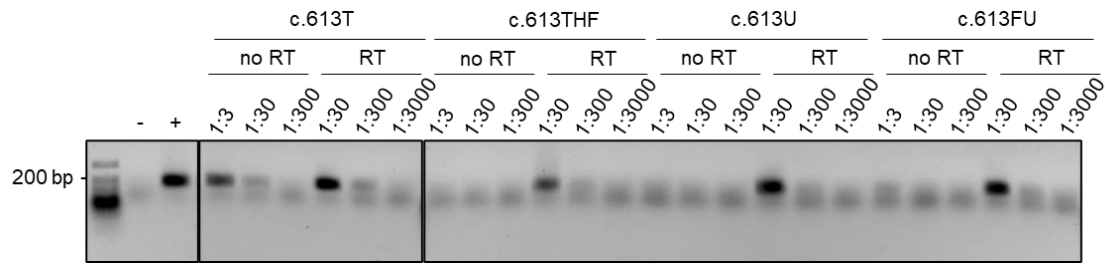


Figure 3-57: Representative analytical PCR of cDNA diluted samples from c.613T/THF/U/FU constructs transfected into HAP cells. 1.5% Agarose gel of RT and no RT samples, -: water control, +: 3 pg of pZAJ_5c template. Fragment size expected: 174 bp.

Once the cDNA had been synthesized, 1:3 dilutions of the cDNA samples worked as a template for the first PCR amplification round using a set of 8 primers as described in 2.2.27. Each of these primers contains a part that anneals with the template sequence and an additional part called adaptor. The adaptors were stretches of 20 additional nucleotides that anneal with the primers designed for the subsequent round of PCR. In addition, these primers had been designed to accommodate 0, 2, 4, 5, or 6 random nucleotides between the annealing (red) and the adaptor (yellow) part (**Figure 3-58A**). This allowed the generation of shorter and longer amplicons from the same template, which prevented the overlap between the sequences that might be misinterpreted by the sequencing system. Indeed, all amplicons resulting from this reaction had a length between 174-185 bp. As we use the product of this PCR as a template for the second PCR amplification, the success of the reaction was analyzed only by loading the negative (water) and positive (pZAJ_5c) control into a 1.5% agarose gel. As expected, we observed a lack of amplification in the negative control and clear amplification of the pZAJ_5c template within the expected size of 174-185 bp indicating that the reaction worked properly (**Figure 3-58B**).

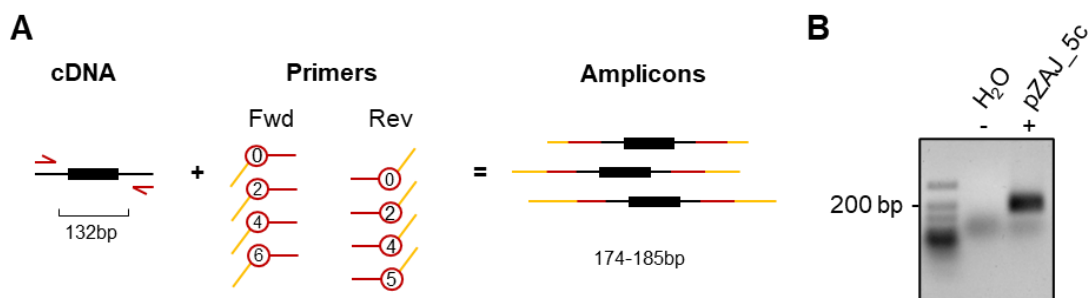


Figure 3-58: Scheme and validation of 1st PCR amplification. (A) Out of scale scheme of the first PCR amplification using the recently synthesized cDNA as a template. The black rectangle indicates the EGFP region of interest; red line: forward (fwd) and reverse (rev) primers; yellow line: adaptors; numbers within the circles correspond to the number of random nucleotides included in each primer to generate a variety of amplicon lengths. (B) 1.5% Agarose gel of PCR products from amplification of negative and positive control using a mix of 4 forward and 4 reverse primers. Fragment size expected: 174-185 bp.

Results

To distinguish individual samples during the sequencing process and the bioinformatic analysis, we used a different combination of primers (indexes) to label each sample in the second PCR round (**Figure 2-2**). All PCR experiments were followed by a DNA purification step as described in 2.2.28 to get rid of the excess of primers and primer dimers that could interfere with the sequencing process. Once again, we verified the correct amplification of the samples by loading the positive and negative control of the second PCR into a 1.5% agarose gel.

To finalize the library preparation, we sent our amplicons for sample quality control. Here, high-sensitive accurate measurement of the DNA concentration by Qubit was performed as well as the content of the DNA sample by Bioanalyzer [2.2.29]. By using the concentrations obtained via Qubit analysis, we created a pool where all samples were adjusted to a final concentration of 4 nM [2.2.30] and sent it for Next Generation Sequencing (NGS) via MiSeq System from Illumina (details about the sequencing procedure are described in 2.2.31).

3.4.2 Adenine is the most frequent ribonucleotide incorporated opposite synthetic abasic sites during transcriptional bypass

We previously reported in section 3.2.3 that erroneous bypass of synthetic AP lesions during transcription led to the synthesis of a fluorescent EGFP in cells deficient in nucleotide excision repair (XP-A cells). To unravel the mutation signature of synthetic AP lesions during transcriptional mutagenesis, we analyzed the transcripts synthesized from pMR_A207P-derived constructs (**Supplementary table 3**). After transfection in duplicates of ts.619G (dG), ts.619THF (THF), and the BER-resistant ts.619S-THF (S-THF) into XP-A cells, one part was fixed for FACS analysis, and another part was used for RNA extraction. By design, constructs generated from this reporter show a gain of EGFP fluorescence if either guanine (G), uracil (U), or adenine (A) are incorporated into the nascent mRNA opposite to the lesion. Expression analyses of the construct containing a single BER-resistant AP lesion (S-THF) at the nucleotide 619 revealed a 55% increase of EGFP fluorescence compared to the reference constructs dG or BER-sensitive AP site (THF) (**Figure 3-59A**). This result indicated that transcriptional bypass of S-THF resulted in the misincorporation of guanine, uracil and/or adenine in, at least, half of the transcripts synthesized.

Prior to the in-depth analysis of the RNA sequencing results, it was necessary to estimate the probability of error and to verify levels of possible contamination. The error introduced by the polymerases during the whole sequencing protocol was estimated by annealing the first 62 bases of each library with a template sequence [2.2.32]. As a result, the mean

error obtained was 12.76 for every 100000 bases sequenced, meaning a 0.0128% error rate with a confidence interval (CI) of 99%. To estimate the DNA contamination that might occur during the RNA extraction process, we further analyzed the sequencing result of one of the control samples *RT* versus *noRT* (ts.619G). After transfection of ts.619G double-stranded construct, it was expected that RNA polymerase II incorporated cytosine opposite to guanine during transcription. Indeed, we detected 91-99% of cytosine signal in all RT samples (**Figure 3-59B**); surprisingly, these sequencing results showed an additional guanine signal between 0.6-8.25%. This guanine fraction was too high and too variable to be an RNA polymerase error while transcribing dG (Imashimizu et al., 2013; Shaw et al., 2002). Moreover, the expression analysis only showed 0.01% of relative EGFP fluorescence which does not correlate with 0.6-8.25% of guanine incorporation that would reactivate the protein fluorescence completely (**Figure 3-9**). Because these cells had not been transfected with any plasmid containing dC in position ts.619, the origin of the guanine signal detected in all three repetitions could not derive from DNA present within the cells that contaminated the RNA extraction process. After studying the corresponding *noRT* samples, we noticed a highly variable guanine signal among experiments. This result together with a variable amplification in all *noRT* samples in the analytical PCR suggested a stochastic source of contamination. This might arise during PCR amplification and originate from a functional EGFP plasmid that contains ts.619C sequence (carry-over contamination). Interestingly, the third replicate of the ts.619G *noRT* sample showed 70% of guanine signal that was absent in the corresponding RT sample. This result suggested that this was the experiment with less DNA contamination during RNA extraction, being the carry-over the major source of DNA template for amplification when cDNA synthesis did not occur (*noRT* sample). Since reverse transcriptase was absent in *noRT* samples, the amplification of the cytosine signal indicated an additional source of DNA contamination with the transfected plasmid during RNA extraction.

Expression analysis of BER-sensitive abasic site (THF) templates did not show any regain of EGFP fluorescence, likely because of its efficient repair within the cells (**Figure 3-59A**). The sequencing result of a single experiment using this template showed a high cytosine signal (95%) within the transcripts at this position. This result confirmed that THF was effectively repaired and substituted by guanine before transcription occurred (**Figure 3-59C**). Once again, guanine signal in both *RT* and *noRT* samples did not correlate with the EGFP fluorescence analysis and indicated carry-over contamination with a functional EGFP plasmid during PCR experiments. In summary, these results indicated that any value gathered below 0.0128% should not be taken into consideration.

Results

Additionally, we assumed stochastic contamination with a wild-type EGFP coding sequence that leads to a variable guanine signal in all samples analysed.

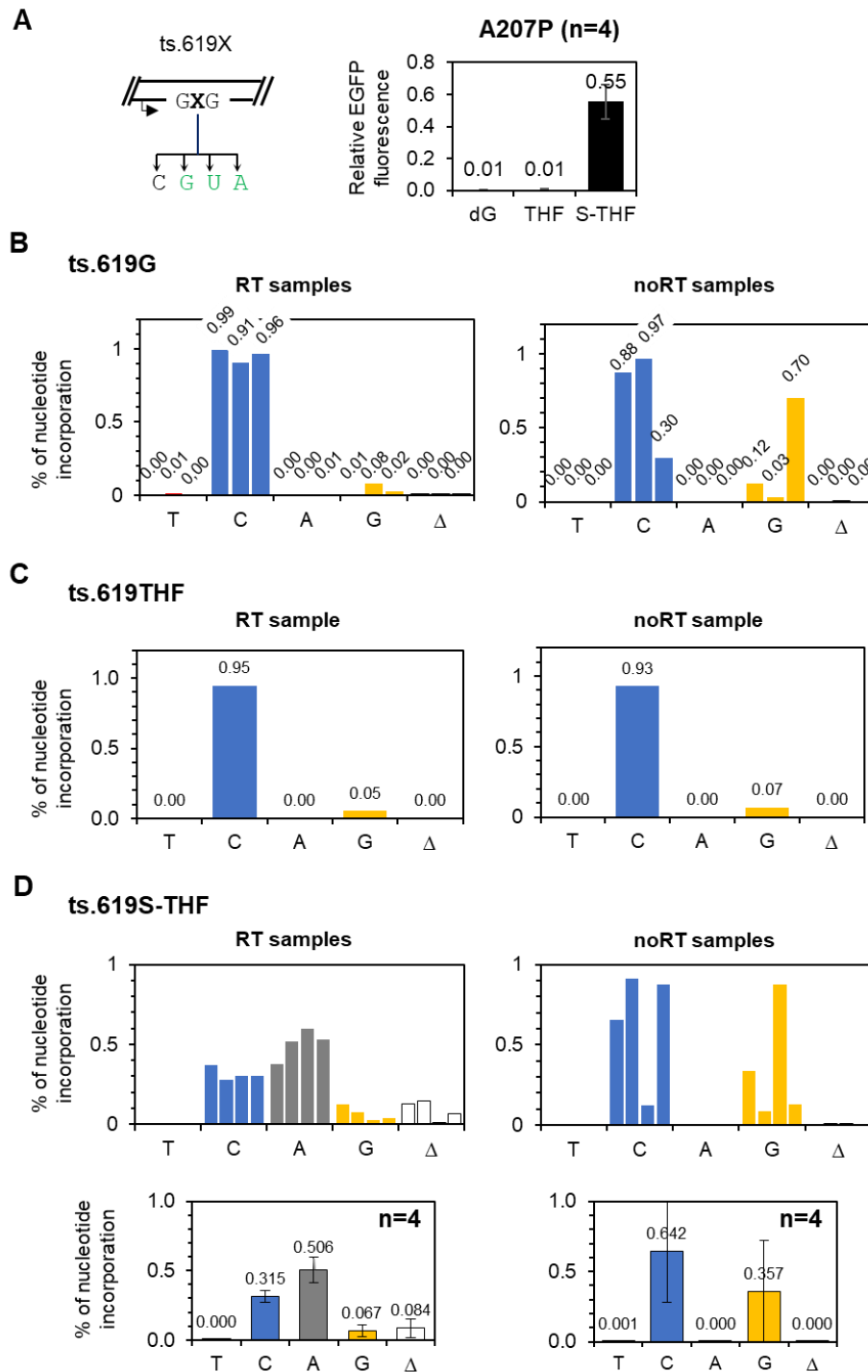


Figure 3-59: RNA sequencing results of ts.619G/THF/S-THF constructs transfected in XP-A cells. T: thymine, C: cytosine, A: adenine, G: guanine, Δ: deletion. **(A)** Simplified scheme of phenotypic outcome after S-THF transcriptional bypass in pMR_A207P plasmid X=S-THF (left). Quantification by FACS of relative EGFP fluorescence of four independent experiments in XP-A cells (right). **(B)** Percentage of nucleotide incorporation opposite dG of RT (left) and no RT (right) samples represented by three individual experiments. **(C)** Percentage of nucleotide incorporation opposite THF of RT (left) and no RT (right) samples. **(D)** Percentage of nucleotide incorporation opposite S-THF lesion of RT (left) and no RT (right) samples, upper graphs represent individual experiments, lower graphs represent the mean \pm SD, n=4.

To uncover the pattern of ribonucleotides incorporated opposite BER-resistant AP sites responsible for the 55% reactivation of the fluorescence (**Figure 3-59A**), we analyzed the RNA sequencing results of the ts.619S-THF sample. Adenine, with at least 50.6%, was the most common nucleotide incorporated opposite to BER-resistant AP lesions during TM (**Figure 3-59D**). Hence, this result accurately correlated with the amount of EGFP fluorescence detected in the FACS analysis. In addition, we observed 8.4% of deletions that were undetectable in noRT samples suggesting that they appeared in the mRNA as a result of the transcriptional bypass of the lesion. These deletions showed a specific pattern that started from the nucleotide previous to the lesion (-1) with a length between 9 and 12 nucleotides (**Figure 3-60**).

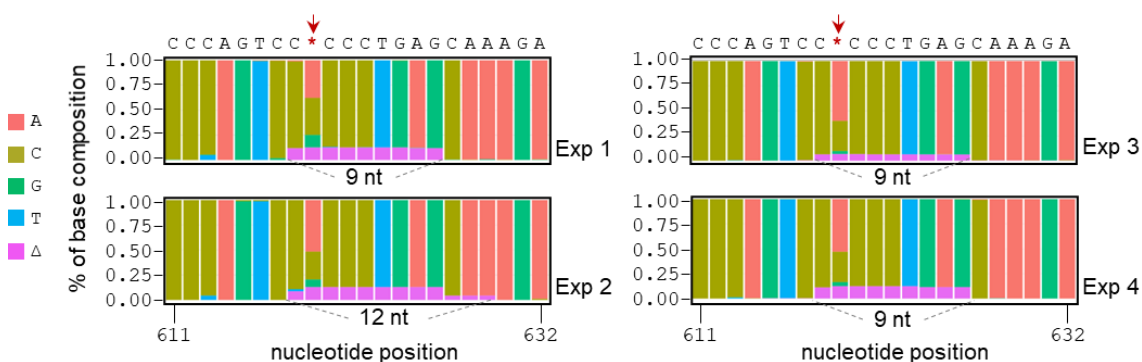


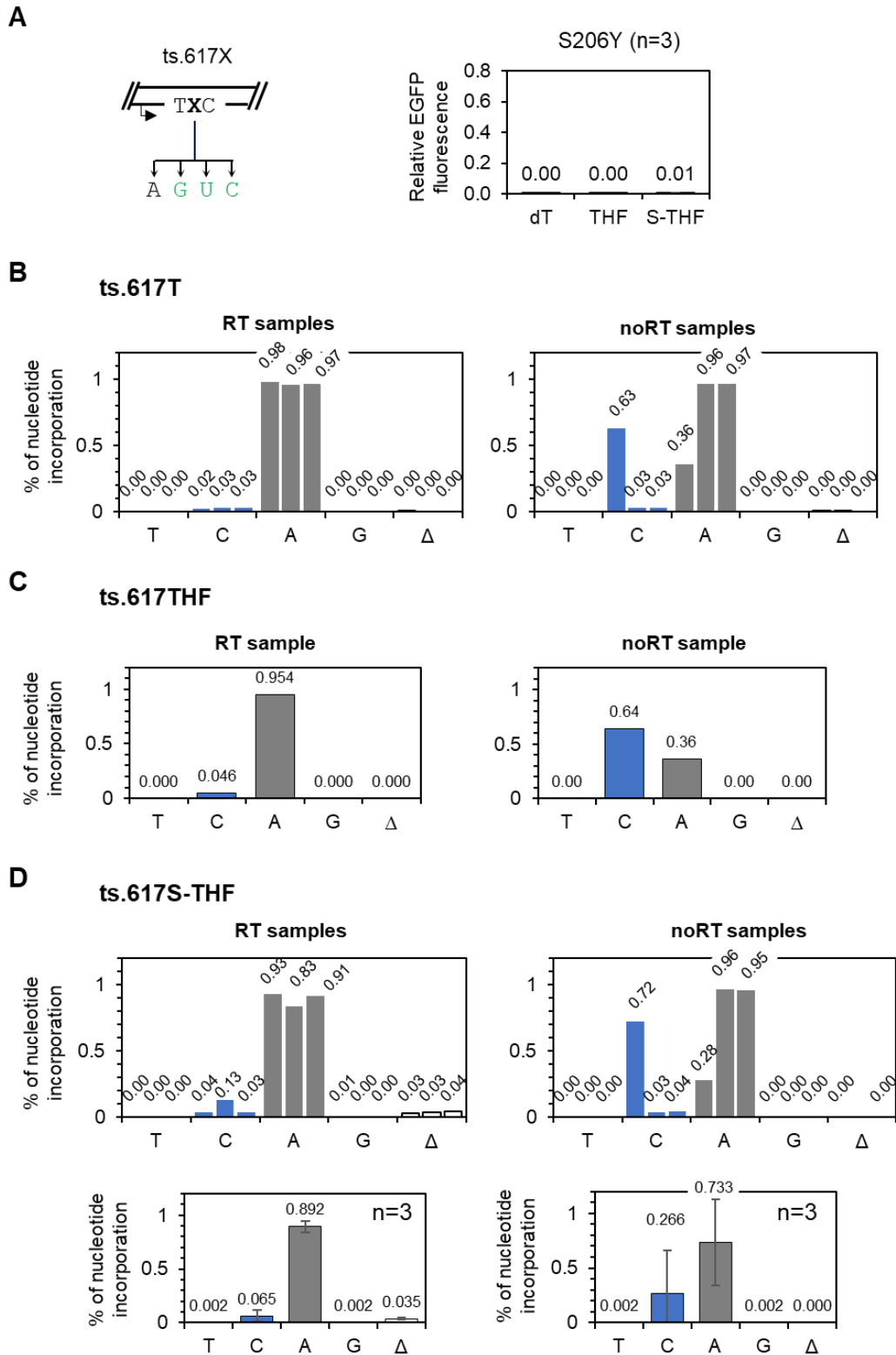
Figure 3-60: S-THF libraries showing % of nucleotide substitution and deletions from the nucleotide 611-632. Red * and arrow indicate the position of S-THF, each bar represents one nucleotide position and the number underneath the composition table indicates the length of the deletions with >1% frequency.

Taking into consideration that the previous results of the control samples (G and THF) indicated that the guanine signal was a carry-over from functional EGFP plasmid, we ascribed the 6.7% of guanine signal to the same carry-over in ts.619S-THF samples. Interestingly, 31.5% of cytosine detected in RT samples did not correspond to the mutation profile of AP sites as a result of the transcriptional bypass as it was also present in noRT samples. In this context, we hypothesized that cytosine signals might arise as a result of contamination with a transfected S-THF construct that has been previously repaired within the cells.

To prove this hypothesis, we decided to use an additional construct derived from an alternative mutated plasmid where the repair of the lesion would not lead to a cytosine signal in the mRNA. Thus, if the cytosine signal persisted, cytosine could be considered as part of the mutation signature of S-THF. However, if the cytosine signal disappeared after placing the S-THF in this position, it would indicate that the lesion was partially repaired within the cells. We, therefore, generated constructs for TM assay starting from the pMR_S206Y (ts.617T) mutant. In this position, reversal to a fluorescent EGFP was only possible if transcriptional bypass occurred through the incorporation of cytosine,

Results

uracil, or guanine opposite to S-THF. Analysis of the relative EGFP fluorescence 24h after transfection, showed a complete lack of fluorescence in all cells transfected with S-THF as well as with THF and dT control constructs (**Figure 3-61A**). This result suggested that either the S-THF lesion was completely repaired restoring dT in position ts.617 or that transcriptional mutagenesis only occurs through adenine incorporation opposite to the lesion. Accordingly, the RNA sequencing analysis showed an extremely large adenine signal (93-98%) in all RT samples transfected with S-THF (**Figure 3-61D**). It is important to note that, contamination with a wild-type EGFP coding sequence would lead to a cytosine signal at this position. As a consequence, a variable cytosine signal was obtained in S-THF samples as well as in ts.617T (**Figure 3-61B**) and ts.617 THF (**Figure 3-61C**) that did not correlate with the results obtained in the FACS analysis. These results confirmed that we have overall carry-over contamination with plasmids encoding for a functional EGFP.



Results by Leen Sarmini

Figure 3-61: RNA sequencing results of ts.617S-THF constructs transfected in XP-A cell-line. T: thymine, C: cytosine, A: adenine, G: guanine, Δ: deletion. **(A)** Simplified scheme of phenotypic outcome after S-THF transcriptional bypass in pMR_S206Y plasmid X=S-THF (left). Quantification by FACS of relative EGFP fluorescence of four independent experiments in XP-A cells (right). **(B)** Percentage of nucleotide incorporation opposite thymine of RT (left) and no RT (right) samples corresponding to three independent experiments. **(C)** Percentage of nucleotide incorporation opposite THF lesion of RT (left) and no RT (right) samples. **(D)** Percentage of nucleotide incorporation opposite S-THF lesion of RT (left) and no RT (right) samples corresponding to three independent experiments.

Results

As we also expected an exclusive adenine signal arising from cells transfected with control templates (T and BER-sensitive THF), we could not confirm whether the 93.1% of adenine in cells transfected with S-THF raised after the repair and/or the bypass of the lesion within the cells. To address this question, it was necessary to compare the results obtained in positions 619 and 617. Thus, if S-THF undergoes complete repair within XP-A cells, we would have observed a complete lack of fluorescence in both positions (**Figure 3-59A** and **Figure 3-61A**) and a complete absence of adenine signal in position 619 (**Figure 3-59D**). Since that was not the case, we assumed that the 93.1% of adenine signal in position 617, raised as a summatory effect of the transcriptional bypass of the lesion accompanied by adenine incorporation and its repair within the cells. A partial repair of the S-THF substrate would also explain the 31.5% of cytosine signal obtained in position 619, pointing to the 50.6% of adenine signal as the mutation signature of S-THF during TM (**Figure 3-59D**). (All results presented in **Figure 3-61** were obtained by Leen Sarmini, Ph.D. student in AG Khobta). Altogether, the results presented in this section indicate that the S-THF lesion is partially repaired within XP-A cells and that its transcriptional bypass results in adenine incorporation.

3.4.3 Mutagenic TLS over synthetic and natural AP lesions results mostly in adenine incorporation

We have previously reported that mutagenic bypass of abasic DNA lesions was occurring during DNA synthesis [3.3.4] and that bypass of a natural AP site led to higher mutagenicity than its synthetic analogue THF (**Figure 3-62A**). To disclose the mutation profile of both types of abasic sites during TLS, we analyzed the transcripts resulting from HAP cells transfected with c.613T, THF, U and 2'FU placed opposite to gapped templates (**Supplementary table 4**).

In order to properly drive conclusions from the sequencing results, it was necessary to study the results obtained from the control samples c.613T and c.613FU. The sequencing results showed variable adenine and guanine signals in all *noRT* samples that indicated two potential sources of contaminations (**Figure 3-62B**). The adenine signals suggested amplification of the plasmid DNA whose gap had been properly repaired within the cells and contaminated the sample during RNA isolation. Conversely, the guanine signals suggested carry-over contamination with a functional EGFP plasmid that contained a cytosine in position 613. Analysis of the RT sample corresponding to the control template containing thymine (c.613T) showed not only the expected adenine signal but also a considerably high guanine signal. DNA polymerases are too accurate to be responsible for such a high percentage of guanine misincorporation opposite to

thymine. Besides, the FACS analysis showed a complete lack of fluorescence in cells transfected with c.613T templates, thus confirming that the high guanine signal present in *noRT*, as well as *RT* samples, emerged from contamination that occurred after RNA extraction. Thereby, the correct repair of the gap in the c.613T template materialized in an average high adenine signal. Similar results were obtained after FACS (**Figure 3-62A**) and sequencing (**Figure 3-62B**) analysis of the transfected c.613FU template. As the 2' fluorine atom protected the uracil base from UDGs activity within the cells, TLS over this uracil residue led to the incorporation of adenine and the maintenance of a non-fluorescent status of the protein. Since guanine incorporation opposite to uracil did not correlate with the lack of fluorescence obtained after FACS analysis, we assumed that the guanine signal appeared due to carry-over contamination with a functional EGFP plasmid.

Interestingly, both synthetic (THF) and natural abasic site (U) templates showed adenine as a preferred nucleotide incorporated during lesion bypass (**Figure 3-62B**). Unfortunately, we could not disregard the guanine signal in these results due to the possibility that part of it belongs to the mutation profile of the lesion. However, 18.4% of EGFP fluorescence detected in cells transfected THF template (**Figure 3-62A**) correlated with the sum of cytosine (14%) and thymine signal (4%) detected in the sequencing results. Indeed, the 23% of guanine signal detected in this sample would have resulted in a much higher EGFP fluorescence in the FACS results. On the contrary, 45.3% of EGFP fluorescence detected in cells transfected with U constructs barely correlates with the sum of cytosine (10%) and thymine (5%) signals. In order to achieve the level of fluorescence detected in c.613U samples, we must assume that part of the guanine signal (40%) was, indeed, part of the mutation profile of natural AP sites during TLS. In summary, these results suggested that the mutation profile varies between synthetic and natural AP lesions, with a high likelihood of being A>C>T> Δ >G opposite to THF and A>G>C>T> Δ opposite to natural abasic sites.

same signal as we detect from the isolated mRNA, analysis of *noRT* samples was no longer necessary.

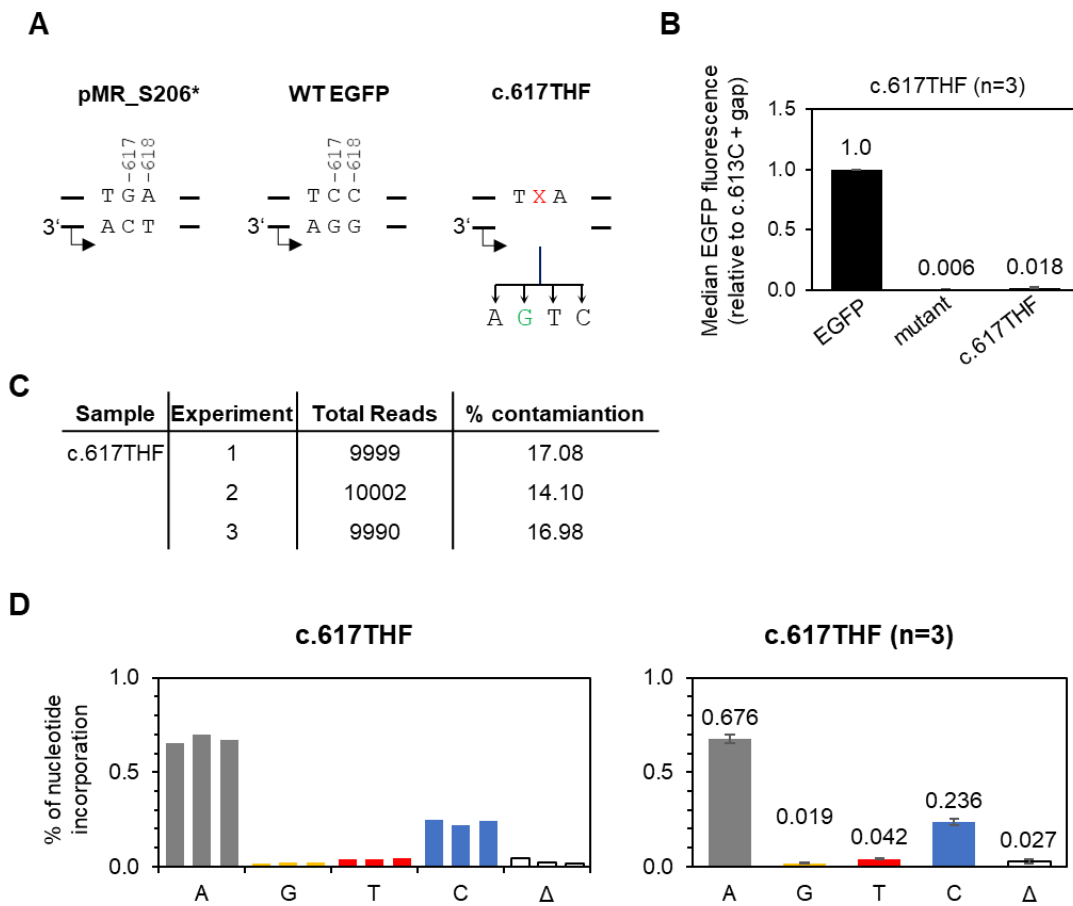


Figure 3-63: RNA sequencing analysis of c.617THF construct in the double mutated plasmid pMR_S206*. (A) Scheme of constructs transfected in HAP cells. (B) Quantification of relative EGFP fluorescence after transfection (C) The raw number of reads obtained in positions 617 and 618 by RNA sequencing analysis (D) Quantification of % of nucleotide incorporated opposite THF in position 617 and opposite to A in position 618 without carry-over.

In pMR_S206* mutant, EGFP fluorescence can only be restored by incorporation of guanine opposite to the lesion (**Table 3-2**). Fluorescence analysis of HAP cells transfected with c.617THF showed a complete lack of EGFP fluorescence (**Figure 3-63B**). This result indicated that either bypass of THF in position 617 did not occur or that only nucleotides that kept the non-fluorescent status of the protein were incorporated opposite to the lesion. Since we have previously demonstrated that THF is bypassed in a mutagenic manner, this lack of EGFP fluorescence suggested that TLS occurred but did not result in guanine incorporation opposite THF; instead, adenine, thymine, or/and cytosine must have been incorporated opposite to the lesion. To confirm this assumption, it was necessary to analyze the RNA sequencing results of cells transfected with c.617THF templates. Here, all reads that contained a guanine signal in position 618 were bioinformatically eliminated as they emerged from wild-type EGFP carry-over

Results

contamination. Thus, we obtained a rather high average of 16.05% of contamination in these samples (**Figure 3-63C**), a percentage that also correlated with the guanine signal detected in c.613THF samples (**Figure 3-62B**). As previously reported in position 613, adenine appeared as the most frequent nucleotide incorporated opposite to synthetic AP sites (67.6%), followed by cytosine (23.6%), thymine (4.2%), deletions 2.7%, and a small proportion of guanine (1.9%) (**Figure 3-63D**). Even though we detected an average of 2.7% of deletions, they showed a length between 18 and 19 nucleotides that would correlate with the size of the gap in the transfected constructs (**Figure 3-64**). Thus, suggesting that the deletion detected by the RNA sequencing analysis might have its origin in a gapped template that had not been properly filled within the cells. Still, this accounts for a very small percentage of the total mutagenic profile of the lesion. Taking together, these results showed that the mutation profile of synthetic (THF) AP lesions adopt the following pattern in human cells: A>C>T>G.

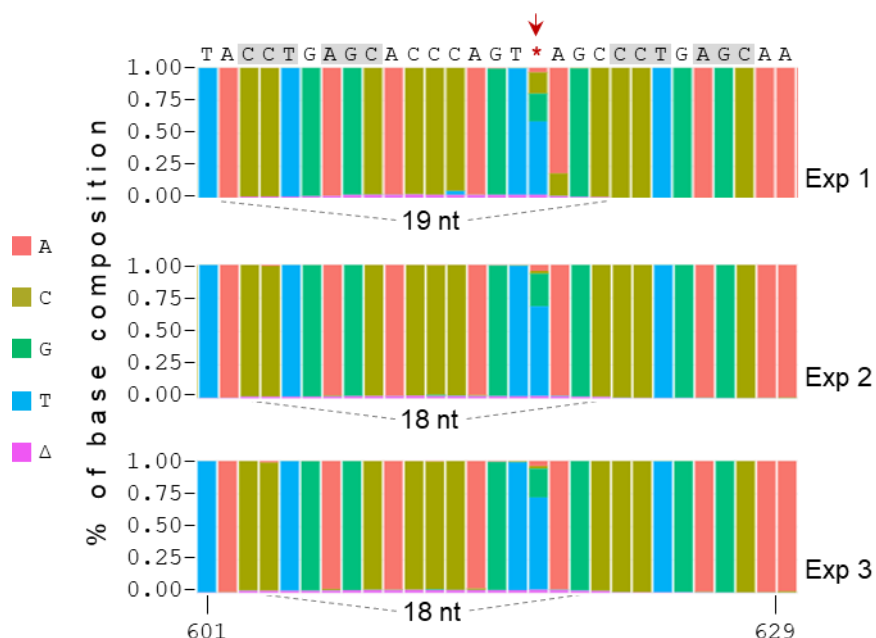


Figure 3-64: c.617 THF libraries showing % of nucleotide substitution and deletions from the nucleotide 601-629. Recognition nicking sites of the *Nt.Bpu10I* (grey shadow in the sequence), red * and arrow indicate the position of the THF, each bar represents one nucleotide position and the number underneath the composition table indicates the length of the deletions with >1% frequency. The bar chart of experiment one shows percentages before extracting the EGFP carry-over while the bar charts of experiments 2 and 3 show percentages after extraction of the carry-over.

3.4.5 The mutation profile of natural AP sites during TLS is A>G>C>T

Since a double mutation within the constructs proved to be very useful for the analysis of the mutation profile of THF AP sites, we aimed to follow the same strategy to study the mutation profile of natural AP sites. Because constructs originated from the EGFP S206* mutant only reactivated the protein fluorescence upon guanine incorporation, we designed an alternative experiment using Q205* mutant as a plasmid backbone (**Figure**

3-65A). As the incorporation of guanine, cytosine and thymine in position ts.613 reactivates the protein fluorescence, by using a modified Q205* mutant we were able to correlate the fluorescence analysis with the RNA sequencing results. After revising silent mutations that we discarded for the design of our mutant vectors in section 3.1.1, we found a position (c.612) that, regardless of the nucleotide present, would encode for a threonine amino acid in position 204 (**Figure 3-65B**). Thus, modification of the nucleotide in position 612 would create an internal control for contamination in the RNA sequencing analysis while at the same time maintain the fluorescence status of the protein. As a result, we generated four TLS constructs harboring guanine instead of cytosine in position c.612 and cytosine (C) (as a WT EGFP construct for normalization of the FACS data), thymine (T), uracil (U) or 2'fluorinated-uracil (FU) in position c.613 opposite to an 18-nucleotide gap (**Figure 3-65C**). An analytical ligation experiment confirmed the presence of the gap in the template strand of the constructs. Lack of cc form of the plasmid in templates incubated with T4 DNA ligase indicated a complete elimination of the 18-mers DNA stretch placed between the nicking sites (**Figure 3-65D**). The resulting constructs were transfected into HAP cells for FACS analysis and preparation of the RNA sequencing library (**Supplementary table 2**). During this second library preparation, we made small adjustments to the reverse transcription protocol that are indicated in 2.2.25.

Results

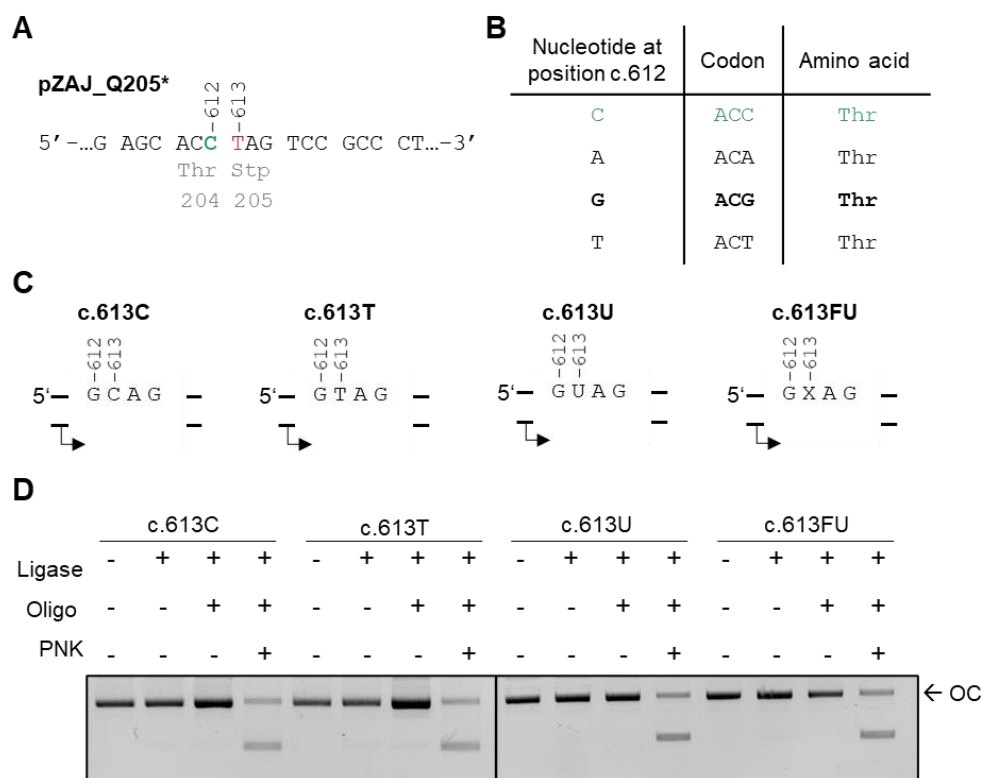


Figure 3-65: Design of constructs with double mutation for detection of mutagenic TLS over natural AP sites. (A) Scheme of the coding strand of the pZAJ_Q205* mutant vector: c.613T (red) shows the position to insert the DNA lesion, c.612C (green) shows the position for a potential second mutation that would serve as an internal control for the RNA sequencing process (B) All possible nucleotide changes in position 612 encode for the same amino acid: green corresponds to the original nucleotide present in position 612 of the EGFP sequence, bold indicates the nucleotide change chose as secondary mutation. (C) Out of scale scheme of the constructs used TLS assays and library preparation. (D) Analytical ligation to verify the presence of a gap in the TS of the constructs. Oligo: oligonucleotides, oc: open circular form of the plasmid.

Expression analysis of the constructs containing U revealed a 37% of reversal to a fluorescent EGFP relative to a wild-type EGFP construct (c.613C) (Figure 3-66A). Restoration of the non-fluorescent status of the protein observed in cells transfected with the glycosylase resistant construct c.613FU (1.6%) indicated that a negligible percentage of the mutation frequency observed in c.613U appeared after TLS over uracil residues. Instead, TLS over the resulting natural AP sites was accompanied by a high rate of guanine, thymine, and/or cytosine misincorporation that led to the reactivation of the EGFP fluorescence.

To unravel the mutation profile of the natural AP sites formed as excision of uracil in c.613U templates, we analyzed the percentage of nucleotides incorporated opposite to the lesion via RNA sequencing. Prior to the final analysis, all reads that contained a guanine signal in position 612 were bioinformatically eliminated [2.2.32]. This allowed the removal of all reads that emerged from wild-type EGFP carry-over contamination. The average estimated error for this library was 0.241% with a confidence interval of (CI) of 99%, meaning that results below that percentage were not considered in the overall

analysis. As expected for the control constructs (T and FU), only adenine signals appeared in the three independent repetitions with an average percentage of 97.7 and 95.9 respectively (**Figure 3-66B**). Analysis of the c.613U constructs revealed adenine as the most frequent nucleotide incorporated opposite to physiological AP sites (43.5%), followed by guanine (23.5%), cytosine (20%), then thymine (10.9%), and a small percentage of deletions (2.1%). Altogether, these results indicate that the mutation profile of natural AP sites is A>G>C>T> Δ and confirmed that bypass of synthetic versus natural AP lesions leads to different mutagenic outcomes during TLS.

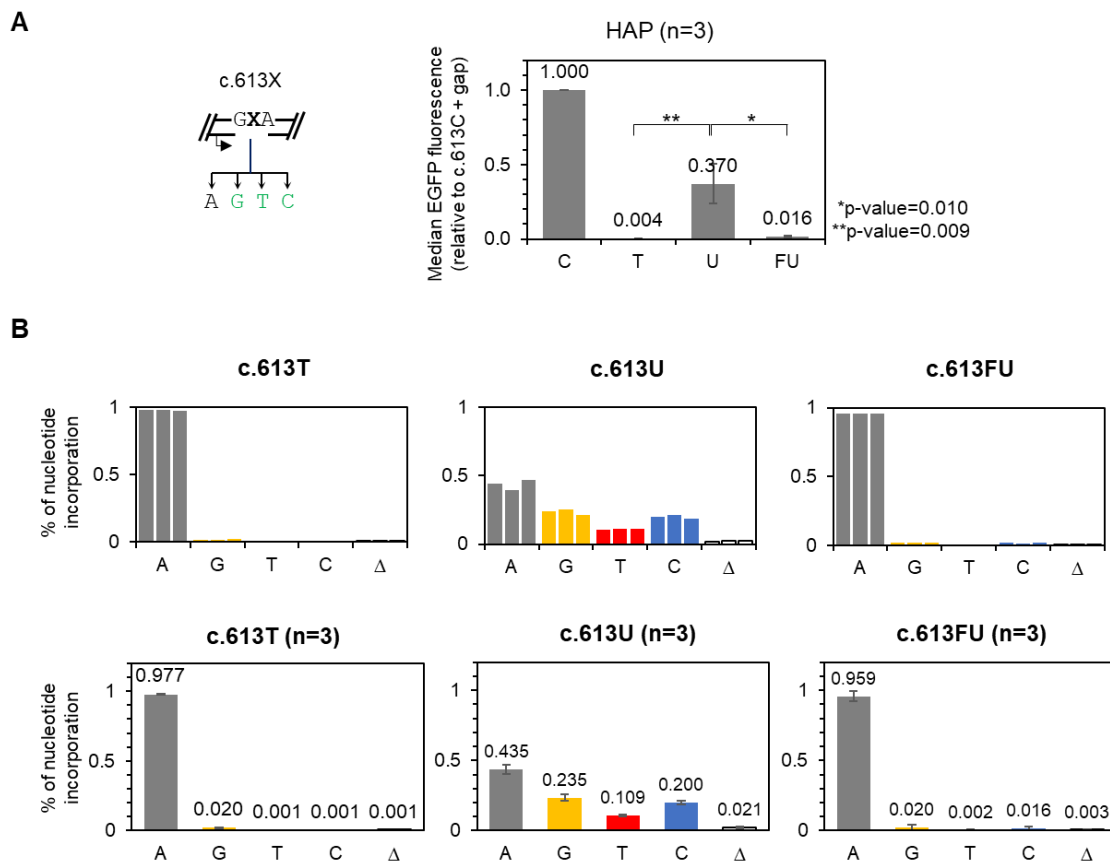


Figure 3-66: FACS analysis and RNA sequencing results of c.612G-c.613T/U/FU constructs for TLS assay transfected into HAP cell-line. (A) Left: Simplified scheme of phenotypic outcome after AP site resulting from uracil excision is bypassed. Right: Relative EGFP fluorescence of three independent experiments (mean \pm SD, n=3). (B) Percentage of nucleotide incorporation opposite to A, U and FU after TLS. Top: Results of the three independent experiments. Bottom: mean \pm SD, n=3. T: thymine, C: cytosine, A: adenine, G: guanine, Δ : deletion

4 Discussion

4.1 A novel approach for direct and sensitive detection of mutagenic bypass of DNA lesions

The investigation presented in section 3.1 describes the design and generation of a set of EGFP mutant vectors that enable the direct detection of mutagenic bypass of DNA lesions during replication and transcription in mammalian cells. Each of these mutants contains a different point mutation in the coding sequence of the EGFP gene that results in the elimination of the protein's fluorescence. Interestingly, any alternative base substitution at the mutated nucleotide leads to the synthesis of a functional EGFP. Thus, by placing a DNA lesion instead of the mutated nucleotide, the reactivation of the protein fluorescence would be a direct indication of the mutagenicity of the lesion during its bypass.

To assess the capacity of DNA or RNA polymerases to bypass a certain type of damage, cell-free systems performing primer extension assays have been generally used (Bré Geon et al., 2003; Choi et al., 2010; Locatelli et al., 2010; Patra et al., 2015; Tornaletti et al., 2006). Although mechanistically very useful, these experiments are performed in a controlled environment that rarely occurs within a living cell. Furthermore, multiple DNA polymerases co-exist in the cellular milieu and the ability of one of them to bypass specific damage *in vitro* does not necessarily correlate with its function under physiological conditions (Avkin et al., 2002; Patra et al., 2015). Inevitably, we considered the use of reporter vectors that can be directly transfected into living cells, since they have been an efficient tool to study DNA lesion bypass during both replication (Avkin et al., 2002; Weerasooriya et al., 2014) and transcription (Nagel et al., 2014; Petrova et al., 2016). However, these assays are technically laborious and specially designed to only detect TM or mutagenic TLS over a specific lesion. Therefore, we aimed to establish a new method that would facilitate the detection of erroneous bypass of damage during DNA and RNA synthesis via fluorescence reactivation assays. This type of assay would require a vector encoding for a loss-of-fluorescence protein that, after the substitution of the mutated nucleotide by a DNA lesion, its mutagenic bypass would lead to the reactivation of the fluorescence. We chose the enhanced green fluorescent protein (EGFP) because its fluorophore has the ability to generate fluorescence that can be efficiently quantified in living cells (Cormack et al., 1996). Moreover, this protein is extremely stable and it does not need cofactors for excitation of the fluorophore (Prasher et al., 1992; Tsien, 1998).

In order to design reporters that encode for a non-fluorescent EGFP protein, we took advantage of two tandem nicking sites located within the gene coding sequence described previously (**Figure 3-1** and (Kitsera et al., 2011; Lühnsdorf et al., 2012)). In particular, this stretch of nucleotides is essential for the correct folding and further excitation of the fluorophore (Brejc et al., 1997). After screening twenty-five nucleotide substitutions confined by the nicking sites, we found eight mutations that reasonably impaired the protein fluorescence (**Figure 3-5**). One of them was previously characterised by the group as a stop codon (Johannes Burggraaff). Four out of the remaining seven substitutions interestingly coded for a proline residue, which cyclic structure confers an inherent rigidity to the polypeptide chain that likely leads to a disrupted EGFP. Because the idea was to substitute the mutated nucleotide with the DNA lesion of interest, we hoped to find several mutations involving different nucleotides. Thus, we created and characterised three new mutant reporters: pMR_Q205P with a A>C mutation at position c.614, S206Y with a C>A mutation at position c.617 and A207P with a G>C mutation at position c.619 of the EGFP coding sequence. These mutants together with the preexisting pZAJ_Q205* with a C>T mutation in position c.613, compiled a set of reporters harboring a point mutation that leads to a non-fluorescent EGFP protein (**Figure 3-7**).

For the proper functioning of the TM and TLS assays, any secondary mutation occurring at the base pair affected by the primary mutation must lead to the synthesis of a fluorescent EGFP. Indeed, any alternative substitution at the mutated position in Q205*, Q205P and A207P vectors led to the synthesis of a fully functional EGFP with a >70% of fluorescent reactivation (**Figure 3-9**). However, the sensitivity of the pMR_S206Y reporter was limited since it did not show complete EGFP restoration when A and C, encoding for phenylalanine and cysteine residues respectively, were incorporated at position 617 in the template strand. It is likely that because the serine residue in position 206 (S206) is essential for the activation of the fluorescence of the WT protein, disruptions at this position have a high impact on the fluorescent outcome. S206 forms hydrogen bonds with water molecules that are essential for the maintenance of the neutral form of the fluorophore as well as for the transition from the neutral to the ionized form (Brejc et al., 1997; Tsien, 1998). Thus, it is expected that two aromatic amino acid residues with hydrophobic side chains, as tyrosine (ts.617T) and phenylalanine (ts.617A), affect the overall structure of the fluorophore (Anjana et al., 2012) depicting 0% and 2% of relative EGFP fluorescence respectively (**Figure 3-9**). Interestingly, we detected a 22% of residual EGFP fluorescence when a cytosine was incorporated at that position (ts.617C) encoding for a cysteine residue (**Figure 3-9**). Although both serine

Discussion

and cysteine residues have similar chemical structures and are hydrophilic, the oxygen present in the side chain of the serine residues possesses higher electronegativity compared to the sulfur present in cytosine residues. That might cause the formation of more hydrogen bonds between the serine side chain and water molecules, possibly resulting in overall higher fluorescence. In summary, the range of base pairs covered by the available mutants allows the analyses of mutagenic TLS at DNA damage affecting A, C or T in the coding strand. Conversely, transcriptional mutagenesis can be investigated for DNA lesions affecting A, G or T in the transcribed strand.

The pMR_S206A with a T>G mutation in position c.616 was the only mutant reporter with a mutated guanine in the coding strand. Unfortunately, transfection into HeLa cells showed that this substitution resulted in a partially fluorescent EGFP protein (31%) (**Figure 3-7**). Unlike the other hydrophobic residues substituted in this position (tyrosine and phenylalanine), the alanine residue in position 206 depicted partial fluorescence. This suggests that aromatic molecules in the side chain might have a higher impact than their water affinity in decreasing the fluorescence of the protein.

Because pMR_S206A could not be used for fluorescence reactivation assays due to its residual fluorescence, we additionally designed an alternative mutant containing a guanine mutation in the template strand. The pMR_S206* contains, not only one but two consecutive point mutations in positions c.617C>G and c.618C>A, that encode for a stop codon in transcribed mRNA. However, two of the three possible alternative substitutions at position 617 encoded for a stop codon in this mutant (**Figure 3-10A**). Also, the substitution of the mutated cytosine in position ts.617 by an adenine (ts.617C>A) encoding for a leucine residue resulted in an almost complete lack of fluorescence (**Figure 3-10B**). Consequently, this plasmid can be used to detect guanine misincorporation during replicative bypass and cytosine misincorporation during transcriptional bypass occurring at the lesion introduced at position 617. Interestingly, this double-mutated vector proved to be very useful for the RNA sequencing experiment. Here, the presence of a mutated nucleotide (c.618C>A) next to the target position used for inserting the DNA lesion (c.617) worked as an internal control. Thus, it allowed to filter out reads that showed cytosine signal in position 618 as they were attributable to contamination with WT EGFP (**Figure 3-63**). As a result, we developed a set of mutant reporters lacking EGFP fluorescence that can be used for the analysis of the mutagenicity of a broad range of DNA lesions in human cells.

Abasic DNA lesions are the most frequent damages arising in the DNA under physiological conditions and their non-bulky character makes them a perfect target for

bypass events during replication and transcription. Because AP sites do not form Watson-Crick base pairs with any of the DNA bases, these events result in the systematic misincorporation of nucleotides opposite to the lesion. Thus, transcriptional bypass of abasic sites leads to the synthesis of mutant mRNA (Doetsch, 2002; Viswanathan A et al., 1999). Similarly, nucleotide misincorporations opposite to abasic sites during DNA synthesis are used as a template for the subsequent repair of the lesion generating perpetual mutations within the genome (Choi et al., 2010; Nair et al., 2009; Otsuka et al., 2002; Patra et al., 2015).

Although extensive investigations have been performed in order to unravel the consequences of transcriptional and replicational bypass of abasic sites, there is a lack of consensus regarding the bypass mechanism and the mutation profiles of this lesion. In order to avoid the effective repair that abasic DNA lesions endure within the cells, most experiments are performed *in vitro*. Alternatively, *in vivo* studies generally use yeast as a model organism. Unlike the embryonic lethality showed in mammalian cells depleted of APE1 (Xanthoudakis et al., 1996), *apn1* and *apn2* deficient strains in yeast were viable (Guillet Marie & Boiteux Serge, 2002). This allows the study of the abasic sites bypass in a BER-impaired background where the half-life of the lesion is extended and the outcomes of bypass events are easily detected. To provide a more accurate picture of the bypass mechanisms of AP sites in higher eukaryotes, we created a suitable system to study the mutagenic bypass of abasic lesions using human cells.

The availability of the Nb.Bpu10I and Nt.Bpu10I nicking sites in all mutant reporters allows the insertion of synthetic oligonucleotides containing an abasic site at the mutated position in either DNA strand (Lühnsdorf et al., 2012). Consequently, the same EGFP mutant reporter can be used to study mutagenic TLS or TM simply by introducing the DNA lesion in the non-template (NTS) or the template strand (TS), respectively (**Figure 4-1**).

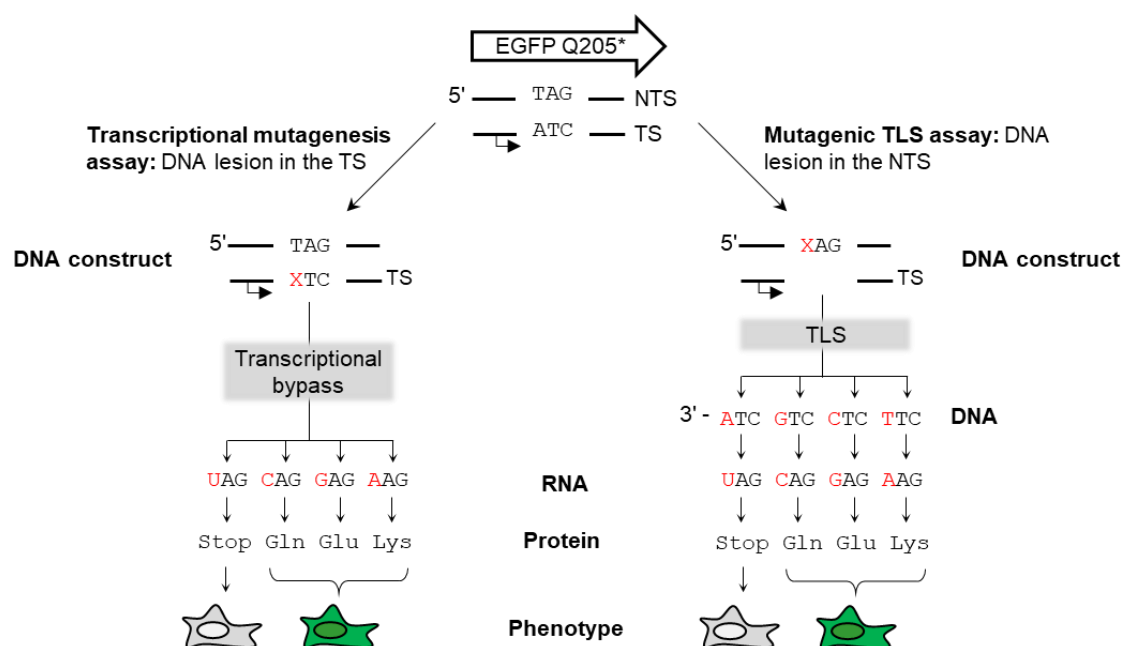


Figure 4-1: Scheme of the enhanced green fluorescent protein (EGFP) Q205* vector and its derived constructs for transcriptional mutagenesis (TM) and mutagenic translesion DNA synthesis (TLS) assays. DNA constructs are transfected into human cells, the outcomes of the bypass of a DNA lesion (X) are shown at DNA, RNA, protein and phenotypic levels. Modified from M. Rodriguez-Alvarez et al., 2020.

To test the sensitivity of the newly generated reporters, we inserted abasic DNA lesions at the mutated positions for the detection of mutagenic TLS (**section 3.3**) and TM (**section 3.2**). Then we characterised the mutagenicity of AP sites by EGFP reactivation analysis and determined their mutation profile by RNA sequencing of the transcripts derived from these cells (**section 3.4**).

4.2 Transcriptional bypass of synthetic abasic lesion leads to adenine misincorporation into mutant mRNA

By using pZAJ_Q205* and pMR_A207P EGFP mutant reporters generated in 3.1, we detected erroneous bypass of abasic DNA lesions occurring during transcription. Furthermore, we found synthetic AP sites that escape repair undergo transcriptional mutagenesis through mostly adenine incorporation in the nascent mRNA.

The major complication about studying the bypass of abasic sites during transcription is their rapid repair within the cells, which impedes the obtention of valid data prior to the modification of the lesion. Abasic DNA lesions are effectively repaired by BER, where APE1 cleaves the 5'-phosphodiester bond to the damage (Dempfle & Harrison, 1994). Alternatively, AP lyases can remove abasic sites by incising 3' to the lesion. These enzymes generate single-strand breaks (SSB) that result in ends compatible with further repair proteins (Matsumoto Y & Kim K, 1995; Sun et al., 1995). Thus, most of the data currently available about transcriptional bypass of abasic lesions has been obtained *in*

vitro where the absence of repair proteins allows the study of the process in a more direct way. It has been previously reported that transcription through abasic DNA lesions induces TM *in vitro* by using RNA polymerases from unicellular organisms (Chen & Bogenhagen, 1993; Wei Wang et al., 2018; Zhou & Doetsch, 1993) as well as from mammalian cells (Kuraoka et al., 2003). All these reports demonstrated the capacity of RNA polymerases from different organisms to bypass abasic DNA lesions; still, these experiments were highly dependent on the experimental conditions and inconsistencies exist regarding the mutagenic signature of the lesion during transcriptional bypass. Both, cytosine (Kuraoka et al., 2003) and adenine (Chen & Bogenhagen, 1993; Wei Wang et al., 2018; Zhou & Doetsch, 1993) have been reported to be incorporated opposite abasic sites during TM. Transcriptional bypass of AP sites was further confirmed *in vivo* by using unicellular organisms such as *E.Coli* (Clauson et al., 2010) and *S. cerevisiae* (Kim & Jinks-Robertson, 2010; Yu et al., 2003) where the time frame available to study these bypass events is extended in a BER impaired background. However, there is still no evidence of transcriptional mutagenesis occurring at abasic DNA lesions in human cells.

An alternative approach to avoid the repair of AP sites at a cellular level is to use the tetrahydrofuran (THF) analog of the abasic sites that, besides its inherent higher stability, is resistant to β -elimination reactions. Thus, THF constructs originated from pZAJ_Q205* or pMR_A207P and treated with AP-lyases such EndoIII and Fpg showed complete resistance to cleavage (**Figure 3-24**). Still, THF is a good substrate of AP endonucleases (Takeshita et al., 1987; Wilson III David M et al., 1995) as demonstrated by total incision of the same constructs upon incubation with APE1 (**Figure 3-15 and Figure 3-19**) and Endo IV (**Figure 3-19**). Therefore, the second challenge arises regarding the protection of AP sites against APE1-mediated repair. It was previously shown that substitution of the 5' phosphate backbone to an abasic lesion with a phosphorothioate linkage prevents its cleavage by AP endonucleases (Allgayer et al., 2016; Wilson III David M et al., 1995). Indeed, *in vitro* digestion assays showed a very efficient incision of THF upon incubation with AP endonuclease 1 (APE1), which was further avoided by sulfurization of the phosphorothioate linkage 5' to the lesion (S-THF) (**Figure 3-16**). By using THF AP sites protected with phosphorothioate linkage, we created BER-resistant constructs (S-THF) that can be transfected into human cells in order to study transcriptional mutagenesis.

The complete absence of EGFP signal obtained from cells transfected with BER-sensitive AP constructs (THF), confirmed the robust activity of BER removing this type of lesion and inserting the original nucleotide that eliminates the protein fluorescence (**Figure 3-17 and Figure 3-20**). Consistent with this idea, BER-resistant AP sites (S-

Discussion

THF) led to the reactivation of EGFP fluorescence when transfected into HeLa (**Figure 3-17 and Figure 3-20**), MRC5 and XP-A cells (**Figure 3-21**). This fluorescent phenotype indicated that transcriptional bypass of BER-resistant AP sites results in the synthesis of mutant mRNA also in human cells.

To further characterise the mutation profile of AP sites during TM, we aimed to find the nucleotide insertion pattern of the transcriptional machinery opposite to the lesion. RNA sequencing of the transcripts derived from XP-A cells showed that adenine is the most frequent nucleotide incorporated opposite synthetic abasic sites during transcriptional mutagenesis (**Figure 3-59D and Figure 3-61D**). This is in line with results obtained using alternative transcriptional polymerases from *E.Coli* and *S.cerevisiae* (Clauson et al., 2010; Wei Wang et al., 2018). However, a study using mammalian RNA pol II in transcription elongation assays reported cytosine incorporation opposite THF (Kuraoka et al., 2003). Such *in vitro* experiments are highly dependent on the dNTP pool added that might influence the mutagenic outcome of the lesion. This demonstrates that it is necessary to develop tools to assess erroneous bypass of DNA lesions in human cells. In summary, data presented in section 3.4.2 indicated that human RNA polymerase II follows the A-rule while bypassing synthetic abasic sites.

RNA sequencing results showed exclusive adenine incorporation opposite S-THF that substituted a thymine residue (**Figure 3-61D**). On the other hand, adenine and cytosine signals emerged when the lesion was placed instead of a guanosine residue (**Figure 3-59D**). If cytosine was part of the mutation profile of BER-resistant AP sites during TM, we would observe a cytosine signal regardless of the plasmid backbone used to create the construct. Besides, cytosine incorporation opposite to S-THF in position ts.617 would lead to the synthesis of a fluorescent EGFP. Since the expression analysis showed a complete lack of fluorescence (**Figure 3-61A**), we concluded that cytosine is not part of the mutation profile of synthetic AP sites during TM. Instead, the 31.5% of cytosine signal observed by transfecting ts.619G>S-THF construct might arise as a result of the repair of the lesion (**Figure 3-59D**). This would lead to a guanine residue in the template strand and therefore, a cytosine signal in the RNA sequencing analysis. In position ts.617T>S-THF the repair of the lesion would lead to a thymine residue in the template strand and a subsequent adenine signal in the transcripts. Thus, the 89.2% of adenine signal observed in this position (**Figure 3-61D**) likely emerged as a sum of the repair and the bypass of the lesion accompanied by adenine incorporation. To sum up, by comparing results obtained in both constructs, we not only demonstrated that transcriptional bypass of THF is accompanied by adenine incorporation, but also that this lesion is partially repaired within the cells.

Given that S-THF is resistant to BER and XP-A cells have impaired NER pathway, the question about which mechanism is accountable for the repair of these lesions seems unavoidable. It would be possible that the synthetic oligonucleotides contain a mixture of Rp and Sp phosphorothioate isomers which directly affects the incision activity of APE1 towards THF (20-fold for Rp isomers and >10000-fold for Sp isomers) (Wilson III David M et al., 1995). Thus, oligonucleotides containing Rp isomers of THF would still be substrates of AP endonucleases to some extent. One could also speculate that the repair of THF occurred because of the progressive decay of phosphorothioate linkages over time. Therefore, mRNA arising after transcription of accurately repaired THF templates would coexist together with mutant mRNA from the templates that have not yet been repaired. Future studies addressing this point could examine the reversal to a fluorescent EGFP phenotype in a time course. If transcriptional mutagenesis signal decreases over time, that would indicate a decay of the template containing phosphorothioate linkages.

All results discussed above in this section (4.2) were obtained using templates harboring synthetic abasic lesions (THF). However, natural AP sites created in pZAJ_Q205* mutant did not induce any EGFP reactivation by TM 24 hours after transfection, even when chemically protected from BER pathway in NER deficient cells (**Figure 3-26**). This result differed from other published reports using unicellular organisms, where they indicated that TM occurs at natural AP sites (Clauson et al., 2010; Yu et al., 2003). Interestingly, the digestion assay showed that natural AP sites protected by phosphorothioate linkages in 5' as well as in 3' (S-AP-S) were resistant to AP endonuclease and lyases activity (**Figure 3-25**). One possible explanation could be that physiological abasic sites arrest transcription by RNAPII, as previously shown in cell-free systems (Tornaletti et al., 2006). However, high levels of fluorescence were obtained when natural AP site constructs were generated from a functional EGFP vector as pZAJ_5c (**Figure 3-31**). This result indicates that the arrest of RNAPII at the site of damage might be possible at an early stage, but it got resolved (either through repair or bypass) before the 24 hours time point. Because of the high level of fluorescence detected in those specific constructs, we also conclude that the template harboring natural AP sites was successfully transfected, entering the nucleus of the cells and not degraded during the transfection process. By using a wild-type EGFP vector as a backbone for the construction of the templates, we also detected an overall lower EGFP signal in cells transfected with BER-sensitive AP sites (AP) (**Figure 3-31**). This effect has been already characterised by the group and raises because of the repair of the lesion interfering with the transcription process (Lühnsdorf et al., 2014).

Discussion

In summary, the absence of EGFP fluorescence and therefore, TM in cells transfected with natural AP sites suggests a very efficient repair of the lesion that restores the non-fluorescent character of the construct (**Figure 3-27**). The different TM outcomes generated by synthetic versus natural AP sites might be a consequence of their inherent differences during BER repair. It has been proposed that THF lesions are repaired via long-patch BER (FEN1 dependent) while natural AP sites are generally substrates of short-patch BER (Biade et al., 1998; Matsumoto Y & Kim K, 1995; Szczesny et al., 2008). Consistent with this idea, Rp isomers of S-THF that are cleaved by APE1 would lead to a 5'dRP residue that contains a sulfur instead of an oxygen atom and that may affect its subsequent processing, for instance, via FEN1 cleavage. Since natural AP sites would not need the function of FEN1 for a proper repair via short-patch BER, these AP sites would be more efficiently repaired within the cell. However, it remains to be determined whether repair of natural AP sites takes place through BER despite the presence of phosphorothioate linkages or through an alternative mechanism in a BER and NER-independent manner.

Using uracil lesion for our TM assay, we observed a small amount of transcriptional mutagenesis generated from constructs containing phosphorothioate 5' or 5' and 3' to the uracil base that was not detectable using natural AP site templates (**Figure 3-32B**). Thus, the small percentage (9-10% in XP-A) of residual fluorescence detected might arise as a result of uracil transcriptional bypass prior to its cleavage by UDGs (**Figure 3-32A**). Since this effect is only detected in constructs where the phosphodiester linkages are substituted by phosphorothioates, it seems likely that the sulfurization of the phosphodiester bond somehow hinders the function of the UDGs, delaying the entire repair process. Both guanine and adenine have been previously reported to be incorporated opposite to uracil during TM (Kuraoka et al., 2003; Viswanathan A et al., 1999). Besides, both residues would lead to a fluorescent EGFP protein in this position (**Figure 3-14**) which would explain the small percentage of fluorescence depicted by these constructs. It is worth mentioning that we did not observe significant differences between NER proficient or deficient cell lines, suggesting NER pathway is not necessary for the efficient repair of either uracil or natural AP sites in this context.

4.2.1 Nucleotide excision repair (NER) works as a backup repair mechanism for synthetic AP sites in human cells

Several studies performed in yeast provided insights that suggested the involvement of NER in the repair of AP sites placed in the transcribed strand of an active gene (Kim & Jinks-Robertson, 2010; Swanson et al., 1999; Torres-Ramos et al., 2000). To investigate this matter in a human cell environment, we examined the accumulation of mutant

transcripts after transcriptional bypass of templates carrying BER-resistant AP sites in NER deficient cell lines. Here, the relative EGFP fluorescence obtained was inversely proportional to the repair of the lesion in each specific genetic background. Deficiency in either GG-NER (XP-C) or TC-NER (CS-A and CS-B) was sufficient to enhance TM over synthetic AP sites, 4-fold and 2-fold respectively (**Figure 3-22**). Interestingly, complete NER deficiency in XP-A cells resulted in a tremendous reversal (8-fold) to a fluorescent EGFP when compared to repair proficient MRC5 cells (**Figure 3-21**). This result indicated that the enhanced fluorescent phenotype was directly linked to NER deficiency. Besides, the further complementation of these cells with XPA protein reverted the mutagenic phenotype by half (**Figure 3-21**). This partial reversal might not be complete due to the limitation of the experiment, where the levels of XPA protein synthesized from the exogenous plasmid are not expected to fully compensate for the XPA levels of a healthy cell. Still, this result clearly demonstrates the implication of XP-A and to a broader extent, the NER pathway, in the repair of THF AP sites in human cells. Nevertheless, the specific NER subpathway involved in the repair of THF remains uncertain since an impairment of GG-NER and to a lesser extent, TC-NER showed significant alterations of TM levels. Because the cell models used in this experiment were not isogenic, it might be possible that the differences in their genetic background prevent us from obtaining a more definitive result. It would be interesting to create CSA and XPC KOs in the MRC-5 background to repeat the assay and hopefully determine the NER subpathway responsible for THF repair. Additional support for the involvement of NER in the repair of synthetic abasic DNA lesions has been obtained within the group using an alternative gene reactivation principle (Kitsera et al., 2019). Thus, the results presented here represent the first proof of NER working as a backup mechanism for the repair of synthetic AP sites in human cells.

In summary, we demonstrated the usefulness of the reporters generated in 3.1 to study transcriptional mutagenesis of abasic sites in human cells. Thus, it was evident that the use of synthetic abasic lesions as THF did not mirror the transcriptional outcome of natural abasic sites. We showed that THF AP sites that escape BER and NER repair, undergo transcriptional bypass through mostly adenine incorporation in the nascent mRNA. Furthermore, by using the EGFP mutant plasmids as a tool to study repair, we have provided novel proof of the involvement of NER in the repair of synthetic AP lesions (**Figure 4-2**).

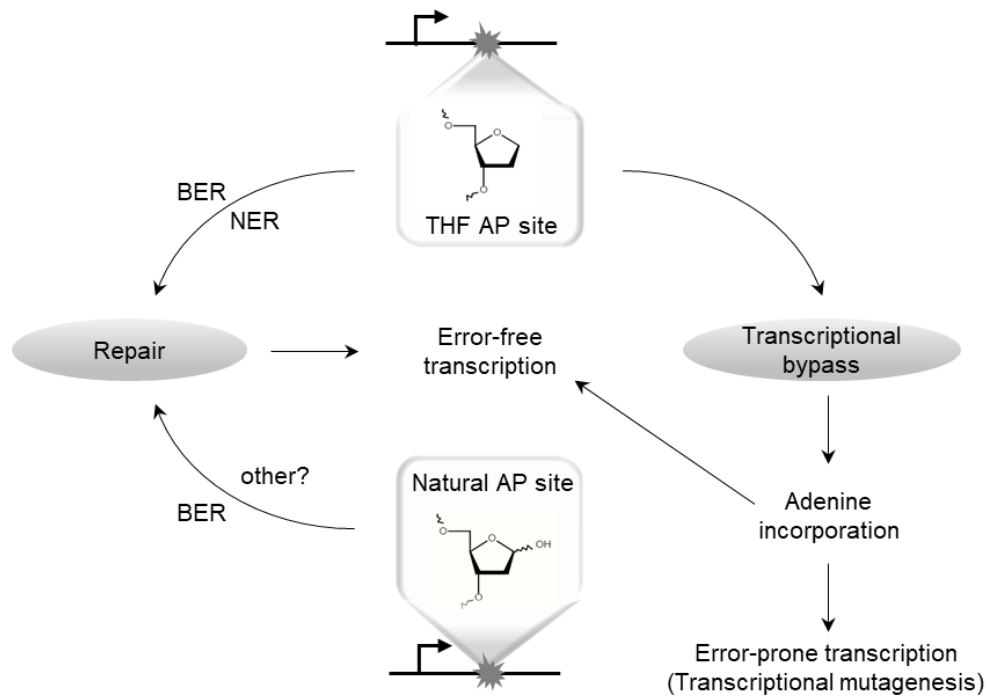


Figure 4-2: Model of transcriptional outcomes of abasic DNA lesions placed in the template strand of a gene. BER: Base excision repair; NER: Nucleotide excision repair; AP: apurinic/aprimidinic site.

4.3 Translesion synthesis over abasic lesions mostly leads to adenine incorporation; however, mutation profiles differ between synthetic and natural AP sites

By using all EGFP mutant reporters generated in 3.1, we showed that mutagenic translesion synthesis (TLS) occurs at synthetic and natural abasic DNA lesions though mostly adenine incorporation. Furthermore, we characterized the mutation profile of both types of lesions. While A>G>C>T nucleotides are incorporated opposite natural AP sites during TLS, synthetic abasic lesion (THF) showed a strong decrease in guanine incorporation: A>C>T>G.

The main adversity when studying replicational bypass over abasic DNA lesions is the absence of a proper system to address TLS in human cells in a direct manner. The vast majority of data available in the literature about TLS of AP sites is obtained via primer extension assays or gap-directed TLS with human purified proteins *in vitro*. As a result, several replicative DNA polymerases have been reported to bypass abasic DNA lesions (Beard et al., 2009; Choi et al., 2010; Villani et al., 2014). In addition, TLS polymerases such as DNA pol η (Choi et al., 2010; Patra et al., 2015; Villani et al., 2014), DNA pol ι (Choi et al., 2010; Nair et al., 2009; Zhang et al., 2001), DNA pol κ (Choi et al., 2010) and Rev1 (Choi et al., 2010; Lin et al., 1999; Nair et al., 2011) were also shown to be

capable of bypassing AP sites *in vitro*. These experiments demonstrate the capacity of a specific human DNA polymerase to bypass this type of lesions; however, they are performed by using one polymerase at a time. Consequently, the capacity of these proteins to bypass the damage does not indicate its function under physiological conditions where all polymerases coexist inside the cell. Several *in vivo* studies using *S. cerevisiae* as a model organism reported translesion synthesis occurring at synthetic (Otsuka et al., 2002), as well as natural AP sites (Chan et al., 2013; Gibbs & Lawrence, 1995). All reported adenine or cytosine as the preferred nucleotide incorporated opposite to the lesion during TLS and two of them (Chan et al., 2013; Otsuka et al., 2002) showed the presence of Rev1 as a requirement for this cytosine incorporation. Even though experiments in yeast provided valuable data about TLS of AP sites in eukaryotic cells, this organism favors homology-dependent over the translesion synthesis (TLS) mechanisms more commonly used in mammalian cells (Quinet et al., 2018; Swanson et al., 1999). Besides, higher organisms possess multiple TLS enzymes with characteristic efficiency and fidelity profiles that are not present in lower eukaryotes (Choy JY 2010, Shibutani S 2004). As a result, techniques based on vectors harboring abasic lesions transfected into human cells emerged as a promising tool to study the replicational bypass of this lesion (Avkin et al., 2002; Shachar et al., 2009; Weerasooriya et al., 2014). By using these techniques, several groups demonstrated that bypass of synthetic AP lesions occurs mainly through adenine incorporation. However, TLS reporters described in the literature also require shuttling into bacteria in order to quantify TLS outcomes. Thus, a reporter system for the direct detection of mutagenic TLS in mammalian cells confers a benefit to this research. Furthermore, it brings the opportunity to study the bypass of natural AP sites in human cells, which data is currently missing in the literature.

The mutant EGFP reporters generated in 3.1 worked as an excellent tool to study mutagenic TLS over abasic DNA lesions via reporter reactivation assays. For detection of mutagenic TLS occurring directly at the damaged template, it was necessary to place an abasic DNA lesion instead of the mutated nucleotide in the coding strand and opposite to a gap. The gap must be filled by DNA synthesis within the cells and only by erroneous bypass of the lesion, the fluorescence of the EGFP would be reactivated (**Figure 3-35A**). Thus, we detected mutagenic TLS occurring at synthetic and natural AP lesions regardless of the mutant reporter and the cell line used (**Figure 3-47**, **Figure 3-48** and **Figure 3-50**). Additionally, we have demonstrated that both types of AP sites led to erroneous DNA synthesis mostly through adenine incorporation (**Figure 3-63** and **Figure 3-66**). Regarding THF AP sites, this result agrees with previous reports demonstrating the preferential incorporation of dAMP opposite to this lesion (Avkin et al., 2002; Shachar

Discussion

et al., 2009; Weerasooriya et al., 2014). Interestingly, incorporation of adenine opposite natural AP sites has never been reported in human cells. Because most abasic lesions emerge from spontaneous depurination (Loeb & Preston, 1986; Schaaper et al., 1983) and cytosine deaminations (Krokan et al., 2002), the incorporation of adenine opposite to these lesions makes them extremely mutagenic.

Replicative polymerases are well known for their strong preference for adenine incorporation opposite to AP sites, a mechanism termed “the A-rule” (Lawrence et al., 1990; Mozzherin et al., 1997; Shibutani et al., 1997). B- and X-family polymerases such as pol α , pol δ (Choi et al., 2010) and pol β (Villani et al., 2014) have been reported to bypass synthetic abasic DNA lesions *in vitro*. In addition, upon treatment with a specific inhibitor of B-family DNA polymerases, strong inhibition of THF bypass has been observed (Avkin et al., 2002). Since none of the studies using human TLS polymerases in primer extension assays have reported the incorporation of adenine opposite to this lesion, we conclude that replicative polymerases are mainly responsible for it. Therefore, it seems logical that higher organisms possess multiple TLS enzymes in order to prevent replicative DNA polymerases to induce mutagenicity while bypassing these lesions.

Even though adenine prevails as the preferred nucleotide incorporated opposite AP sites in this work (**Figure 3-63 and Figure 3-66**), regain of EGFP fluorescence transfecting Q205* mutant (where C, G or T incorporation reverts to a fluorescent protein) certainly demonstrates that, at least, one of the other three alternative nucleotides is also incorporated opposite to the lesion (**Figure 3-39 and Figure 3-51**). In this position, FACS analysis showed higher mutagenesis of natural AP sites compared to their broadly used synthetic analogue tetrahydrofuran. Likewise, HAP cells transfected with THF in position c.617* generated very low fluorescence (**Figure 3-48**) compared to samples transfected with uracil substrate in the same position (**Figure 3-50**). It is important to note that constructs generated from this pMR_S206* (c.617*G) mutant vector, only show reactivation of the fluorescence upon guanine incorporation opposite to the lesion (**Figure 3-10**). Thus, this result suggests that guanine incorporation is more common opposite natural AP sites than opposite THF. The RNA sequencing analysis clearly confirmed the disparities between the mutagenic footprint of natural AP sites (A>G>C>T) and synthetic THF AP sites (A>C>T>G) (**Figure 3-62, Figure 3-63 and Figure 3-66**). Differences that have been previously observed in yeast transfected with oligonucleotides carrying THF as well as natural AP sites (Otsuka et al., 2002).

If we consider that adenine incorporation opposite to AP sites is exclusively performed by replicative polymerases, guanine, cytosine and/or thymine incorporation would result

from a bypass performed by specialized TLS polymerases. Rad18 is considered an essential protein for the activation of the translesion synthesis (TLS) mechanism through PCNA monoubiquitylation (Haracska et al., 2002; Watanabe et al., 2004; Wilkinson et al., 2020). Indeed, we observed a reduction of the EGFP fluorescence in c.613THF and c.613U in absence of Rad18 compared to WT cells (**Figure 3-52**) suggesting that the incorporation of dG, dT and dC in position 613 was reduced in Rad18 KO cells. The role of Rad18 inducing mutagenic bypass of other DNA lesions (UV or MMS induced) was previously demonstrated by Tateishi et al., 2003 in mouse embryonic fibroblasts (MEFs). Interestingly, results in position 619 where the incorporation of adenine results in a fluorescent EGFP did not show any significant decrease of fluorescence in Rad18 KO cells compared to HAP (**Figure 3-53**). Since the incorporation of adenine by replicative polymerases is not expected to be reduced in absence of Rad18, it seems that the overall bypass of the lesion is not affected in these cells. Instead, a reduction in dG, dT and dC incorporation opposite to the lesion might be compensated by incorporation of adenine through replicative polymerases. An observation confirmed by the RNA sequencing data where the incorporation of adenine opposite to the lesion increases in absence of Rad18 (**Supplementary figure 3**). It is important to note that lack of Rad18 reduced the mutagenic bypass of abasic lesions by almost half in position 613 (**Figure 3-52**); however, 13.4 and 33.5% of fluorescence were remaining for THF and U substrates, respectively. This is not the first study that showed mutagenic TLS occurring in absence of Rad18 protein; by transfecting plasmids harboring other DNA lesions into MEFs Hendel et al., 2011 showed that 25% of TLS remained in absence of Rad18. It seems clear that, although not essential, Rad18 is required for maximal TLS. Still, whether residual TLS arises because PCNA monoubiquitination is not essential for TLS as previously suggested by other groups (Gervai et al., 2017; Hendel et al., 2011) or because an alternative E3-ubiquitin ligase modifies PCNA protein (Simpson et al., 2006) demands further investigation. In summary, these results suggest that incorporation of cytosine, thymine and guanine opposite AP lesions is reduced in absence of Rad18.

Because of the differences observed between THF and natural AP sites, we will now address the discussion about the mutation profile of these lesions separately. By using all EGFP mutants to generate constructs harboring THF, we detected the highest fluorescent phenotype in A207P (c.619C>THF) followed by Q205P (c.614C>THF) (**Figure 3-47 and Figure 3-48**). Both mutant reporters maintain their non-fluorescent status only if incorporation of guanine occurs opposite to the mutated site (**Table 3-2**). Conversely, by using S206* mutant (c.617*THF) where only guanine incorporation opposite THF reactivates the protein fluorescence, we observed an almost complete lack

Discussion

of fluorescence. These results suggested that the incorporation of guanine opposite THF is rather unusual, as further confirmed by the sequencing analysis that showed only 1.9% of guanine incorporation (**Figure 3-63**). As opposed to results published by *Otsuka 2002 et al.* where the nucleotide positioned 5' to the THF influenced the mutation profile of the lesion, our sequencing results showed that regardless of the nucleotide placed 5' to the THF AP site (C for Q205* and T for S206*) adenine was invariably the most frequent nucleotide incorporated opposite to THF. This preference for adenine incorporation opposite to THF AP sites has been previously reported by other studies (Avkin et al., 2002; Shachar et al., 2009; Weerasooriya et al., 2014). Given the assumption that replicative polymerases account for 67.6% of adenine incorporation opposite THF, it is likely that the incorporation of the other nucleobases arose as a consequence of damage bypass by TLS polymerases (**Figure 3-63**).

To study which TLS polymerase is accountable for the incorporation of cytosine, thymine and guanine opposite THF, we transfected constructs derived from the pZAJ_Q205* mutant vector into TLS polymerases knock-out (KO) cell lines. By using this reporter, incorporation of adenine opposite to the lesion becomes undetectable and the EGFP fluorescence obtained is the result of cytosine, guanine, and/or thymine incorporation opposite to the lesion. EGFP fluorescence did not vary in absence of pol η and pol κ compared to WT cells, suggesting that these polymerases are not critical for THF bypass. Indeed, primer extension assays showed that human Pol κ was strongly stalled by THF lesions (Choi et al., 2010) and gapped plasmids carrying THF transfected in pol η deficient cells (XP-V) did not show a decrease in the overall TLS (Avkin et al., 2002). In this context, we detected a reduction of EGFP fluorescence in cells depleted of Pol ι and Rev1 compared to HAP which implies a potential role of both polymerases in the mutagenic bypass of THF AP sites (**Figure 3-54**). In fact, *in vitro* experiments have demonstrated the ability of both Pol ι and Rev 1 to bypass THF by thymine and cytosine incorporation, respectively (Choi et al., 2010; Nair et al., 2011). Considering that 23.6% of cytosine incorporation was detected opposite THF (**Figure 3-63**), Rev1 appeared as a strong candidate that might be responsible for THF bypass in human cells. Thus, it would be interesting to repeat transfections into Rev1 KO cells to increase the statistical power of the results. It is also important to consider that redundancy between TLS polymerases to bypass abasic lesions (**Table 1-1**) might affect the quality of the results by only analyzing the protein fluorescence. Therefore, to unravel which specific TLS polymerase is accountable for the mutagenic bypass of THF, future experiments should analyze transcripts derived from the transfection of TLS pol KO cell lines.

RNA sequencing results of cells transfected with templates carrying natural AP sites indicated that adenine is the preferred nucleotide incorporated opposite to this lesion with a 43.5% (**Figure 3-66**). This is in agreement with the few studies that investigated the erroneous bypass of natural AP sites by replicative DNA polymerases *in vitro* (Choi et al., 2010; Sagher & Strauss, 1983) and in lower eukaryotes (Chan et al., 2013). Conversely, cytosine incorporation was also reported opposite natural AP sites using *S. cerevisiae* as a model organism (Gibbs & Lawrence, 1995; Otsuka et al., 2002). Thus, conclusions regarding the choice of nucleotide incorporated opposite to the lesion in lower eukaryotes remain uncertain. In this work, we used templates containing uracil, that after its excision within the cells forms a natural AP site, and glycosylase-resistant uracil (FU) to investigate the mutagenic character of natural abasic sites for the first time in human cells. Unlike THF where only replicative polymerases were reported to incorporate adenine opposite to the lesion, primer extension assays using natural AP sites have shown that human pol η also inserts adenine, besides guanine, opposite to this lesion (Patra et al., 2015). In this connection, if pol η would be partially responsible for AP sites bypass, cells deficient in pol η recruitment, as Rad18 KO, should show a decrease in adenine incorporation. Instead, adenine signal opposite natural AP sites seemed to increase in absence of Rad18 compared to WT cells (**Supplementary figure 3**); however, statistical analysis of these results was not significant. This was likely because of the high sensitivity showed by Rad18 KO cells towards the transfection where the recovery of transfected cells that remained alive was problematic. An alternative possibility would be to analyze transcripts from pol η KO cells transfected with natural AP sites and treated with aphidicolin, a well-established inhibitor of the replicative polymerases (Avkin et al., 2002). Thus, if replicative polymerases are responsible for the adenine incorporation, treatment with the inhibitor should abrogate the adenine signal in treated cells.

Even though adenine incorporation opposite natural AP sites occurs with the highest frequency (42%) (**Figure 3-66**), this signal is lower than with THF (67.6%) (**Figure 3-63**). Because it has been shown that physiological AP sites are natural blocks for the replication machinery (Goodman, 2002; Hübscher et al., 2002), this result suggests that the majority of the natural AP site bypass occurs via TLS polymerases. Besides adenine, the sequencing experiment clearly showed incorporation of guanine (23.5%), cytosine (20%) and thymine (10.9%) opposite to natural AP sites (**Figure 3-66**). Guanine incorporation opposite to this lesion was previously described by Patra et al., 2015 using primer extension assays with human pol η . In yeast, it has been shown that cytosine was part of the mutation profile of natural AP sites in a Rev1 dependent manner (Chan et al.,

Discussion

2013; Otsuka et al., 2002), also confirmed *in vitro* by using human Rev1 (Lin et al., 1999). Therefore, it would be interesting to repeat the sequencing analysis in Pol η and Rev1 KO cells transfected with uracil templates to decipher their role in TLS of natural AP sites.

Unlike the phenotypic results obtained for THF where positions 614 and 619 depicted the highest fluorescence with negligible differences (**Figure 3-48**), the same experiment using natural AP lesions showed an overwhelming reactivation of the EGFP fluorescence in position 619, 2.5x higher than in position 614 (**Figure 3-50**). In both positions, incorporation of adenine, cytosine, or thymine reactivates the protein fluorescence, yet the incorporation of adenine reactivates the protein fluorescence to an estimate of 78% in position 614 while 100% was obtained in position 619 (**Figure 3-9**). Still, this variation between both mutant reporters does not fully explain a 2.5x difference in the TLS assay. The fact that the sequence context was shown to strongly influence the frequency of nucleotide incorporation opposite natural AP sites *in vitro* (Patra et al., 2015) and *in vivo* (Chan et al., 2013) might explain such a difference in our results. However, it is important to note that incorporation of adenine, cytosine and thymine in position ts.619 (pMR_A207P) resulted in mutations whose fluorescence intensity exceeded the level of EGFP fluorescence of the WT EGFP protein (**Figure 3-9**). Therefore, a comparison between different phenotypes at 614 and 619 positions might not suffice to conclude if the sequence context affects the mutagenic profile of the lesion. Additional RNA sequencing analysis of these constructs would be necessary in order to clarify this point.

Altogether, this work presented a set of mutant EGFP reporters that can be used to directly detect mutagenic bypass of AP lesions during DNA synthesis in human cells. By using this approach, we demonstrated that adenine is the most frequent nucleotide incorporated opposite both THF and natural AP sites, most likely by replicative polymerases. Being guanine one of the most commonly lost bases in the genome, adenine insertion by DNA polymerases illustrates the huge mutagenic potential of abasic DNA lesions in human cells. In addition to adenine incorporation, we showed that mutation profiles of natural and synthetic abasic sites are different. While the incorporation of guanine, closely followed by cytosine, and lastly thymine was detected opposite natural AP sites; cytosine led the mutation profile of THF, followed by thymine and a small percentage of guanine (**Figure 4-3**). These results, together with the fact that each polymerase has a characteristic base selectivity opposite AP sites, suggest that bypass of natural and synthetic abasic lesions might occur through different TLS polymerases. Therefore, even though the use of THF AP sites has greatly contributed to the understanding of abasic lesion repair and bypass, differences regarding the

mutagenicity of the lesion observed in this study inevitably question if the use of this synthetic analog is the most accurate alternative to study the mutagenicity of abasic sites in human cells.

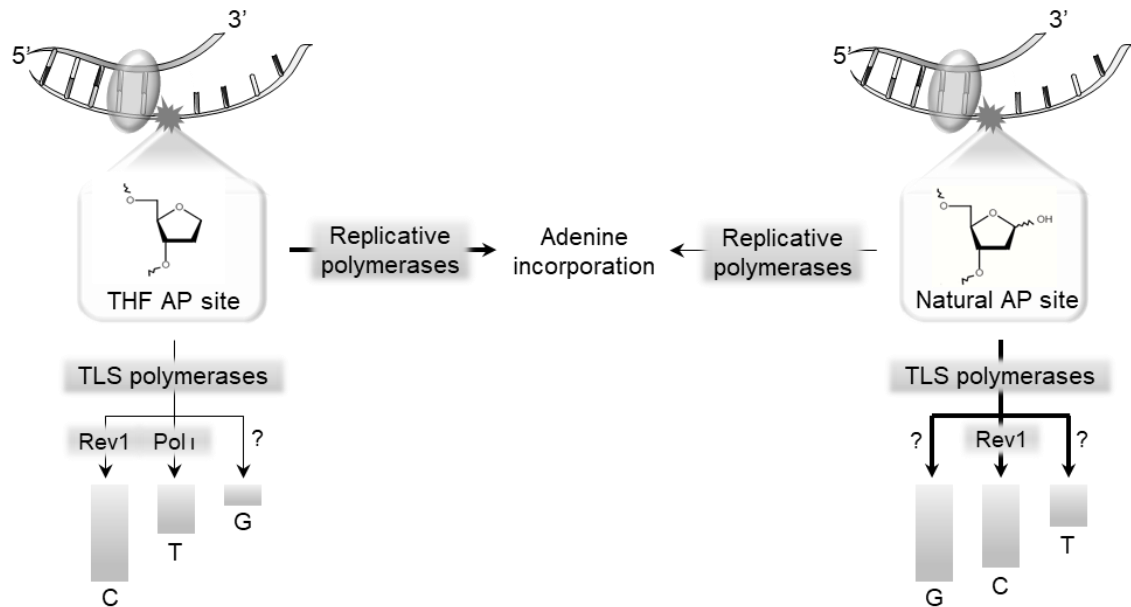


Figure 4-3: Model illustrating the bypass of abasic DNA lesions during DNA synthesis in human cells. Thicker black arrows represent the most common bypass expected for each type of abasic lesion. AP: apurinic/aprimidinic site; pol: polymerases; C: cytosine; G: guanine; T: thymine.

References

- Allgayer, J., Kitsera, N., Bartelt, S., Epe, B., & Khobta, A. (2016). Widespread transcriptional gene inactivation initiated by a repair intermediate of 8-oxoguanine. *Nucleic Acids Research*, *44*(15), 7267–7280. <https://doi.org/10.1093/nar/gkw473>
- Anjana, R., Kirti Vaishnavi, M., Sherlin, D., Pavan Kumar, S., Naveen, K., Sandeep Kanth, P., & Sekar, K. (2012). Hypothesis Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: a study based on orientation and distance. *Print) Bioinformatics*, *8*(24), 1220. www.bioinformatics.net
- Avkin, S., Adar, S., Blander, G., & Livneh, Z. (2002). Quantitative measurement of translesion replication in human cells: Evidence for bypass of abasic sites by a replicative DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(6), 3764–3769. <https://doi.org/10.1073/pnas.062038699>
- Batista, L. F. Z., Kaina, B., Meneghini, R., & Menck, C. F. M. (2009). How DNA lesions are turned into powerful killing structures: Insights from UV-induced apoptosis. In *Mutation Research - Reviews in Mutation Research* (Vol. 681, Issues 2–3, pp. 197–208). <https://doi.org/10.1016/j.mrrev.2008.09.001>
- Beard, W. A., Shock, D. D., Batra, V. K., Pedersen, L. C., & Wilson, S. H. (2009). DNA polymerase β substrate specificity: Side chain modulation of the “a-rule.” *Journal of Biological Chemistry*, *284*(46), 31680–31689. <https://doi.org/10.1074/jbc.M109.029843>
- Beard, W. A., & Wilson, S. H. (2014). Structure and mechanism of DNA polymerase β . *Biochemistry*, *53*(17), 2768–2780. <https://doi.org/10.1021/bi500139h>
- Bennett, R. A. O. (1999). The *Saccharomyces cerevisiae* ETH1 Gene, an Inducible Homolog of Exonuclease III That Provides Resistance to DNA-Damaging Agents and Limits Spontaneous Mutagenesis. *MOLECULAR AND CELLULAR BIOLOGY*, *19*(3), 1800–1809. www.kegg.com
- Biade, S., Sobol, R. W., Wilson, S. H., & Matsumoto, Y. (1998). Impairment of proliferating cell nuclear antigen-dependent apurinic/aprimidinic site repair on linear DNA. *Journal of Biological Chemistry*, *273*(2), 898–902. <https://doi.org/10.1074/jbc.273.2.898>
- Boiteux, S., & Guillet, M. (2004). Abasic sites in DNA: Repair and biological consequences in *Saccharomyces cerevisiae*. In *DNA Repair*. <https://doi.org/10.1016/j.dnarep.2003.10.002>
- Bré Geon, D., Doddrige, Z. A., You, H. J., Weiss, B., & Doetsch, P. W. (2003). Transcriptional Mutagenesis Induced by Uracil and 8-Oxoguanine in *Escherichia coli* dividing cells in artificial environments based on growth-rich media. However, cells living outside of a laboratory environment are not frequently engaged in division and. In *Molecular Cell* (Vol. 12).
- Brégeon, D., & Doetsch, P. W. (2011). Transcriptional mutagenesis: Causes and involvement in tumour development. In *Nature Reviews Cancer* (Vol. 11, Issue 3, pp. 218–227). Nature Publishing Group. <https://doi.org/10.1038/nrc3006>
- Brejč, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormö, M., & Remington, S. J. (1997). Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(6), 2306–2311.

<https://doi.org/10.1073/pnas.94.6.2306>

- Brřgeon, D., & Doetsch, P. W. (2011). Transcriptional mutagenesis: Causes and involvement in tumour development. In *Nature Reviews Cancer* (Vol. 11, Issue 3, pp. 218–227). Nature Publishing Group. <https://doi.org/10.1038/nrc3006>
- Byun, T. S., Pacek, M., Yee, M. C., Walter, J. C., & Cimprich, K. A. (2005). Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes and Development*, 19(9), 1040–1052. <https://doi.org/10.1101/gad.1301205>
- Caldecott, K. W. (2020). Mammalian DNA base excision repair: Dancing in the moonlight. In *DNA Repair* (Vol. 93). Elsevier B.V. <https://doi.org/10.1016/j.dnarep.2020.102921>
- Chan, K., Resnick, M. A., & Gordenin, D. A. (2013). The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis. *DNA Repair*. <https://doi.org/10.1016/j.dnarep.2013.07.008>
- Chang, D. J., Lupardus, P. J., & Cimprich, K. A. (2006). Monoubiquitination of proliferating cell nuclear antigen induced by stalled replication requires uncoupling of DNA polymerase and mini-chromosome maintenance helicase activities. *Journal of Biological Chemistry*, 281(43), 32081–32088. <https://doi.org/10.1074/jbc.M606799200>
- Chen, Y. H., & Bogenhagen, D. F. (1993). Effects of DNA lesions on transcription elongation by T7 RNA polymerase. *Journal of Biological Chemistry*, 268(8), 5849–5855. [https://doi.org/10.1016/s0021-9258\(18\)53397-4](https://doi.org/10.1016/s0021-9258(18)53397-4)
- Choi, J. Y., Lim, S., Kim, E. J., Jo, A., & Guengerich, F. P. (2010). Translesion Synthesis across Abasic Lesions by Human B-Family and Y-Family DNA Polymerases α , δ , η , ι , κ , and REV1. *Journal of Molecular Biology*, 404(1), 34–44. <https://doi.org/10.1016/j.jmb.2010.09.015>
- Clauson, C. L., Oestreich, K. J., Austin, J. W., & Doetsch, P. W. (2010). Abasic sites and strand breaks in DNA cause transcriptional mutagenesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, 107(8), 3657–3662. <https://doi.org/10.1073/pnas.0913191107>
- Copeland, W. C., & Wang, T. S. (1993). Enzymatic Characterization of the Individual Mammalian Primase Subunits Reveals a Biphasic Mechanism for Initiation of DNA Replication*. *The Journal of Biological Chemistry*, 268(35), 26179–26189.
- Cormack, B. P., Valdivia, R. H., & Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP) (GFP mutation; FITC; fluorescence-activated cell sorter; fluorescence intensity). *Gene*, 173, 33–38.
- Daigaku, Y., Davies, A. A., & Ulrich, H. D. (2010). Ubiquitin-dependent DNA damage bypass is separable from genome replication. *Nature*, 465(7300), 951–955. <https://doi.org/10.1038/nature09097>
- de Vries, A., Th van Oostromt, C. M., A Hofhuis, F. M., Dortant, P. M., W Bergt, R. J., de GrulJlt, F. R., Wester, P. W., van KrelJlt, C. F., A Capel, P. J., van Steegt, H. I., & Verbeek, S. J. (1995). Increased susceptibility to ultraviolet-a and carcinogens of mice lacking the DNA excision repair gene XPA. *Nature*, 377, 169–173.
- Dean, W., & Howard-Flanders, P. (1968). Discontinuities in the DNA synthesized in an Excision-defective Strain of *Escherichia coli* following Ultraviolet Irradiation. *J. Mol. Biol.*, 31, 291–304.

References

- Demple, B., & Harrison, L. (1994). REPAIR OF OXIDATIVE DAMAGE TO DNA: Enzymology and Biology. *Annual Review of Biochemistry*, 63(1), 915–948. <https://doi.org/10.1146/annurev.bi.63.070194.004411>
- Diaz, M., Watson, N. B., Turkington, G., Verkoczy, L. K., Klinman, N. R., & Mcgregor, W. G. (2003). Decreased Frequency and Highly Aberrant Spectrum of Ultraviolet-Induced Mutations in the hprt Gene of Mouse Fibroblasts Expressing Antisense RNA to DNA Polymerase β . *Molecular Cancer Research*, 1, 836–847.
- Doetsch, P. W. (2002). Translesion synthesis by RNA polymerases: occurrence and biological implications for transcriptional mutagenesis. In *Mutation Research* (Vol. 510).
- Dutta, S., Chowdhury, G., & Gates, K. S. (2007). Interstrand cross-links generated by abasic sites in duplex DNA. *Journal of the American Chemical Society*, 129(7), 1852–1853. <https://doi.org/10.1021/ja067294u>
- Ezerskyte, M., Paredes, J. A., Malvezzi, S., Burns, J. A., Margison, G. P., Olsson, M., Scicchitano, D. A., & Dreij, K. (2018). O6-methylguanine-induced transcriptional mutagenesis reduces p53 tumor-suppressor function. *Proceedings of the National Academy of Sciences of the United States of America*, 115(18), 4731–4736. <https://doi.org/10.1073/pnas.1721764115>
- Fan, J., Matsumoto, Y., & Wilson, D. M. (2006). Nucleotide sequence and DNA secondary structure, as well as replication protein A, modulate the single-stranded abasic endonuclease activity of APE1. *Journal of Biological Chemistry*, 281(7), 3889–3898. <https://doi.org/10.1074/jbc.M511004200>
- Fousteri, M., Vermeulen, W., van Zeeland, A. A., & Mullenders, L. H. F. (2006). Cockayne Syndrome A and B Proteins Differentially Regulate Recruitment of Chromatin Remodeling and Repair Factors to Stalled RNA Polymerase II In Vivo. *Molecular Cell*, 23(4), 471–482. <https://doi.org/10.1016/j.molcel.2006.06.029>
- Friedberg, E. C. (2005). Suffering in silence: The tolerance of DNA damage. In *Nature Reviews Molecular Cell Biology* (Vol. 6, Issue 12, pp. 943–953). <https://doi.org/10.1038/nrm1781>
- Fung, H., & Demple, B. (2005). A vital role for Ape1/Ref1 protein in repairing spontaneous DNA damage in human cells. *Molecular Cell*, 17(3), 463–470. <https://doi.org/10.1016/j.molcel.2004.12.029>
- Gelfand, C. A., Eric Plum, G., Grollman, A. P., Johnson, F., & Breslauer, K. J. (1998). Thermodynamic Consequences of an Abasic Lesion in Duplex DNA Are Strongly Dependent on Base Sequence †. *Biochemistry*, 37, 7321–7327. <https://pubs.acs.org/sharingguidelines>
- Gervai, J. Z., Gálicza, J., Szeltner, Z., Záborszky, J., & Szüts, D. (2017). A genetic study based on PCNA-ubiquitin fusions reveals no requirement for PCNA polyubiquitylation in DNA damage tolerance. *DNA Repair*, 54, 46–54. <https://doi.org/10.1016/j.dnarep.2017.04.003>
- Gibbs, P. E. M., & Lawrence, C. W. (1995). Novel mutagenic properties of abasic sites in *Saccharomyces cerevisiae*. *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.1995.0430>
- Goodman, M. F. (2002). Error-prone repair DNA polymerases in prokaryotes and eukaryotes. In *Annual Review of Biochemistry* (Vol. 71, pp. 17–50). <https://doi.org/10.1146/annurev.biochem.71.083101.124707>
- Guillet Marie, & Boiteux Serge. (2002). Endogenous DNA abasic sites cause cell death

- in the absence of Apn1 Apn2 and Rad1-Rad10 in *Saccharomyces cerevisiae*_2002_Guillet M and Boiteux S. *The EMBO Journal*, 21(11), 2833–2841.
- Guo Caixia, Fischhaber PL, Luk-Paszyc MJ, Masuda Y, Zhou J, Kamiya K, Kisker C, & Friedberg EC. (2003). Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis. *The EMBO Journal*, 22(24), 6621–6630.
- Haracska, L., Unk, I., Johnson, R. E., Phillips, B. B., Hurwitz, J., Prakash, L., & Prakash, S. (2002). Stimulation of DNA Synthesis Activity of Human DNA Polymerase κ by PCNA. *Molecular and Cellular Biology*, 22(3), 784–791. <https://doi.org/10.1128/mcb.22.3.784-791.2002>
- Hendel, A., Krijger, P. H. L., Diamant, N., Goren, Z., Langerak, P., Kim, J., Reißner, T., Lee, K. young, Geacintov, N. E., Carell, T., Myung, K., Tateishi, S., D'Andrea, A., Jacobs, H., & Livneh, Z. (2011). PCNA ubiquitination is important, but not essential for translesion DNA synthesis in mammalian cells. *PLoS Genetics*, 7(9). <https://doi.org/10.1371/journal.pgen.1002262>
- Hoegge, C., Pfander, B., Moldovan, G.-L., Pyrowolakis, G., & Jentsch, S. (2002). *RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO*. www.nature.com/nature
- Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleš, A. K., ... Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, 12(2), 115–121. <https://doi.org/10.1038/nmeth.3252>
- Hübscher, U., Maga, G., & Spadari, S. (2002). Eukaryotic DNA polymerases. In *Annual Review of Biochemistry* (Vol. 71, pp. 133–163). <https://doi.org/10.1146/annurev.biochem.71.090501.150041>
- Imashimizu, M., Oshima, T., Lubkowska, L., & Kashlev, M. (2013). Direct assessment of transcription fidelity by high-resolution RNA sequencing. *Nucleic Acids Research*, 41(19), 9090–9104. <https://doi.org/10.1093/nar/gkt698>
- Iyama, T., & Wilson, D. M. (2013). DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair*, 12(8), 620–636. <https://doi.org/10.1016/j.dnarep.2013.04.015>
- Jacobs, A. L., & Schär, P. (2012). DNA glycosylases: In DNA repair and beyond. In *Chromosoma* (Vol. 121, Issue 1, pp. 1–20). <https://doi.org/10.1007/s00412-011-0347-4>
- Johnson RE, Kondratick CM, Satya Prakash, & Louise Prakash. (1999). hRAD30 Mutations in the variant form of Xeroderma Pigmentosum. *Science*, 285, 263–265.
- Johnson RE, Washington MT, Haracska L, Prakash S, & Prakash L. (2000). Eukaryotic polymerases ι and ζ act sequentially to bypass DNA lesions_2000_Nature, Johnson RE. *Nature*, 406, 1015–1019.
- Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P. A., Hagen, L., Krokan, H. E., & Slupphaug, G. (2002). hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. *Journal of Biological Chemistry*, 277(42), 39926–39936. <https://doi.org/10.1074/jbc.M207107200>
- Khobta, A., Anderhub, S., Kitsera, N., & Epe, B. (2010). Gene silencing induced by

References

- oxidative DNA base damage: Association with local decrease of histone H4 acetylation in the promoter region. *Nucleic Acids Research*.
<https://doi.org/10.1093/nar/gkq170>
- Kim, N., & Jinks-Robertson, S. (2010). Abasic Sites in the Transcribed Strand of Yeast DNA Are Removed by Transcription-Coupled Nucleotide Excision Repair. *Molecular and Cellular Biology*, 30(13), 3206–3215.
<https://doi.org/10.1128/mcb.00308-10>
- Kitsera, N., Rodriguez-Alvarez, M., Emmert, S., Carell, T., & Khobta, A. (2019). Nucleotide excision repair of abasic DNA lesions. *Nucleic Acids Research*, 47(16), 8537–8547. <https://doi.org/10.1093/nar/gkz558>
- Kitsera, N., Stathis, D., Lühnsdorf, B., Müller, H., Carell, T., Epe, B., & Khobta, A. (2011). 8-Oxo-7,8-dihydroguanine in DNA does not constitute a barrier to transcription, but is converted into transcription-blocking damage by OGG1. *Nucleic Acids Research*, 39(14), 5926–5934. <https://doi.org/10.1093/nar/gkr163>
- Klungland Arne, & Lindahl Tomas. (1997). Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1). *The EMBO Journal*, 16(11), 3341–3348.
- Knobel, P. A., & Marti, T. M. (2011). Translesion DNA synthesis in the context of cancer research. In *Cancer Cell International* (Vol. 11).
<https://doi.org/10.1186/1475-2867-11-39>
- Krokan, H. E., Drabløs, F., & Slupphaug, G. (2002). Uracil in DNA-occurrence, consequences and repair. *Oncogene*, 21, 8935–8948.
<https://doi.org/10.1038/sj.onc>
- Kuraoka, I., Endou, M., Yamaguchi, Y., Wada, T., Handa, H., & Tanaka, K. (2003). Effects of endogenous DNA base lesions on transcription elongation by mammalian RNA polymerase II. Implications for transcription-coupled DNA repair and transcriptional mutagenesis. *Journal of Biological Chemistry*, 278(9), 7294–7299. <https://doi.org/10.1074/jbc.M208102200>
- Landry, J. J. M., Pyl, P. T., Rausch, T., Zichner, T., Tekkedil, M. M., Stütz, A. M., Jauch, A., Aiyar, R. S., Pau, G., Delhomme, N., Gagneur, J., Korbel, J. O., Huber, W., & Steinmetz, L. M. (2013). The genomic and transcriptomic landscape of a hela cell line. *G3: Genes, Genomes, Genetics*, 3(8), 1213–1224.
<https://doi.org/10.1534/g3.113.005777>
- Lange, S. S., Takata, K. I., & Wood, R. D. (2011). DNA polymerases and cancer. In *Nature Reviews Cancer* (Vol. 11, Issue 2, pp. 96–110).
<https://doi.org/10.1038/nrc2998>
- Lawrence, C. W., Borden, A., Banerjee, S. K., & Leclerc, J. E. (1990). Mutation frequency and spectrum resulting from a single abasic site in a single-stranded vector. *Nucleic Acids Research*, 18(8), 2153–2157.
<https://doi.org/10.1093/nar/18.8.2153>
- Lehmann, J., Seebode, C., Smolorz, S., Schubert, S., & Emmert, S. (2017). XPF knockout via CRISPR/Cas9 reveals that ERCC1 is retained in the cytoplasm without its heterodimer partner XPF. *Cellular and Molecular Life Sciences*, 74(11), 2081–2094. <https://doi.org/10.1007/s00018-017-2455-7>
- Levy, D. D., Saijo, M., Tanaka, K., & Kraemer, K. H. (1995). Expression of a transfected DNA repair gene (XPA) in xeroderma pigmentosum group A cells restores normal DNA repair and mutagenesis of UV-treated plasmids. In

- Carcinogenesis* (Vol. 16, Issue 7).
<https://academic.oup.com/carcin/article/16/7/1557/268226>
- Lhomme, J., Constant, J. F., & Demeunynck, M. (1999). Abasic DNA structure, reactivity, and recognition. In *Biopolymers* (Vol. 52, Issue 2, pp. 65–83). John Wiley & Sons Inc. [https://doi.org/10.1002/1097-0282\(1999\)52:2<65::AID-BIP1>3.0.CO;2-U](https://doi.org/10.1002/1097-0282(1999)52:2<65::AID-BIP1>3.0.CO;2-U)
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760.
<https://doi.org/10.1093/bioinformatics/btp324>
- Lin, W., Xin, H., Zhang, Y., Wu, X., Yuan, F., & Wang, Z. (1999). The human REV1 gene codes for a DNA template-dependent dCMP transferase. *Nucleic Acids Research*, 27(22), 4468–4475.
- Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, 362, 709–715.
- Lindahl, T., & Andersson Annika. (1972). Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. *Biochemistry*, 11(19), 3618–3623.
<https://pubs.acs.org/sharingguidelines>
- Liu, J., Zhou, W., & Doetsch, P. W. (1995). RNA Polymerase Bypass at Sites of Dihydrouracil: Implications for Transcriptional Mutagenesis. *MOLECULAR AND CELLULAR BIOLOGY*, 15(12), 6729–6735.
- Livneh, Z., Ziv, O., & Shachar, S. (2010). Multiple two-polymerase mechanisms in mammalian translesion DNA synthesis. In *Cell Cycle* (Vol. 9, Issue 4, pp. 729–735). Taylor and Francis Inc. <https://doi.org/10.4161/cc.9.4.10727>
- Locatelli, G. A., Pospiech, H., Tanguy Le Gac, N., Van Loon, B., Hubscher, U., Parkkinen, S., Syvˆaoja, J. E., Syvˆaoja, S., & Villani, G. (2010). Effect of 8-oxoguanine and abasic site DNA lesions on in vitro elongation by human DNA polymerase ϵ in the presence of replication protein A and proliferating-cell nuclear antigen. *Biochem. J*, 429, 573–582. <https://doi.org/10.1042/BJ20100405>
- Loeb, L. A., & Preston, B. D. (1986). MUTAGENESIS BY APURINIC/ APYRIMIDINIC SITES. In *Ann. Rev. Genet* (Vol. 20). www.annualreviews.org
- Lühnsdorf, B., Epe, B., & Khobta, A. (2014). Excision of uracil from transcribed DNA negatively affects gene expression. *Journal of Biological Chemistry*.
<https://doi.org/10.1074/jbc.M113.521807>
- Lühnsdorf, B., Kitsera, N., Warken, D., Lingg, T., Epe, B., & Khobta, A. (2012). Generation of reporter plasmids containing defined base modifications in the DNA strand of choice. *Analytical Biochemistry*. <https://doi.org/10.1016/j.ab.2012.03.001>
- Macville, M., Schröck, E., Padilla-Nash, H., Keck, C., Ghadimi, B. M., Zimonjic, D., Popescu, N., & Ried, T. (1999). Comprehensive and Definitive Molecular Cytogenetic Characterization of HeLa Cells by Spectral Karyotyping. *CANCER RESEARCH*, 59, 141–150.
- Malfatti, M. C., Balachander, S., Antoniali, G., Koh, K. D., Saint-Pierre, C., Gasparutto, D., Chon, H., Crouch, R. J., Storici, F., & Tell, G. (2017). Abasic and oxidized ribonucleotides embedded in DNA are processed by human APE1 and not by RNase H2. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkx723>
- Marenstein, D. R., Wilson, D. M., & Teebor, G. W. (2004). Human AP endonuclease (APE1) demonstrates endonucleolytic activity against AP sites in single-stranded

References

- DNA. *DNA Repair*, 3(5), 527–533. <https://doi.org/10.1016/j.dnarep.2004.01.010>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, 17(1), 10–12. <http://www-huber.embl.de/users/an->
- Masutani, C., Araki, M., Yamada, A., Kusumoto, R., Nogimori, T., Maekawa, T., Iwai, S., & Hanaoka, F. (1999). Xeroderma pigmentosum variant (XP-V) correcting protein from HeLa cells has a thymine dimer bypass DNA polymerase activity. In *The EMBO Journal* (Vol. 18, Issue 12).
- Matsumoto Y, & Kim K. (1995). Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. *Science*, 269, 699–702.
- McCulloch, S. D., & Kunkel, T. A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. In *Cell Research* (Vol. 18, Issue 1, pp. 148–161). <https://doi.org/10.1038/cr.2008.4>
- McIntyre, J. (2020). Polymerase iota - an odd sibling among Y family polymerases. In *DNA Repair* (Vol. 86). Elsevier B.V. <https://doi.org/10.1016/j.dnarep.2019.102753>
- Minko, I. G., Harbut, M. B., Kozekov, I. D., Kozekova, A., Jakobs, P. M., Olson, S. B., Moses, R. E., Harris, T. M., Rizzo, C. J., & Lloyd, R. S. (2008). Role for DNA polymerase κ in the processing of N2-N 2-guanine interstrand cross-links. *Journal of Biological Chemistry*, 283(25), 17075–17082. <https://doi.org/10.1074/jbc.M801238200>
- Mohni, K. N., Wessel, S. R., Zhao, R., Wojciechowski, A. C., Luzwick, J. W., Layden, H., Eichman, B. F., Thompson, P. S., Mehta, K. P. M., & Cortez, D. (2019). HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell*, 176(1–2), 144-153.e13. <https://doi.org/10.1016/j.cell.2018.10.055>
- Morgan, M., Pagès, H., Obenchain, V., & Hayden, N. (2021). *Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import*. <https://doi.org/10.18129/B9.bioc.Rsamtools>
- Mozzherin, D. J., Shibutani, S., Tan, C. K., Downey, K. M., & Fisher, P. A. (1997). Proliferating cell nuclear antigen promotes DNA synthesis past template lesions by mammalian DNA polymerase δ . *Proceedings of the National Academy of Sciences of the United States of America*, 94(12), 6126–6131. <https://doi.org/10.1073/pnas.94.12.6126>
- Nagel, Z. D., Margulies, C. M., Chaim, I. A., McRee, S. K., Mazzucato, P., Ahmad, A., Abo, R. P., Butty, V. L., Forget, A. L., & Samson, L. D. (2014). Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(18). <https://doi.org/10.1073/pnas.1401182111>
- Nair, D. T., Johnson, R. E., Prakash, L., Prakash, S., & Aggarwal, A. K. (2009). DNA Synthesis across an Abasic Lesion by Human DNA Polymerase ι . *Structure*, 17(4), 530–537. <https://doi.org/10.1016/j.str.2009.02.015>
- Nair, D. T., Johnson, R. E., Prakash, L., Prakash, S., & Aggarwal, A. K. (2011). DNA synthesis across an abasic lesion by yeast Rev1 DNA polymerase. *Journal of Molecular Biology*, 406(1), 18–28. <https://doi.org/10.1016/j.jmb.2010.12.016>
- Obeid, S., Blatter, N., Kranaster, R., Schnur, A., Diederichs, K., Welte, W., & Marx, A. (2010). Replication through an abasic DNA lesion: Structural basis for adenine selectivity. *EMBO Journal*, 29(10), 1738–1747.

<https://doi.org/10.1038/emboj.2010.64>

- Ogi, T., Shinkai, Y., Tanaka, K., & Ohmori, H. (2002). Polkappa protects mammalian cells against the lethal and mutagenic effects of benzo[a]pyrene. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15548–15553. <https://doi.org/10.1073/pnas.222377899>
- Ohashi, E., Murakumo, Y., Kanjo, N., Akagi, J.-I., Masutani, C., Hanaoka, F., & Ohmori, H. (2004). Interaction of hREV1 with three human Y-family DNA polymerases. *Genes to Cells*, *9*, 523–531. <https://doi.org/10.1111/j.1365-2443.2004.00747.x>
- Otsuka, C., Sanadai, S., Hata, Y., Okuto, H., Noskov, V. N., Loakes, D., & Negishi, K. (2002). Difference between deoxyribose- and tetrahydrofuran-type abasic sites in the in vivo mutagenic responses in yeast. In *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkf666>
- Parker, J. L., & Ulrich, H. D. (2009). Mechanistic analysis of PCNA poly-ubiquitylation by the ubiquitin protein ligases Rad18 and Rad5. *EMBO Journal*, *28*(23), 3657–3666. <https://doi.org/10.1038/emboj.2009.303>
- Patra, A., Zhang, Q., Lei, L., Su, Y., Egli, M., & Guengerich, F. P. (2015). Structural and kinetic analysis of nucleoside triphosphate incorporation opposite an abasic site by human translesion DNA polymerase. *Journal of Biological Chemistry*, *290*(13), 8028–8038. <https://doi.org/10.1074/jbc.M115.637561>
- Petrova, L., Gran, C., Bjoras, M., Doetsch, P. W., & Volkert, M. R. (2016). Efficient and Reliable Production of Vectors for the Study of the Repair, Mutagenesis, and Phenotypic Consequences of Defined DNA Damage Lesions in Mammalian Cells. *PLOS One*. <https://doi.org/10.1371/journal.pone.0158581>
- Prakash, S., Johnson, R. E., & Prakash, L. (2005). EUKARYOTIC TRANSLESION SYNTHESIS DNA POLYMERASES: Specificity of Structure and Function. *Annual Review of Biochemistry*. <https://doi.org/10.1146/annurev.biochem.74.082803.133250>
- Prakash, S., & Prakash, L. (2002). Translesion DNA synthesis in eukaryotes: A one- or two-polymerase affair. In *Genes and Development* (Vol. 16, Issue 15, pp. 1872–1883). <https://doi.org/10.1101/gad.1009802>
- Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., & Cormier, M. J. (1992). Primary structure of the *Aequorea victoria* green-fluorescent protein (Bioluminescence; Cnidaria; aequorin; energy transfer; chromophore; cloning). In *Biochemistry and Molecular Biology. Mayo Foundation* (Vol. 111, Issue 908).
- Pursell, Z. F., & Kunkel, T. A. (2008). Chapter 4 DNA Polymerase ϵ . A Polymerase of Unusual Size (and Complexity). In *Progress in Nucleic Acid Research and Molecular Biology* (Vol. 82, pp. 101–145). [https://doi.org/10.1016/S0079-6603\(08\)00004-4](https://doi.org/10.1016/S0079-6603(08)00004-4)
- Quinet, A., Lerner, L. K., Martins, D. J., & Menck, C. F. M. (2018). Filling gaps in translesion DNA synthesis in human cells. In *Mutation Research - Genetic Toxicology and Environmental Mutagenesis* (Vol. 836, pp. 127–142). Elsevier B.V. <https://doi.org/10.1016/j.mrgentox.2018.02.004>
- Quinet, A., Martins, D. J., Vessoni, A. T., Biard, D., Sarasin, A., Stary, A., & Menck, C. F. M. (2016). Translesion synthesis mechanisms depend on the nature of DNA damage in UV-irradiated human cells. *Nucleic Acids Research*, *44*(12), 5717–5731. <https://doi.org/10.1093/nar/gkw280>
- Quiñones, J. L., & Demple, B. (2016). When DNA repair goes wrong: BER-generated

References

- DNA-protein crosslinks to oxidative lesions. In *DNA Repair*.
<https://doi.org/10.1016/j.dnarep.2016.05.014>
- Rodriguez-Alvarez, M., Kim, D., & Khobta, A. (2020). EGFP reporters for direct and sensitive detection of mutagenic bypass of DNA lesions. *Biomolecules*, *10*(6).
<https://doi.org/10.3390/biom10060902>
- Rodriguez-Alvarez, Marta, Kim, D., & Khobta, A. (2020). EGFP reporters for direct and sensitive detection of mutagenic bypass of DNA lesions. *Biomolecules*, *10*(6), 1–19. <https://doi.org/10.3390/biom10060902>
- Sagher, D., & Strauss, B. (1983). Insertion of Nucleotides Opposite Apurinic/Apyrimidinic Sites in Deoxyribonucleic Acid during in Vitro Synthesis: Uniqueness of Adenine Nucleotides. *Biochemistry*, *22*, 4518–4526.
<https://pubs.acs.org/sharingguidelines>
- Sasatani, M., Zaharieva, E. K., & Kamiya, K. (2020). The in vivo role of Rev1 in mutagenesis and carcinogenesis. In *Genes and Environment* (Vol. 42, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s41021-020-0148-1>
- Saxowsky, T. T., & Doetsch, P. W. (2006). RNA polymerase encounters with DNA damage: Transcription-coupled repair or transcriptional mutagenesis? In *Chemical Reviews*. <https://doi.org/10.1021/cr040466q>
- Saxowsky, T. T., Meadows, K. L., Klungland, A., & Doetsch, P. W. (2008). 8-Oxoguanine-mediated transcriptional mutagenesis causes Ras activation in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(48), 18877–18882.
<https://doi.org/10.1073/pnas.0806464105>
- Schaaper, R. M., Kunkel, T. A., & Loeb, L. A. (1983). Infidelity of DNA synthesis associated with bypass of apurinic sites (depurination/bacteriophage OX174/transversion mutagenesis/DNA polymerases/SOS repair). *Proc. Natl Acad. Sci. USA*, *80*, 487–491.
- Schmitt, M. W., Matsumoto, Y., & Loeb, L. A. (2009). High fidelity and lesion bypass capability of human DNA polymerase δ . *Biochimie*, *91*(9), 1163–1172.
<https://doi.org/10.1016/j.biochi.2009.06.007>
- Schröder, A. S., Parsa, E., Iwan, K., Traube, F. R., Wallner, M., Serdjukow, S., & Carell, T. (2016). 2'-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes. *Chemical Communications*, *52*(100), 14361–14364. <https://doi.org/10.1039/c6cc07517g>
- Semlow, D. R., Zhang, J., Budzowska, M., Drohat, A. C., & Walter, J. C. (2016). Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. *Cell*, *167*(2), 498-511.e14. <https://doi.org/10.1016/j.cell.2016.09.008>
- Shachar, S., Ziv, O., Avkin, S., Adar, S., Wittschieben, J., Reißner, T., Chaney, S., Friedberg, E. C., Wang, Z., Carell, T., Geacintov, N., & Livneh, Z. (2009). Two-polymerase mechanisms dictate error-free and error-prone translesion DNA synthesis in mammals. *EMBO Journal*, *28*(4), 383–393.
<https://doi.org/10.1038/emboj.2008.281>
- Sharma, S., Helchowski, C. M., & Canman, C. E. (2013). The roles of DNA polymerase ζ and the Y family DNA polymerases in promoting or preventing genome instability. *Mutation Research*, *743–744*, 97–110.
<https://doi.org/10.1016/j.mrfmmm.2012.11.002>
- Shaw, R. J., Bonawitz, N. D., & Reines, D. (2002). Use of an in vivo reporter assay to

- test for transcriptional and translational fidelity in yeast. *Journal of Biological Chemistry*, 277(27), 24420–24426. <https://doi.org/10.1074/jbc.M202059200>
- Shibutani, S., Takeshita, M., & Grollman, A. P. (1997). Translesional Synthesis on DNA Templates Containing a Single Abasic Site A MECHANISTIC STUDY OF THE “A RULE”*. *The Journal of Biological Chemistry*, 272(21), 13916–13922. <http://www-jbc.stanford.edu/jbc/>
- Simpson, L. J., Ross, A. L., Szüts, D., Alviani, C. A., Oestergaard, V. H., Patel, K. J., & Sale, J. E. (2006). RAD18-independent ubiquitination of proliferating-cell nuclear antigen in the avian cell line DT40. *EMBO Reports*, 7(9), 927–932. <https://doi.org/10.1038/sj.embor.7400777>
- Snowden A, Kow YW, & Van Houten B. (1990). Damage Repertoire of the Escherichia coli UvrABC Nuclease Complex Includes Abasic Sites, Base-Damage Analogues, and Lesions Containing Adjacent 5' or 3' Nicks. *Proc. Natl. Acad. Sci. US.A*, 736, 7251–7259. <https://pubs.acs.org/sharingguidelines>
- Sobol RW, Prasad R, Evenski A, Baker A, Yang XP, Horton JK, & Wilson SH. (2000). The lyase activity of the DNA repair protein b-polymerase protects from DNA-damage-induced cytotoxicity. *Nature*, 405, 807–810.
- Spivak, G. (2015). Nucleotide excision repair in humans. In *DNA Repair* (Vol. 36, pp. 13–18). Elsevier B.V. <https://doi.org/10.1016/j.dnarep.2015.09.003>
- Spivak, G. (2016). Transcription-coupled repair: an update. In *Archives of Toxicology*. <https://doi.org/10.1007/s00204-016-1820-x>
- Stelter, P., & Ulrich, H. D. (2003). Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. *Nature*, 425, 188–191. www.nature.com/nature
- Sugasawa, K., Ng, J. M. Y., Masutani, C., Iwai, S., Van Der Spek, P. J., André, #, Eker, P. M., Hanaoka, F., Bootsma, D., & Hoeijmakers, J. H. J. (1998). Xeroderma Pigmentosum Group C Protein Complex Is the Initiator of Global Genome Nucleotide Excision Repair or endogenous genotoxic agents. DNA damage interferes with the primary DNA functions, including transcription and replication, which can lead to acute cell death. In addition, many DNA-damaging agents are also mutagens. The biological consequences of mutations. In *Molecular Cell* (Vol. 2).
- Sugitani, N., Sivley, R. M., Perry, K. E., Capra, J. A., & Chazin, W. J. (2016). XPA: A key scaffold for human nucleotide excision repair. In *DNA Repair* (Vol. 44, pp. 123–135). Elsevier B.V. <https://doi.org/10.1016/j.dnarep.2016.05.018>
- Sun, B., Latham, K. A., Dodson, M. L., & Lloyd, R. S. (1995). Studies on the catalytic mechanism of five DNA glycosylases: Probing for enzyme-DNA imino intermediates. *Journal of Biological Chemistry*, 270(33), 19501–19508. <https://doi.org/10.1074/jbc.270.33.19501>
- Swanson, R. L., Morey, N. J., Doetsch, P. W., & Jinks-Robertson, S. (1999). Overlapping Specificities of Base Excision Repair, Nucleotide Excision Repair, Recombination, and Translesion Synthesis Pathways for DNA Base Damage in *Saccharomyces cerevisiae*. In *MOLECULAR AND CELLULAR BIOLOGY* (Vol. 19, Issue 4).
- Szczesny, B., Tann, A. W., Longley, M. J., Copeland, W. C., & Mitra, S. (2008). Long patch base excision repair in mammalian mitochondrial genomes. *Journal of Biological Chemistry*, 283(39), 26349–26356.

References

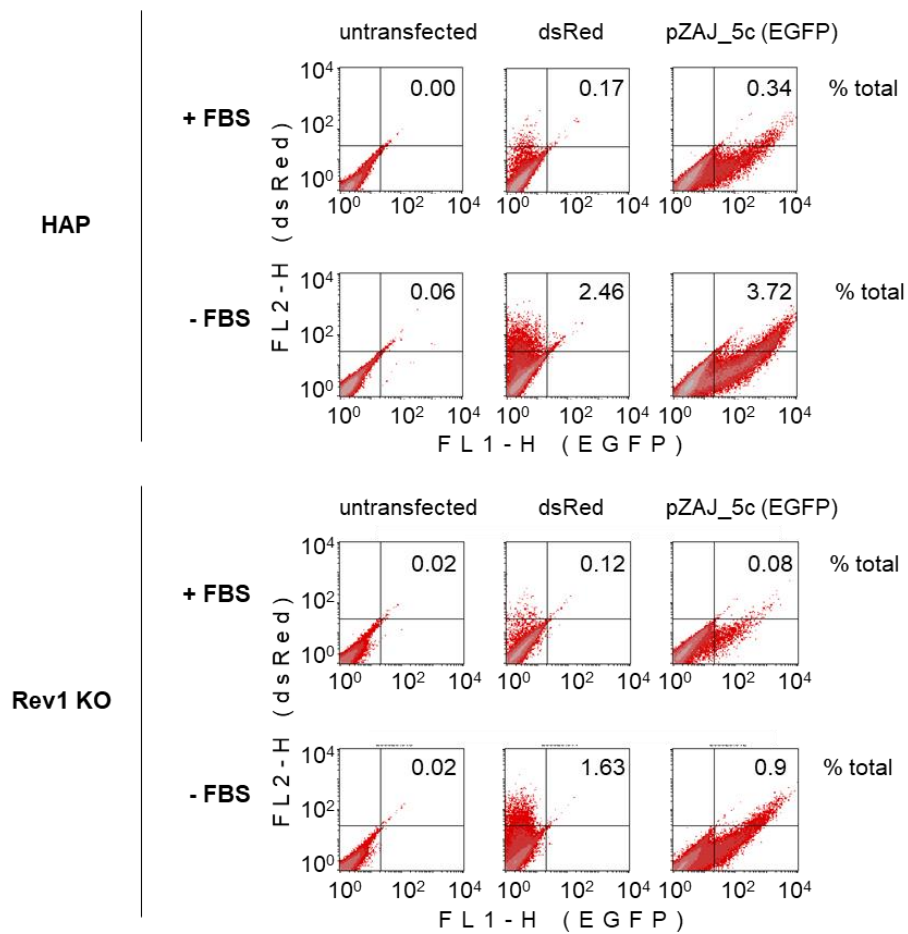
- <https://doi.org/10.1074/jbc.M803491200>
- Takeshita, M., Chang, C.-N., Johnson, F., Will, S., & Grollmans, A. P. (1987). Oligodeoxynucleotides Containing Synthetic Abasic Sites. *The Journal of Biological Chemistry*, 262(21), 10171–10179.
- Tamura, K., Kaneda, M., Futagawa, M., Takeshita, M., Kim, S., Nakama, M., Kawashita, N., & Tatsumi-Miyajima, J. (2019). Genetic and genomic basis of the mismatch repair system involved in Lynch syndrome. *International Journal of Clinical Oncology*, 24(9), 999–1011. <https://doi.org/10.1007/s10147-019-01494-y>
- Tateishi, S., Niwa, H., Miyazaki, J.-I., Fujimoto, S., Inoue, H., & Yamaizumi, M. (2003). Enhanced Genomic Instability and Defective Postreplication Repair in RAD18 Knockout Mouse Embryonic Stem Cells. *Molecular and Cellular Biology*, 23(2), 474–481. <https://doi.org/10.1128/mcb.23.2.474-481.2003>
- Tissier A, Frank, E. G., Mcdonald, J. P., Iwai, S., Hanaoka, F., & Woodgate, R. (2000). Misinsertion and bypass of thymine-thymine dimers by human DNA polymerase δ . *The EMBO Journal*, 19(19), 5259–5266.
- Tornaletti, S., Maeda, L. S., & Hanawalt, P. C. (2006). Transcription arrest at an abasic site in the transcribed strand of template DNA. *Chemical Research in Toxicology*. <https://doi.org/10.1021/tx060103g>
- Torres-Ramos, C. A., Johnson, R. E., Prakash, L., & Prakash, S. (2000). Evidence for the Involvement of Nucleotide Excision Repair in the Removal of Abasic Sites in Yeast. In *MOLECULAR AND CELLULAR BIOLOGY* (Vol. 20, Issue 10).
- Tsien, R. Y. (1998). THE GREEN FLUORESCENT PROTEIN. *Annu. Rev. Biochem*, 67, 509–544.
- Vaisman, A., & Woodgate, R. (2017). Translesion DNA polymerases in eukaryotes: what makes them tick? In *Critical Reviews in Biochemistry and Molecular Biology*. <https://doi.org/10.1080/10409238.2017.1291576>
- Villani, G., Shevelev, I., Orlando, E., Pospiech, H., Syvaaja, J. E., Markkanen, E., Hubscher, U., & Le Gac, N. T. (2014). Gap-directed translesion DNA Synthesis of an abasic site on circular DNA templates by a human replication complex. *PLoS ONE*, 9(4), e93908. <https://doi.org/10.1371/journal.pone.0093908>
- Viswanathan A, You HJ, & Doetsch PW. (1999). Phenotypic Change Caused by Transcriptional Bypass of Uracil in Nondividing Cells. *Science*, 284, 159–162. <http://science.sciencemag.org/>
- Wallace, S. S. (2014). Base excision repair: A critical player in many games. *DNA Repair*. <https://doi.org/10.1016/j.dnarep.2014.03.030>
- Wang, Wei, Walmacq, C., Chong, J., Kashlev, M., & Wang, D. (2018). Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), E2538–E2545. <https://doi.org/10.1073/pnas.1722050115>
- Wang, Wenjie, Sheng, W., Yu, C., Cao, J., Zhou, J., Wu, J., Zhang, H., & Zhang, S. (2015). REV3L modulates cisplatin sensitivity of non-small cell lung cancer H1299 cells. *Oncology Reports*, 34(3), 1460–1468. <https://doi.org/10.3892/or.2015.4121>
- Watanabe, K., Tateishi, S., Kawasuji, M., Tsurimoto, T., Inoue, H., & Yamaizumi, M. (2004). Rad18 guides pol η to replication stalling sites through physical interaction and PCNA monoubiquitination. *EMBO Journal*, 23(19), 3886–3896. <https://doi.org/10.1038/sj.emboj.7600383>

- Weerasooriya, S., Jasti, V. P., & Basu, A. K. (2014). Replicative bypass of abasic site in escherichia coli and human cells: Similarities and differences. *PLoS ONE*, 9(9), e107915. <https://doi.org/10.1371/journal.pone.0107915>
- Wiederhold, L., Leppard, J. B., Kedar, P., Karimi-Busheri, F., Rasouli-Nia, A., Weinfeld, M., Tomkinson, A. E., Wilson, S. H., & Mitra, S. (2004). AP Endonuclease-Independent DNA Base Excision Repair in Human Cells ing mitochondrial oxidative phosphorylation. Addition-ally, ROS are also produced during inflammatory re-sponses and result from exposure to ionizing radiation and a variety of chemicals. The DNA damage induced Tadahide Izumi, 1,2 Rajendra Prasad, 4 by ROS includes strand breaks and a number of base. In *Molecular Cell* (Vol. 15).
- Wilkinson, N. A., Mnuskin, K. S., Ashton, N. W., & Woodgate, R. (2020). Ubiquitin and ubiquitin-like proteins are essential regulators of DNA damage bypass. In *Cancers* (Vol. 12, Issue 10, pp. 1–20). MDPI AG. <https://doi.org/10.3390/cancers12102848>
- Wilson, D. M. (2003). Properties of and substrate determinants for the exonuclease activity of human apurinic endonuclease Ape1. *Journal of Molecular Biology*, 330(5), 1027–1037. [https://doi.org/10.1016/S0022-2836\(03\)00712-5](https://doi.org/10.1016/S0022-2836(03)00712-5)
- Wilson III David M, Takeshita Masaru, Grollman Arthur P, & Demple Bruce. (1995). Incision Activity of Human Apurinic Endonuclease (Ape) at Abasic Site Analogs in DNA. *Biological Chemistry*, 270(27), 16002–16007.
- Wojtaszek, J. L., Chatterjee, N., Najeeb, J., Ramos, A., Lee, M., Bian, K., Xue, J. Y., Fenton, B. A., Park, H., Li, D., Hemann, M. T., Hong, J., Walker, G. C., & Zhou, P. (2019). A Small Molecule Targeting Mutagenic Translesion Synthesis Improves Chemotherapy. *Cell*, 178(1), 152-159.e11. <https://doi.org/10.1016/j.cell.2019.05.028>
- Xanthoudakis, S., Smeyne, R. J., Wallace, J. D., Curran, T., & Burns, J. J. (1996). The redox/DNA repair protein, Ref-i, is essential for early embryonic development in mice (redox regulation/DNA repair/A/P endonuclease/gene targeting/oxidative stress). In *Biochemistry* (Vol. 93).
- Xiao W, & Chow BL. (1998). Synergism between yeast nucleotide and base excision repair pathways in the protection against DNA methylation damage. *Current Genetics*, 33, 92–99.
- Yamanaka, K., Chatterjee, N., Hemann, M. T., & Walker, G. C. (2017). Inhibition of mutagenic translesion synthesis: A possible strategy for improving chemotherapy? *PLoS Genetics*, 13(8). <https://doi.org/10.1371/journal.pgen.1006842>
- Yu, S.-L., Lee, S.-K., Johnson, R. E., Prakash, L., & Prakash, S. (2003). The Stalling of Transcription at Abasic Sites Is Highly Mutagenic. *Molecular and Cellular Biology*, 23(1), 382–388. <https://doi.org/10.1128/mcb.23.1.382-388.2003>
- Zeman, M. K., & Cimprich, K. A. (2014). Causes and consequences of replication stress. In *Nature Cell Biology* (Vol. 16, Issue 1, pp. 2–9). <https://doi.org/10.1038/ncb2897>
- Zhang, Y., Yuan, F., Wu, X., Taylor, J.-S., & Wang, Z. (2001). Response of human DNA polymerase I to DNA lesions. In *Nucleic Acids Research* (Vol. 29, Issue 4).
- Zheng, Y., & Sheppard, T. L. (2004). Half-Life and DNA Strand Scission Products of 2-Deoxyribonolactone Oxidative DNA Damage Lesions. *Chemical Research in Toxicology*, 17(2), 197–207. <https://doi.org/10.1021/tx034197v>
- Zhou, W., & Doetsch, P. W. (1993). Effects of abasic sites and DNA single-strand

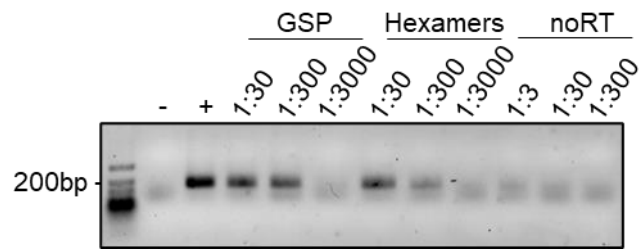
References

breaks on prokaryotic RNA polymerases (apurinic/aprimidinic sites/transcription/DNA damage). *Proc. Natl. Acad. Sci. USA*, 90, 6601–6605.

Appendix I: Supplementary figures and tables

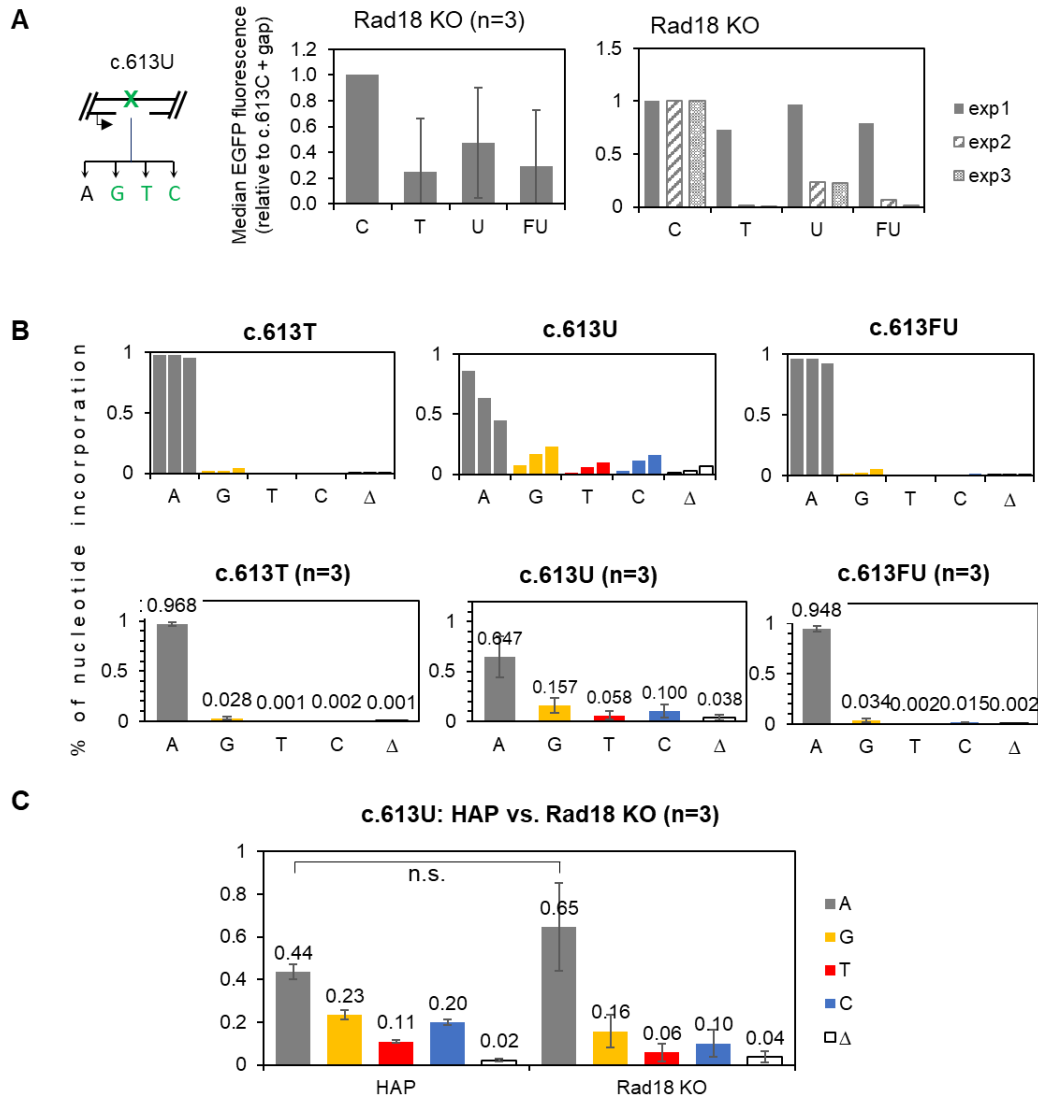


Supplementary figure 1: Transfection in FBS-supplemented vs. FBS-free medium conditions. Optimization for transfection for HAP and Rev1 KO cell-lines. Scatter plots 24h after transfection with dsRed (red) and pZAJ_5c (green) plasmids. For each cell line: Top: Transfection in medium supplemented with FBS. Bottom: Cells were incubated for 4h after transfection with FBS-free medium before changing the medium for an FBS-supplemented one. Percentage of transfected cells is indicated in the upper right panel.



Supplementary figure 2: Optimization of cDNA synthesis of RNA obtained from *ts.619dG* construct transfected in XP-A cells. 1.5% Agarose gel of PCR products from amplification after cDNA synthesis of dG samples using GSP (gene specific primer, random hexamers and the control sample without reverse transcriptase (noRT)).

Appendix I: Supplementary figures and tables



Supplementary figure 3: FACS analysis and RNA sequencing results of c.612G-c.613T/U/FU constructs for TLS assay transfected into Rad18 KO cell-line. (A) Left: Simplified scheme of phenotypic outcome after AP site resulting from uracil excision is bypassed. Right: Relative EGFP fluorescence of three independent experiments (mean \pm SD, n=3). (B) Percentage of nucleotide incorporation opposite to A, U and FU after TLS. Top: Results of the three independent experiments. Bottom: mean \pm SD, n=3. (C) Comparison of the mutant profile of natural AP sites in HAP vs. Rad18 KO cells, n.s. (non-significant two-tailed T-Test) T: thymine, C: cytosine, A: adenine, G: guanine, Δ : deletion

Sample number	Cell line	Type of experiment	Plasmid	Construct	RT/noRT	Index p5	Index p7
1	HAP	TLS	5c	c.613C	RT	D501	D701
2	HAP	TLS	5c		noRT	D501	D702
3	HAP	TLS	Q205X	c.613T	RT	D501	D703
4	HAP	TLS	Q205X		noRT	D501	D704
5	HAP	TLS	Q205X	c.613THF	RT	D501	D705
6	HAP	TLS	Q205X		noRT	D501	D706
7	HAP	TLS	Q205X	c.613U	RT	D501	D707
8	HAP	TLS	Q205X		noRT	D501	D708
9	HAP	TLS	Q205X	c.613FU	RT	D501	D709
10	HAP	TLS	Q205X		noRT	D501	D710
11	XP-A	TM	A207P	ts.619G	RT	D501	D711
12	XP-A	TM	A207P		noRT	D501	D712
13	XP-A	TM	A207P	ts.619S-THF	RT	D502	D701
14	XP-A	TM	A207P		noRT	D502	D702
15	HAP	TLS	Q205X	c.613T	noRT	D502	D703
16	HAP	TLS	Q205X		RT	D502	D704
17	HAP	TLS	Q205X	c.613THF	noRT	D502	D705
18	HAP	TLS	Q205X		RT	D502	D706
19	HAP	TLS	A207P	c.619THF	noRT	D502	D707
20	HAP	TLS	A207P		RT	D502	D708
21	HAP	TLS	S206X	c.617THF	noRT	D502	D709
22	HAP	TLS	S206X		RT	D502	D710
23	Rad18 KO	TLS	Q205X	c.613T	noRT	D502	D711
24	Rad18 KO	TLS	Q205X		RT	D502	D712
25	Rad18 KO	TLS	Q205X	c.613THF	noRT	D503	D701
26	Rad18 KO	TLS	Q205X		RT	D503	D702
27	Rad18 KO	TLS	A207P	c.619THF	noRT	D503	D703
28	Rad18 KO	TLS	A207P		RT	D503	D704
29	Rad18 KO	TLS	S206X	c.617THF	noRT	D503	D705
30	Rad18 KO	TLS	S206X		RT	D503	D706
31	HAP	TLS	Q205X	c.613T	RT	D503	D707
32	HAP	TLS	Q205X		noRT	D503	D708
33	HAP	TLS	Q205X	c.613THF	RT	D503	D709
34	HAP	TLS	Q205X		noRT	D503	D710
35	HAP	TLS	Q205X	c.613U	RT	D503	D711
36	HAP	TLS	Q205X		noRT	D503	D712
37	HAP	TLS	Q205X	c.613fU	RT	D504	D701
38	HAP	TLS	Q205X		noRT	D504	D702
39	Pol i KO	TLS	Q205X	c.613T	RT	D504	D703
40	Pol i KO	TLS	Q205X		noRT	D504	D704
41	Pol i KO	TLS	Q205X	c.613THF	RT	D504	D705
42	Pol i KO	TLS	Q205X		noRT	D504	D706
43	Rev 1 KO	TLS	Q205X	c.613T	RT	D504	D707
44	Rev 1 KO	TLS	Q205X		noRT	D504	D708
45	Rev 1 KO	TLS	Q205X	c.613THF	RT	D504	D709
46	Rev 1 KO	TLS	Q205X		noRT	D504	D710

Appendix I: Supplementary figures and tables

47	HAP	TLS	Q205X	c.613T	RT	D504	D711
48	HAP	TLS	Q205X		noRT	D504	D712
49	HAP	TLS	Q205X	c.613THF	RT	D505	D701
50	HAP	TLS	Q205X		noRT	D505	D702
51	HAP	TLS	Q205X	c.613U	RT	D505	D703
52	HAP	TLS	Q205X		noRT	D505	D704
53	HAP	TLS	Q205X	c.613fU	RT	D505	D705
54	HAP	TLS	Q205X		noRT	D505	D706
55	HAP	TLS	Q205X	c.613Tg	RT	D505	D707
56	HAP	TLS	Q205X		noRT	D505	D708
57	HAP	TLS	A207P	c.619Tg	RT	D505	D709
58	HAP	TLS	A207P		noRT	D505	D710
59	XP-A	TM	A207P	ts.619G	RT	D505	D711
60	XP-A	TM	A207P		noRT	D505	D712
61	XP-A	TM	A207P	ts.619THF	RT	D506	D701
62	XP-A	TM	A207P		noRT	D506	D702
63	XP-A	TM	A207P	ts.619S-THF	RT	D506	D703
64	XP-A	TM	A207P		noRT	D506	D704
65	Rad18 KO	TLS	A207P	c.613T	RT	D506	D705
66	Rad18 KO	TLS	A207P		noRT	D506	D706
67	Rad18 KO	TLS	A207P	c.613THF	RT	D506	D707
68	Rad18 KO	TLS	A207P		noRT	D506	D708
69	Rad18 KO	TLS	A207P	c.613U	RT	D506	D709
70	Rad18 KO	TLS	A207P		noRT	D506	D710
71	XP-A	TM	A207P	ts.619S-THF	RT	D506	D711
72	XP-A	TM	A207P		noRT	D506	D712
73	XP-A	TM	A207P	c.613T	RT	D507	D701
74	XP-A	TM	A207P	c.619S-THF	noRT	D507	D702
75	XP-A	TM	S206Y	ts.617T	RT	D507	D703
76	XP-A	TM	S206Y		noRT	D507	D704
77	XP-A	TM	S206Y	ts.617Tg	RT	D507	D705
78	XP-A	TM	S206Y		noRT	D507	D706
79	XP-A	TM	S206Y	ts.617 2'fTG	RT	D507	D707
80	XP-A	TM	S206Y		noRT	D507	D708
81	XP-A	TM	S206Y	ts.617T	RT	D507	D709
82	XP-A	TM	S206Y		noRT	D507	D710
83	XP-A	TM	S206Y	ts.617THF	RT	D507	D711
84	XP-A	TM	S206Y		noRT	D507	D712
85	XP-A	TM	S206Y	ts.617S-THF	RT	D508	D701
86	XP-A	TM	S206Y		noRT	D508	D702
87	XP-A	TM	S206Y	ts.619G	RT	D508	D703
88	XP-A	TM	S206Y		noRT	D508	D704
89	XP-A	TM	S206Y	ts.619S-THF	RT	D508	D705
90	XP-A	TM	S206Y		noRT	D508	D706

Supplementary table 1: List of samples sequenced in the first RNA library preparation. TLS: translesion synthesis; TM: transcriptional mutagenesis. RT: reverse transcription performed using gene-specific primer and 150ng RNA as template; noRT: without reverse transcription reaction.

Sample number	Cell-line	Type of experiment	Plasmid	Construct	RT/noRT	Index p5	Index p7
1	HAP	TLS	Q205X	c612G c613T	RT	D501	D701
2	HAP	TLS	Q205X	c612G c613U	RT	D501	D702
3	HAP	TLS	Q205X	c612G c613FU	RT	D501	D703
4	Rad18 KO	TLS	Q205X	c612G c613T	RT	D501	D704
5	Rad18 KO	TLS	Q205X	c612G c613U	RT	D501	D705
6	Rad18 KO	TLS	Q205X	c612G c613FU	RT	D501	D706
7	HAP	TLS	Q205X	c612G c613T	RT	D501	D707
8	HAP	TLS	Q205X	c612G c613U	RT	D501	D708
9	HAP	TLS	Q205X	c612G c613FU	RT	D501	D709
10	HAP	TLS	Q205X	c613TG	RT	D502	D701
11	HAP	TLS	S206X	c.617THF	RT	D502	D702
12	Rad18 KO	TLS	Q205X	c612G c613T	RT	D502	D703
13	Rad18 KO	TLS	Q205X	c612G c613U	RT	D502	D704
14	Rad18 KO	TLS	Q205X	c612G c613FU	RT	D502	D705
15	Rad18 KO	TLS	Q205X	c613TG	RT	D502	D706
16	Rad18 KO	TLS	S206X	c.617THF	RT	D502	D707
17	HAP	TLS	Q205X	c612G c613T	RT	D502	D708
18	HAP	TLS	Q205X	c612G c613U	RT	D502	D709
19	HAP	TLS	Q205X	c612G c613FU	RT	D503	D701
20	HAP	TLS	Q205X	c613TG	RT	D503	D702
21	HAP	TLS	S206X	c.617THF	RT	D503	D703
22	Rad18 KO	TLS	Q205X	c612G c613T	RT	D503	D704
23	Rad18 KO	TLS	Q205X	c612G c613U	RT	D503	D705
24	Rad18 KO	TLS	Q205X	c612G c613FU	RT	D503	D706
25	Rad18 KO	TLS	Q205X	c613TG	RT	D503	D707
26	Rad18 KO	TLS	S206X	c.617THF	RT	D503	D708
27	XP-A	TM	S206Y	ts617T	RT	D503	D709
28	XP-A	TM	S206Y		noRT	D504	D701
29	XP-A	TM	S206Y	ts617S-THF	RT	D504	D702
30	XP-A	TM	S206Y		noRT	D504	D703
31	XP-A	TM	S206Y	ts617T	RT	D504	D704
32	XP-A	TM	S206Y		noRT	D504	D705
33	XP-A	TM	S206Y	ts617S-THF	RT	D504	D706
34	XP-A	TM	S206Y		noRT	D504	D707

Supplementary table 2: List of samples sequenced in the second RNA library preparation. TLS: translesion synthesis; TM: transcriptional mutagenesis. RT: reverse transcription performed using random hexamers and 400ng RNA as template; noRT: without reverse transcription reaction.

Appendix I: Supplementary figures and tables

Cell-line	Construct	RT/noRT	Sample number	Library	Total Reads	Mutation profile					
						T	C	A	G	Δ	
XP-A	ts.619G	RT	11	1	81932	0.0003	0.9908	0.0007	0.0062	0.0045	
		RT	59	1	3368	0.0103	0.9072	0.0000	0.0825	0.0007	
		RT	87	1	97	0.9647	0.0068	0.0241	0.0014	0.9970	
		noRT	12	1	130750	0.0006	0.8757	0.0005	0.1226	0.0000	
		noRT	60	1	117057	0.0003	0.9678	0.0007	0.0299	0.0014	
		noRT	88	1	135431	0.0010	0.2962	0.0013	0.7015	0.0000	
	ts.619THF	RT	61	1	205	0.0000	0.9463	0.0000	0.0537	0.0000	
		noRT	62	1	175434	0.0002	0.9283	0.0011	0.0696	0.0000	
	ts.619S-THF	RT	13	1	289	0.0000	0.3728	0.3763	0.1254	0.1246	
		RT	63	1	93	0.0000	0.2796	0.5161	0.0753	0.1448	
		RT	71	1	321	0.0000	0.3053	0.5296	0.0405	0.0671	
		RT	89	1	14786	0.0010	0.3029	0.6006	0.0284	0.0005	
		noRT	14	1	133932	0.0007	0.6583	0.0003	0.3399	0.0000	
		noRT	64	1	139500	0.0004	0.9128	0.0008	0.0855	0.0002	
	ts.617T	noRT	72	1	165	0.0000	0.8727	0.0000	0.1273	0.0000	
		noRT	90	1	14786	0.0011	0.1225	0.0008	0.8754	0.0008	
		RT	81	1	3443	0.0019	0.0212	0.9770	0.0020	0.0005	
		RT	27	2	10038	0.0028	0.0272	0.9572	0.0024	0.0003	
		RT	31	2	10038	0.0007	0.0318	0.9691	0.0021	0.0009	
		noRT	82	1	107179	0.0000	0.6320	0.3630	0.0001	0.0001	
	ts.617THF	noRT	28	2	10043	0.0009	0.0318	0.9648	0.0025	0.0001	
		noRT	32	2	10049	0.0004	0.0292	0.9684	0.0021	0.0000	
		RT	83	1	432	0.0000	0.0463	0.9537	0.0000	0.0000	
		noRT	84	1	82720	0.0000	0.6385	0.3615	0.0000	0.0003	
		ts.617S-THF	RT	85	1	422	0.0011	0.0362	0.9310	0.0052	0.0282
			RT	29	2	10028	0.0023	0.1257	0.8322	0.0006	0.0340
	RT		33	2	10040	0.0027	0.0325	0.9132	0.0006	0.0437	
	noRT		86	1	88104	0.0000	0.7205	0.2783	0.0011	0.0001	
		noRT	30	2	10041	0.0019	0.0330	0.9627	0.0025	0.0000	
		noRT	34	2	10036	0.0017	0.0417	0.9539	0.0025	0.0003	

Supplementary table 3: RNA sequencing results of TM samples. Column library indicates the results that were obtained in the first or second library preparation. RT: reverse transcription performed; noRT: without reverse transcription reaction.

Cell-line	Construct	RT/noRT	Sample number	Library	Internal mutation	Contamination (%)	Contamination filtered out	Total Reads	Mutation profile					
									A	G	T	C	Δ	
HAP	c.613T	RT	3	1	no	unknown	no	3011	0.9827	0.0083	0.0007	0.0076	0.0007	
		RT	16	1	no	unknown	no	163757	0.9155	0.0828	0.0005	0.0003	0.0009	
		RT	31	1	no	unknown	no	102256	0.5132	0.4860	0.0005	0.0003	0.0000	
		RT	47	1	no	unknown	no	32162	0.6074	0.3909	0.0012	0.0005	0.0000	
		RT	1	2	yes (c.612G)	13.84	yes	10026	0.9775	0.0189	0.0009	0.0017	0.0011	
		RT	7	2	yes (c.612G)	5.02	yes	10038	0.9784	0.0191	0.0003	0.0010	0.0012	
		RT	17	2	yes (c.612G)	5.75	yes	10026	0.9741	0.0234	0.0005	0.0016	0.0004	
		noRT	4	1	no	unknown	no	164073	0.9586	0.0406	0.0004	0.0004	0.0000	
		noRT	15	1	no	unknown	no	140529	0.7132	0.2857	0.0005	0.0006	0.0000	
		noRT	32	1	no	unknown	no	71013	0.1014	0.8981	0.0002	0.0003	0.0000	
	c.613THF	noRT	48	1	no	unknown	no	71356	0.5726	0.4264	0.0002	0.0006	0.0002	
		RT	5	1	no	unknown	no	116660	0.6522	0.0345	0.0524	0.2154	0.0455	
		RT	18	1	no	unknown	no	113585	0.6111	0.1575	0.0472	0.1383	0.0460	
		RT	33	1	no	unknown	no	113032	0.4085	0.4045	0.0421	0.1275	0.0174	
		RT	49	1	no	unknown	no	43647	0.5768	0.3196	0.0161	0.0769	0.0107	
		noRT	6	1	no	unknown	no	12391	0.5724	0.3978	0.0079	0.0068	0.0150	
		noRT	17	1	no	unknown	no	51989	0.1822	0.8142	0.0003	0.0000	0.0033	
		noRT	34	1	no	unknown	no	116299	0.0984	0.8993	0.0013	0.0010	0.0000	
		noRT	50	1	no	unknown	no	18953	0.5084	0.4899	0.0013	0.0004	0.0000	
		c.613U	RT	7	1	no	unknown	no	76902	0.4950	0.2109	0.0823	0.1629	0.0490
RT	35		1	no	unknown	no	129624	0.2547	0.6016	0.0440	0.0878	0.0120		
RT	51		1	no	unknown	no	70804	0.5093	0.3765	0.0370	0.0591	0.0180		
RT	2		2	yes (c.612G)	4.26	yes	10018	0.4431	0.2401	0.1028	0.1999	0.0141		
RT	8		2	yes (c.612G)	6.84	yes	10000	0.3970	0.2538	0.1135	0.2142	0.0215		
RT	18		2	yes (c.612G)	5.43	yes	10006	0.4660	0.2098	0.1119	0.1855	0.0268		
noRT	8		1	no	unknown	no	3737	0.6048	0.3583	0.0059	0.0124	0.0186		
noRT	36		1	no	unknown	no	53765	0.0559	0.9430	0.0010	0.0001	0.0000		
noRT	52		1	no	unknown	no	58378	0.4603	0.5374	0.0020	0.0004	0.0000		
c.613FU	RT		9	1	no	unknown	no	68355	0.9788	0.0139	0.0006	0.0055	0.0012	
	RT	37	1	no	unknown	no	65430	0.5306	0.4624	0.0004	0.0053	0.0013		
	RT	53	1	no	unknown	no	94481	0.7024	0.2934	0.0014	0.0024	0.0004		
	RT	3	2	yes (c.612G)	4.21	yes	10025	0.9597	0.0192	0.0012	0.0189	0.0011		
	RT	9	2	yes (c.612G)	18.45	yes	10016	0.9603	0.0188	0.0026	0.0128	0.0056		
	RT	19	2	yes (c.612G)	3.77	yes	10026	0.9572	0.0217	0.0023	0.0173	0.0015		
	noRT	10	1	no	unknown	no	12585	0.6139	0.3813	0.0008	0.0040	0.0000		
	noRT	38	1	no	unknown	no	71069	0.0683	0.9313	0.0003	0.0001	0.0000		
	noRT	54	1	no	unknown	no	41616	0.4318	0.5665	0.0015	0.0002	0.0000		
	c.617THF	RT	22	1	yes (c.618A)	15.59	yes	9980	0.6563	0.0168	0.0393	0.2457	0.0419	
RT		11	2	yes (c.618A)	14.1	yes	10002	0.7002	0.0192	0.0392	0.2181	0.0234		
RT		21	2	yes (c.618A)	16.98	yes	9990	0.6719	0.0201	0.0460	0.2448	0.0171		
noRT		21	1	yes (c.618A)	unknown	no	76488	0.0049	0.9772	0.0175	0.0004	0.0000		
c.619THF	RT	20	1	no	unknown	no	130307	0.6044	0.0275	0.0797	0.2491	0.0025		
	noRT	19	1	no	unknown	no	47417	0.0064	0.0305	0.0005	0.9601	0.0010		
Rad18 KO	c.613T	RT	24	1	no	unknown	no	166508	0.7891	0.2081	0.0008	0.0004	0.0016	
		RT	65	1	no	unknown	no	33945	0.4672	0.5318	0.0007	0.0003	0.0000	
		RT	4	2	yes (c.612G)	7.38	yes	10027	0.9778	0.0192	0.0007	0.0010	0.0013	
		RT	12	2	yes (c.612G)	2.95	yes	10020	0.9761	0.0212	0.0006	0.0012	0.0009	
		RT	22	2	yes (c.612G)	unknown	yes	10032	0.9497	0.0439	0.0011	0.0037	0.0017	
		noRT	23	1	no	unknown	no	91619	0.3385	0.6606	0.0005	0.0004	0.0000	
		noRT	66	1	no	unknown	no	92887	0.2039	0.7954	0.0005	0.0001	0.0001	
		c.613THF	RT	26	1	no	unknown	no	14276	0.0527	0.9251	0.0047	0.0148	0.0028
			RT	67	1	no	unknown	no	52006	0.4391	0.4309	0.0298	0.0818	0.0184
			noRT	25	1	no	unknown	no	142129	0.0696	0.9294	0.0008	0.0002	0.0000
	noRT		68	1	no	unknown	no	35858	0.1985	0.7998	0.0017	0.0000	0.0000	
	c.613U	RT	69	1	no	unknown	no	77563	0.3582	0.5270	0.0433	0.0600	0.0115	
		RT	5	2	yes (c.612G)	12.62	yes	10008	0.8608	0.0766	0.0164	0.0306	0.0156	
		RT	13	2	yes (c.612G)	6.6	yes	9986	0.6332	0.1654	0.0597	0.1114	0.0303	
		RT	23	2	yes (c.612G)	17.34	yes	9952	0.4480	0.2282	0.0990	0.1577	0.0672	
	c.613FU	noRT	70	1	no	unknown	no	40630	0.1490	0.8502	0.0002	0.0002	0.0004	
		RT	6	2	yes (c.612G)	11.52	yes	10029	0.9614	0.0206	0.0023	0.0134	0.0023	
		RT	14	2	yes (c.612G)	6.5	yes	10024	0.9623	0.0233	0.0015	0.0107	0.0022	
		RT	24	2	yes (c.612G)	unknown	yes	10027	0.9205	0.0572	0.0016	0.0195	0.0011	
	c.619THF	RT	28	1	no	unknown	no	167420	0.5738	0.0282	0.0741	0.3053	0.0000	
noRT		27	1	no	unknown	no	38891	0.0060	0.0829	0.0010	0.9092	0.0186		
c617THF	RT	30	1	yes (c.618A)	27.97	yes	136452	0.6129	0.0162	0.0363	0.3120	0.0226		
		16	2	yes (c.618A)	25.5	yes	9946	0.6933	0.0200	0.0513	0.1956	0.0398		
		26	2	yes (c.618A)	22.17	yes	9930	0.6849	0.0202	0.0427	0.2235	0.0287		
	noRT	29	1	no	unknown	no	48874	0.0016	0.9789	0.0167	0.0012	0.0016		

Supplementary table 4: RNA sequencing results of TLS samples converted to DNA. Column library indicates the results that were obtained in the first or second library preparation. RT: reverse transcription performed; noRT: without reverse transcription reaction.

Appendix II: Sequences of plasmid vectors in FASTA format

Appendix II: Sequences of plasmid vectors in FASTA format**pZAJ_5c: Length 4880bp**

```

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCATATATGGAGTCCGCGTACATAACTACGGTAAATGGCCCGCTGGCTGACCG
CCCAACGACCCCGCCATTGACGCTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGCTCAATGGGTGGAGTATTTACGGT
AAACTGCCACTTGGCAGTACATCAAGTGTATCATATGCAAGTACGCCCTTATTGACGCTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTATGCGGTTTTGGCAGTACATCAATGGCGTGGGA
TAGCGGTTTGACTCAGGGGATTTCCAAGTCTCCACCCATTGACGCTCAATGGGAGTTGTTTTGGCACCAAAATCAACGGGACTTCCAAAAATGTCGTA
ACAACCTCCGCCCATTTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGCTATATAAGCAGAGCTGGTTTTAGTGAACCGTCAGATCCGCTAGCGCTA
CCGGTCGCGACTTGTGAGCAAGGGCAGGAGCTGTTCCACCGGGTGGAGCTCTCTGTCGAGCTGGACGGCGACGTAACAGGCCCAAGTTCAGCGT
TGTCCGGCAGGGCAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGGCCCTGGCCACCCCTCGTGAC
CACCTGACCTACGGCGTGCAGTGTCTCAGCCGTACCCGACACATGAAGCAGCAGACTTCTCAAGTCCGCCATGCCGGAAGGCTACGTCAGGAG
CGCACCATCTTCTCAAGGACGACGGCAACTACAAGACCAGCGCCGAGGTTGAAGTTCGAGGGCGACACCTGGTGAACCGCATCGAGCTGAAGGGCATCG
ACTTCAAGGAGGACGGCAACATCTGGGGCACAAAGCTGGAGTACAACACAAGCCACAACGCTATATCATGGCCGACAAGCAGAAGAACGGCAATCAA
GGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCTGACGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGTGCTG
CCCGACAACCACTACCTGAGCACCAGTCCGCCCTGAGCAAAAGACCCCAACGAGAAGCGCATACATGCTCTGCTGGAGTTCGTGACCGCCCGCGGGA
TCACTCTCGGCATGGACGAGCTGTACAAGTACTAGCATCTACACATTGATCTAGCAGAAGCACAGGTCGACGGGTGACGGTCCATCCGCTCTCTGGG
CACAAAGGCATGGGACGCTGCCATCTCTGCTCCTCCACTCCGGCGGGGAGCCATGGCTCTGGATCTGCTTCATGAGTGAAGTACAGTACAGGCT
CAAGCTTCAAGTTCGAGTTCGACGGTACCGGGCCCGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACACATTTGTAGAGGTTTTAC
TTGCTTAAAAAACCTCCACACCTCCCCCTGAACCTGAAACATAAAAATGAATGCAATTTGTTGTTAACTTGTATTGTCAGCTTATAATGGTTACAA
ATAAAGCAATAGCATCAAAAATTCACAAAATAAGCATTTTTCTACTGCATTAGTTGTTGTTGTTCCAAACTCATCAATGTATCTTAACGCGTAAAT
TGTAAAGCTTAAATTTTTGTTAAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCA
AAAGAATAGACCGAGATAGGGTTGAGTGTGTTCCAGTTTGAACAAGAGTCCACTATTAAGAAGCTGGACTCCAACGTCAAAGGGCGAAAAACCGTCT
ATCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTGGGGTCGAGGTGCCGTAAGACACTAAATCGGAACCTCAAAGGGAGCCCGG
ATTTAGAGCTTGACGGGAAAGCCGGCAACGTGGCGAGAAGGAAGGAAGAAAGCGAAAGGAGCGGGCGTAGGGCGTGGCAAGTGTAGCGGTACAG
CTGGCGGTAAACCACCAACCCCGCCGCTTAATGCGCCGTACAGGGCGGCTCAGGTGGCACTTTTCGGGAAATGTGCGGGAAACCCCTATTTGTTAT
TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATATATGAAAAAGGAAGAGTCTGAGCGGAAAGAAC
CAGCTGTGGAATGTGTGTCAGTTAGGGTGTGAAAGTCCCAAGGCTCCCAAGCAGGCAAGATGCAAAAGCATGATCTCAATAGTCAAGAACCCAGT
GTGAAAGTCCCAAGGCTTCTGTCAGAGCAGAGTATGCAAAAGCATGCTCAATTAAGTCAAGCAACATAGTCCCGCCCTAACTCCGCCATCCCGCC
CCTAACTCCGCCAGTTCGCCCATCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGCAAGGCGGAGGGCCGCTCGGCTCTGAGCTATTCCAG
AAGTAGTGAAGAGGCTTTTTGGAGGCCTAGGCTTTTGCAAAAGTGCATCAAGAGACAGGATGAGGATCGTTTCGATGATTGAACAAGATGGATTGCAC
GCAGGTTCTCCGGCCGCTGGGTGGAGAGGCTATTCCGGTATGACTGGGCAACAGACAATCGGCTGCTGATGCCCGGTGTTCCGGCTGTACGCGC
AGGGCGCCCGGTTCTTTTTGTCAAGACCCGACCTGTCCGGTCCCTGAATGAACTCAAGACGAGGCAAGCGCGGCTATCGTGGCTGGCCACGACGGCGT
TCCTTGGCAGCTGTGCTCGAGTGTGCTGTAAGCGGGAAGGACTGGCTGCTATTGGGCGAAGTCCCGGGCAGGATCTCTGTGATCTCACCTTGTCT
CTTGGCAGAAAGTATCCATCATGGCTGATGCTATGCGCGGCTGCATACGCTTGTATCCGGTACCTGCCATTTCGACCACCAAGCGAAACATCGCATCG
AGCGAGCAGTACTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACGTTCCGCCAGGCTCAA
GGCAGCATGCCGACGGCAGGATCTCGTCTGACCCATGGCGATGCTGCTTGGCGAATATCATGGTGGAAAAATGGCCGCTTTCTGATTATCATCAAC
TGTGGCCGGTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCAATGGGCTGACCGCTTCTCG
TGCTTACGGTATCGCCGCTCCCGATTCGCAAGCGCATCGCCTTCTATCGCCTTCTTACGAGTCTTCTGAGCGGACTCTGGGGTTCGAAATGACCGCAC
CAAGCGACGCCAACCTGCCATCAGGATTCGATTCACCCGCGCTTCTATGAAAGTGGGGTTCGGAATCGTTTTCCGGGACCGCCGCTGGATGA
TCTCCAGCGCGGGGATCTCATGCTGGAGTCTTCCGCCACCTTCCGGGAGGCTCAACTGAAACACGGAAGGAAAGCAATACCGGAAGCAACCCGCTAT
GACGGCAATAAAAAGACAGAAATAAACGACCGGTGTTGGTCTGTTGTTATAAACGCGGGTTCGGTCCAGGGCTGGCACTCTGTCGATACCCACCG
AGACCCATTTGGGGCAATACGCCCGGCTTCTTCTTTTTCCCAACCCACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCAGCCCAAGCTCGGGCGG
CAGGCCCTGCCATAGCCCTAGGTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTTAAATTAAGAGGATCAGGTTGAAGTCTTTTTGATAA
TCTCATGACCAAAATCCCTTAACTGAGTTTTTCGTCCACTGAGCGTCAAGCCCGTAGAAAAGATCAAAGGATCTTCTGAGATCTTTTTTCTGCGC
GTAACTGCTGCTTGAACAACAAAAAACCCGCTACAGCGGTGGTTTTGTTGCGGATCAAGAGCTACCAACTTTTTTCCGAAGGTAACGGTTCAG
CGAGAGCGCAGATACCAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCAGCCTACATACCTGCTGCTAAT
CTGTTACCAAGTGGCTGCTGCGAGTGGCGATAAGTCTGTCTTACCGGGTTGACTCAAGACGATAGTTACCGGATAAGGCGCAGGCTCGGGCTCAGCG
GGGGTTCGTGCACACAGCCAGCTTGGAGCGAAGCAGCTACACCGAAGTACGATACCTACAGCGTGAAGTATGAGAAAGCGCCAGCTTCCGAAAGGGA
GAAAGGCGGACAGGATTCGCGTAAGCGCAGGGTTCGGAACAGGAGAGCGCACGAGGGAGCTTCAAGGGGAAACCGCTGGTATCTTTATAGTCTGTGCGG
GTTTCGCCACCTCTGACTTGAAGCTGATTTTTGTGATGCTCGTCAAGGGGCGGAGGCTATGAAAAACGCCAGCAACCGCCCTTTTACGGTTCCTG
GCCTTTGCTGGCCTTTTCTCATACATGTTCTTCTGCGTTATCCCTGATTTCTGGATAACCGTATTACCGCATGCA

```

pZAJ_Q205*: Length 4880bp

```

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCATATATGGAGTCCGCGTACATAACTACGGTAAATGGCCCGCTGGCTGACCG
CCCAACGACCCCGCCATTGACGCTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGCTCAATGGGTGGAGTATTTACGGT
AAACTGCCACTTGGCAGTACATCAAGTGTATCATATGCAAGTACGCCCTTATTGACGCTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTATGCGGTTTTGGCAGTACATCAATGGCGTGGGA
TAGCGGTTTGACTCAGGGGATTTCCAAGTCTCCACCCATTGACGCTCAATGGGAGTTGTTTTGGCACCAAAATCAACGGGACTTCCAAAAATGTCGTA
ACAACCTCCGCCCATTTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGCTATATAAGCAGAGCTGGTTTTAGTGAACCGTCAGATCCGCTAGCGCTA
CCGGTCGCGACTTGTGAGCAAGGGCAGGAGCTGTTCCACCGGGTGGTGGCCATCTGTCGAGCTGGACGGCGACGTAACAGGCCCAAGTTCAGCGT
TGTCCGGCAGGGCAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGGCCCTGGCCACCCCTCGTGAC
CACCTGACCTACGGCGTGCAGTGTCTCAGCCGTACCCGACACATGAAGCAGCAGACTTCTCAAGTCCGCCATGCCGGAAGGCTACGTCAGGAG
CGCACCATCTTCTCAAGGACGACGGCAACTACAAGACCAGCGCCGAGGTTGAAGTTCGAGGGCGACACCTGGTGAACCGCATCGAGTGAAGGGCATCG
ACTTCAAGGAGGACGGCAACATCTGGGGCACAAAGCTGGAGTACAACACAACGCTATATCATGGCCGACAAGCAGAAGAACGGCAATCAA
GGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCTGACGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGTGCTG
CCCGACAACCACTACCTGAGCACCAGTCCGCCCTGAGCAAAAGACCCCAACGAGAAGCGCATACATGCTCTGCTGGAGTTCGTGACCGCCCGCGGGA
TCACTCTCGGCATGGACGAGCTGTACAAGTACTAGCATCTACACATTGATCTAGCAGAAGCACAGGTCGACGGGTGACGGTTCATCCGCTCTCTGGG
CACAAAGGCATGGGACGCTGCCATCTCTGCTCCTCCACTCCGGCGGGGAGCCATGGCTCTGGATCTGCTTCATGAGTGAAGTACAGTACAGGCT
CAAGCTTCAAGTTCGAGTTCGACGGTACCGGGCCCGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACACATTTGTAGAGGTTTTAC
TTGCTTAAAAAACCTCCACACCTCCCCCTGAACCTGAAACATAAAAATGAATGCAATTTGTTGTTAACTTGTATTGTCAGCTTATAATGGTTACAA
ATAAAGCAATAGCATCAAAAATTCACAAAATAAGCATTTTTCTACTGCATTAGTTGTTGTTGTTCCAAACTCATCAATGTATCTTAACGCGTAAAT
TGTAAAGCTTAAATTTTTGTTAAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCA
AAAGAATAGACCGAGATAGGGTTGAGTGTGTTCCAGTTTGAACAAGAGTCCACTATTAAGAAGCTGGACTCCAACGTCAAAGGGCGAAAAACCGTCT

```

ATCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTGGGGTTCGAGGTGCCGTAAGCACTAAATCGGAACCTAAAGGGAGCCCCG
 ATTTAGAGCTTGACGGGAAAGCCGGCGAACCTGGCGAGAAAGGAAGGAAGAAAGCAAGGAGCGGGCGCTAGGGCGTGGCAAGTGTAGCGGTACG
 CTGGCGGTAAACCACCAACCCCGCCGCTTAATGCGCCCTACAGGGCGGTGACGTGGCACTTTTCGGGGAATGTGCGGGAACCTATTTGTTTAT
 TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCTGAGCGGAAAGAAC
 CAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAGTCCCCAGGCTCCCCAGCAGCAGAAGTATGCAAAAGCATGCATCTCAATAGTCAGCAACCAAGT
 GTGGAAGTCCCAGCAGCAGGAGTATGCAAAAGCATCAATAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCTCCCGCC
 CCTAACTCCGCCAGTTCCGCCATTCTCCGCCCATGGCTGACTAATTTTTTTATTTATGCAAGAGGCGAGGGCCGCTCGGCTCGAGCTATTCCAG
 AAGTAGTGAGGAGGCTTTTTGGAGGCTTAGGCTTTGCAAAAGTGCATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCAC
 GCAGTTCTCCGGCCGCTGGGTGGAGAGGCTATTCCGGCTATGACTGGGCACAACAGCAATCGGCTGCTGATGCCCGGCTTCCGGCTGTACAGCGC
 AGGGGCGCCGGGTTCTTTGTCAAGACCGACCTGTCCGGTCCCTGAATGAAGTCAAGACAGGAGCGCGGCTATCGTGGCTGGCCACGACGGCGT
 TCCTTGGCAGCTGTGCTCGAGTGTCTGACTGAAGCGGGAAGGACTGGCTGCTATTGGGCGAAGTCCGGGGCAGGATCTCTGTCATCTACCTTGTCT
 CCTGCCGAGAAAGTATCCATCATGGCTGATGCTATGCGGGCTGCATACGCTTGTATCCGGCTACTGCCATTTCAGCCACCAAGCGAAACATCGCATCG
 AGCGAGCAGTACTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCCGACCGCAACTGTTCCGCCAGGCTCAA
 GCGAGGATGCCCAGCGGGGATCTCATGTCTGAGTCTTCCGCCACCTAGGGGAGGCTAAGTCCGGAATATCATGGTGGAAAAATGGCCGCTTTCTGATTCTCGAC
 TGTGGCCGGTGGGTGTGGCGGACCGTATCAGGACATAGCGTTGGCTACCGTGTATTTGCTGAAGAGCTTGGCGGCAATGGGCTGACCGCTTCTCG
 TGCTTACGGTATCGCCGCTCCCGATTGCGAGCGCATCGCTTCTATCGCTTCTTGACGAGTCTTCTGAGCGGAGCTCTGGGTTGCAAAATGACCGAC
 CAAGCGACGCCAACCTGCCATCAGGATTTCCAGCCCGCCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGA
 TCCTCCAGCGGGGATCTCATGTCTGAGTCTTCCGCCACCTAGGGGAGGCTAAGTGAACACCGGAAGGAGACAATACCGGAAGGAACCCGCGCTAT
 GACGGCAATAAAAAGACAGAATAAAACGACCGTGTGGTCTTGTTCATAAAGCGGGTTCGGTCCAGGGCTGGCACTCTGTCGATACCCACCG
 AGACCCATTGGGGCAATACGCCCGCTTCTTCTTTTTCCCAACCCACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCAGCCAACTCGGGCGG
 CAGGCCCTGCCATAGCTCAGTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTTAATTTAAAGGATCTAGGTGAAGATCTTTTTGATAA
 TCTCATGACAAAATCCCTTAACGTGAGTTTTCCGCTGAGCTGAGCGTCAAGCCCTAGAAAAGATCAAAAGGATCTTTTGAGATCTTTTTCTGGC
 GTAATCTGCTGTGCAAAACAAAAAACCCCGTACCAGCGGTGGTGTGTGGCGGATCAAGAGCTACCACTCTTTTTCCGAAGGTAAGTGGCTTCA
 GCAGAGCGCAGATACCAATCTGCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACATACCTCGCTCTGCTAAT
 CCTGTTACAGTGGCTTCCAGTGGCGATAAGTCTGTCTTACCGGGTGGACTCAAGACGATAGTTACCGGATAAGGCGAGCGGTGGGCTCGGCTCAAGC
 GGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAAGTGTGATACCTACAGCGTGAAGTATGAGAAAGCGCCACGCTTCCGAAAGGGA
 GAAAGCGGACAGGATCCGGTAAGCGCAGGGTCCGGAACGAGGAGCGCACGAGGAGCTTCCAGGGGAAACCGCTGGTATCTTTATAGTCTGTCCG
 GTTTCGCCACCTCTGACTTGAGCGTGCATTTTTGTGATGCTCGTCAAGGGGCGGAGGCTATGAAAAACCGCAGCAACCGCGCTTTTTACGGTCTCTG
 GCCTTTTGTGCGCTTTTGTCTACATGTTCTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCATGCA

pMR_Q205P: Length 4880bp

TAGTTAATAAGTAATCAATACGGGGTCAATTAGTTCATAGCCATATATGGAGTCCCGGTACATAACTACGGTAAATGGCCCGCTGGCTGACCG
 CCAACGACCCCGCCATTGACGTCAATAATGACGTATGTTCCATAGTAAAGCCCAATAGGGACTTTCCATTGACGTCAATGGGTGGATATTTACGGT
 AAATGCCACTTGGCAGTACATCAAGTGTATCATATGCAAGTACGCCCTTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
 CATGACCTTATGGGACTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTATGCGGTTTTGGCAGTACATCAATGGCGTGGGA
 TAGCGGTTGACTCAGGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTGCTGA
 ACAACTCCGCCCATTTGACGCAAAATGGGCGTAGGCTGTACGGTGGGGTGTCTATAAAGCAGAGCTGGTTTAGTGAACCTGCAGTCCGATCCGCTA
 CCGGTCCGACCATGGTGAAGGAGGCGAGGAGTGTACCAGGGTGGTCCCATCTGGTGAAGTGGACCGGCAAGCTGCCCTGGCCACCCCTCGTGAC
 TGTCCGGGAGGGGCGAGGCGATGCCACTACGGCAAGTGAACCTGAAGTTCATCTGACACCGGCAAGCTGCCCTGGCCACCCCTCGTGAC
 CACCCTACTCAGGCGTGCAGTGTCCAGCCGCTACCCGACCATGAAGCAGCAGACTTCTCAAGTCCGCATGCCGAAGGCTACGTCAGGAG
 CGCCACTCTTTTCAAGGACCGGCAACTCAAGACCCCGCGCGAGGTGAAGTTCGAGGGGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
 ACTTCAAGGAGGACGGCAACATCTGGGGCACAAGTGGAGTACAACACAGCCACACGCTATATCATGGCCGACAAGCAGAAGAAGGCAATCAA
 GGTGAATTCAGATCCGCCACAACATCGAGGACGGCAGCGTGAAGTCCGCCGACCACTACCAGCAGAAACCCCATCGCCGACGGCCCGTGTGCTG
 CCGACAACCACTACTGAGCACCCTCCGCTCCGCTGAGCAAAAGCCCAACGAGGAGGCGGATCACATGGTCTGTTGGAGTTCGTGAACCGCCCGGGGA
 TCACCTCGGCATGGACGAGTGTACAAGTACTAGCATCTACACATTGATCCTAGCAGAAGCAGGCTCGAGGGTGAAGGCTCCATCCGCTCTCTGGG
 CACAAGGATGGGCGAGTGCATCATCTGCTCCTCCACTCCGGCGGAGGCAATGGCTCTGGATCTGCTTATGAGTGAAGTACTCAGATCTGAGCT
 CAAGCTTCAATCTCGAGTGCAGGTTACCGGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCAATTTGTAGAGGTTTTAC
 TTGCTTTAAAAAACCTCCCAACCTCCGCTGAACTGAAACATAAAATGCAATGTTGTTGTTAACTTGTATTGAGCTTATGAGCTTATAATAA
 ATAAAGCAATAGCATCAAAATTTCAAAATAAAGCATTTTTTCTGCTTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTAAACGCTAAAT
 TGTAAAGCTTAAATTTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCA
 AAAGAATAGACCGAGATAGGTTGAGTGTGTTCCAGTTTGGAAACAAGAGTCCCACTATAAAGAACGTGGACTCCAACGTCAAAGGGGCAAAAACCGTCT
 ATCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAGTTTTGGGGTTCGAGGTGCCGTAAGCACTAAATCGGAACCTCAAAGGAGCCCGCCG
 ATTTAGAGCTTGACGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGAAGAAAGCAAGGAGCGGGCGCTAGGGCGTGGCAAGTGTAGCGGTACG
 CTGGCGTAAACCACCAACCCCGCCGCTTAATGCGCCCTACAGGGCGGTGACGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTTGTTTAT
 TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCTGAGCGGAAAGAAC
 CAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAGTCCCCAGGCTCCCCAGCAGGAGAAGTATGCAAAAGCATGCATCTCAATAGTCAGCAACCATAGT
 GTGAAAGTCCCAGGCTCCCCAGCAGGAGAAGTATGCAAAAGCATGCATCTCAATAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCATCCCGCC
 CCTAAGTCCGCCAGTTCCGCCATTCTCCGCCCATGGCTGACTAATTTTTTTATTTATGCAAGAGGCGAGGGCCGCTCGGCTCTGAGCTATTCCAG
 AAGTAGTGAGGAGGCTTTTTGGAGGCTAGGCTTTTTGCAAAAGTGCATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCAG
 GCAGGTTCTCCGGCCGCTGGGTGGAGAGGCTATTCCGGCTATGACTGGGCACAACAGACAATCGGCTGCTGATGCCCGGCTGTTCCGGCTGTACAGCG
 AGGGCGCCCGGTTCTTTTGTCAAGACCGACCTGTCCGGTCCCTGAATGAAGTCAAGACGAGGAGCGCGGCTATCGTGGCTGGCCACGACGGCGGT
 TCCTTGGCAGCTGTGCTCGAGTGTCTGACTGAAGCGGGAAGGACTGGCTGCTATTGGGCGAAGTCCGGGGCAGGATCTCTGTCATCTACCTTGTCT
 CCTGCCGAAAGTATCCATCATGGCTGATGCTATGCGGGCTGCATACGCTTGTATCCGGCTACTGCCATTTCAGCCACCAAGCGAAACATCGCATCG
 AGCGAGCAGTACTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCCGACCGCAACTGTTCCGCCAGGCTCAA
 GCGAGCATGCCGAGCGGAGGATCTCGTGTGACCCATGGCGATGCTGCTTGGCGAATATCATGGTGGAAAAATGGCCGCTTTCTGGATTATCGAC
 GTGGCCGGTGGGTGTGGCGGACCGTATCAGGACATAGCGTTGGCTACCCGCTGATTTGCTGAAGAGCTTGGCGGCAATGGGCTGACCGCTTCTCG
 TGCTTTACGGTATCGCCGCTCCCGATTGCGAGCGCATCGCTTCTATCGCTTCTGACGAGTCTTCTGAGCGGAGCTCTGGGGTTCGAAATGACCGAC
 CAAGCGACGCCAACCTGCCATCAGGATTTCCAGCCCGCCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGA
 TCCTCCAGCGGGGATCTCATGTCTGAGTCTTCCGCCACCTAGGGGAGGCTAAGTGAACACCGGAAGGAGACAATACCGGAAGGAACCCCGGCTAT
 GACGGCAATAAAAAGACAGAATAAAACGACCGTGTGGTCTTGTTCATAAAGCGGGTTCGGTCCAGGGCTGGCACTCTGTCGATACCCACCG
 AGACCCATTGGGGCAATACGCCCGCTTCTTCTTTTTCCCAACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCAGCCAGCCGCGGCTAT
 CAGGCCCTGCCATAGCTCAGTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTTAATTTAAAGGATCTAGGTGAAGATCTTTTTGATAA
 TCTCATGACAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAAGCCCTAGAAAAGATCAAAAGGATCTTTCTGAGATCTTTTTTCTGGC
 GTAATCTGCTGTGCAAAACAAAAAACCCCGTACCAGCGGTGGTGTGTTGGCGGATCAAGAGCTACCACTCTTTTTCCGAAGGTAAGTGGCTTCA
 GCAGAGCGCAGATAAAATATGCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACATACCTCGCTCTGCTAAT
 CCTGTTACAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTGGACTCAAGACGATAGTTACCGGATAAGGCGAGCGGTGGGCTGAACG
 GGGGTTCTGTCACAGCCAGCTTGGAGCGAACGACCTACACCGAAGTGTGATACCTACAGCGTGAAGTATGAGAAAGCGCCACGCTTCCGAAAGGGA

Appendix II: Sequences of plasmid vectors in FASTA format

GAAAGGCGGACAGGTATCCGGTAAAGCGGAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCAGGGGGAAACCGCTGGTATCTTTATAGTCTGTCCG
 GTTTCGCCACCTCTGACTTGAGCGTGCATTTTGTGATGCTCGTCAAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACCGCCCTTTTACGGTTCCTG
 GCCTTTTGTGGCCTTTTGTCTCACATGTTCTTCTGCTGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCATGCA

pMR_S206Y: Length 4880bp

TAGTTATTAATAGTAATCAATTACGGGGTCAATTAGTTCATAGCCATATATGGAGTTCGCGT TACATAACTTACGGTAAATGGCCGCTGGCTGACCG
 CCCAACGACCCCGCCATTGACGCTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGCTCAATGGGTGGGATTTTACGGT
 AAATGCCCCACTTGGCAGTACATCAAGTGTATCATATGCAAGTACGCCCTTATTGACGCTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
 CATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATAGTCAATGCTATTACCATGGTATGCGGTTTTGGCAGTACATCAATGGGCGTGGTA
 TAGCGGTTTGTACTCAGGGGATTTCCAAGTCTCCACCCATTGACGCTCAATGGGAGTGTGTTTTGGCACAAAATCAACGGGACTTTCAAAAATGTCGTA
 ACAACTCCGCCCATGACGCAAAATGGGCGTAGGCGTACGGTGGGAGGTCTATATAAGCAGAGTGGTTTGTAGTGAACCGTACAGTCCGCTAGCGCTA
 CGGGTCCGCCACCATGGTGAAGGCGAGGAGCTGTTCACCGGGTGGTCCCATCTGGTCAAGCTGGACGGCGACGTAACGGCCACAAGTTTCCAGCG
 TGTCCGGCAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGGCCCTGGCCACCTCGT
 CACCTGACCTACGGCGTGCAGTGTTCAGCCGTACCCGACCATGAAGCAGCAGCACTTCTCAAGTCCGCCATGCCGAAAGCTACGTCCAGGAG
 CGCACCATTCTTCAAGGACGACGGCAACTACAAGACCCGCGCGAGGTGAAGTTCGAGGGCGACACCTGGTGAACCGCATCGAGCTGAAGGGCATCG
 ACTTCAAGGAGGACGGCAACATCTGGGGCACAAGCTGGAGTACAACAGCACAACGCTATATATGGCCGACAAGCAGAAAGACGGCATCAA
 GGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGTGTG
 CCCGACAACCACTACCTGAGCACCAGTACGCCCTGAACTGAAACGCAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCTGACCGCCGCGGGA
 TCACCTCGGCATGGACGAGTGTACAAGTACTAGCATCTACACATTGATCTAGCAGAAGCAGAGGTGCAGGGTACGGTCCATCCGCTCTCTGGG
 CACAAGGATGGGCGAGCTGCCATCATCTGCTCTCCACCTCCGGGGGAAGCCATGGCTCTGGATCTGCTTCATGAGTGAAGTACTCAGATCTCGAGCT
 CAAGCTTCGAATCTCGACTCGAGGTACCGGGGCGGGGATCCACGGATCTAGATAACTGATCATAATCAGCCATACCAATTTGTAGAGGTTTTAC
 TTGCTTTAAAAAACCTCCACACCTCCCTTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTGTCAGCTTATAATGGTACAA
 ATAAAGCAATAGCATCACAATTTCAAAAATAAAGCATTTTTTCTACGCTTCTAGTTGTGGTTTTGTCCAACTCATCAATGTATCTTAAACGCTAAAT
 TGTAAAGGTTAATTTTTGTTAAAACTCGCTTAAATTTTTGTTAAATCAGCTCATTTTTTAAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCA
 AAAGAATAGACCAGATAGGGTTGAGTGTGTTCCAGTTTGGAAACAAGAGTCCACTATTAAGAAGCTGGACTCAAACGTCAAAGGGCGAAAAACCGTCT
 ATCAGGGCGATGGCCACTACGTGAACCATCACCTAATCAAGTTTTTGGGGTTCGAGGTGCGCTAAAGCACTAAATCGGAACCTTAAAGGGAGCCCGG
 ATTTAGAGCTTTCAGGGGAAAGCCGGCGAAGCTGGCGAGAAGGAAAGGAAAGGAAAGGAGGAGCGGCTGAGGGCGCTGGCAAGTGTAGCGGTACAG
 CTGCGCGTAACCAACACCCCGCGCTTAAATGCGCCGCTACAGGGCGCGTCAAGTGGCACTTTTCCGGGAAATGTGCGCGGAACCCCTATTTGTTTTAT
 TTTTCTAAATACATTTCAAAATGTATCGCTCATGAGCAATAACCCGTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCTCGAGCGGAAAGAAC
 CAGCTGTGGAATGTGTGTCAGTTAGGGTGTGAAAGTCCCAAGGCTCCCAAGCAGGAGATGCAAAAGCATGATCTCAATAGTACGCAACCAAGT
 GTGAAAGTCCCAAGCTCCCAAGCAGGAGATGCAAAAGCATGATCTCAATAGTACGCAACCATAGTCCCGCCCTAACTCCGCCATCCCGCC
 CCTAACTCCGCCAGTTCGCCCATCTCCGCCCATGGCTGACTAATTTTTTTATTTATGCAAGGGCCGAGGGCCGCTCGGCTCTGAGCTATTCCAG
 AAGTGAAGGAGGAGTCTTTTGGAGGCGTAGGCTTTTGAAGATCTCATGAGAGCAGGATGAGGATCGTTTTGCGATGATTGAACAAGATGGATTGCG
 CGAGGTTCTCCGGCCGTTGGTGGAGAGGCTATTGCGCTGACTGGGCACAACAGACAATGGCTGCTGATGCCCGGTTTCCGGCTGTGACGGC
 AGGGCGCCCGGTTCTTTTGTCAAGACCCGACTGTCCGGTCCCTGAATGAACTGCAAGAGCAGGAGCGCGGCTATCGTGGCTGGCCACGACGGCGT
 TCCTTGGCAGCTGTGCTCGAGTGTCTACTGAAGCGGGAAGGACTGGCTGCTATTGGGCGAAGTCCCGGGCAGGATCTCTGTCTATCTACCTTGTCT
 CCTGCCGAGAAAGTATCCATCATGGCTGATGCTATGCGCGGCTGCATACGCTTGTATCCGGCTACCTGCCATTGCAACCAACGAAAGCAATCGCATCG
 AGCGAGCAGTACTCGGATGGAAGCCGCTTGTGATGATGATCTGGCAGGAGCATAGGGGCTCGGCCAGCGCAACTGTTCGCCAGGCTCAA
 GGCAGCATGCCCGAGGCGAGGATCTCGTGTGACCCATGGCGATGCTGCTTGGCGAATATCATGGTGGAAAAATGGCCGCTTTCTGGATTATCGAC
 TGTGGCCGCTGGTGTGGCGGACCGCTATCAGGACATAGCCTTGGCTACCGGATATTGCTGAAGAGCTTGGCGGCAATGGGCTGACCCGCTCTCTCG
 TGCTTTACGGTATCGCCGCTCCCGATTGCGAGCGCATCGCTTCTACGCTTCTTGACGAGTCTCTTTCGAGCGGGACTCTGGGGTTTCAAAATCGCCAC
 CAAGCGACGCCAACCTGCGCATCAGGATTTGATTTCCACCGCGCTTCTATGAAAGGTTGGGCTTCGGAAATCGTTTTCCGGGACGCCGCTGGATGA
 TCCTCCAGCGCGGGGATCTCATGCTGGAGTCTTCCGCCACCTAGGGGAGGCTAACTGAAACACGGGAAGGAGCAATACCGGAAGGAACCCGCGCTAT
 GACGGCAATAAAAAGCAGAAATAAAGCAGCAGGCTGTTGGTGTGTTGTTCAATAAACCGGGGTTCCGGTCCAGGGCTGGCACTGTGCGATACCGCCAG
 AGACCCATTGGGGCAATACGCCCGCTTCTTCTTTTCCCAACCCACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCAGCAACGTCGGGGCGG
 CAGGCCCTGCCATAGCCTCAGTACTCATATACTTTAGATTGATTTAAACTTCATTTTTAATTTAAAGGATCTAGGTGAAGATCTTTTGTATAA
 TCTCATGACCAAAATCCCTTAACTGAGTGTTCGTTCCACTGAGCGTACAGCCCGTAGAAAAGATCAAAGGATCTTCTTGTAGATCTTTTTTCTGCGC
 GTAATCTGCTGCTTGTCAAAACAAAACCCCGCTACAGCGGTGGTGTGTTGTTGCGGATCAAGAGCTTCCAACTCTTTTTCCAGAGGTAACGCTTCA
 CGAGAGCGAGATACCAAACTGTCTTCTAGTGTAGCGTAGTTAGGCCACCATTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAAT
 CCTGTTACCACTGGCTGCTGCCAGTGGCGATAAGTCTGCTTACCGGGTTGACTCAAGACGATAGTTACCGGATAAGGCGACGGGTCGGGCTGAACG
 GGGGTTCTGTGACACAGCCAGCTTGGAGCGAAGCACTACACCGAATGAGATACCTACAGCTGAGCTATGAGAAAGCGCCAGCTTCCGAAAGGGA
 GAAAGGCGGACAGTATCCGGTAAAGCGGAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCTGGTATCTTTATAGTCTGTCGG
 GTTTCGCCACCTCTGACTTGAGCGTGCATTTTGTGATGCTCGTCAAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACCGCGCTTTTACGGTTCCTG
 GCCTTTTGTGGCCTTTTGTCTCACATGTTCTTCTGCTGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCATGCA

pMR_S206*: Length 4880bp

TAGTTATTAATAGTAATCAATTACGGGGTCAATTAGTTCATAGCCATATATGGAGTTCGCGT TACATAACTTACGGTAAATGGCCGCTGGCTGACCG
 CCCAACGACCCCGCCATTGACGCTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGCTCAATGGGTGGGATTTTACGGT
 AAATGCCCCACTTGGCAGTACATCAAGTGTATCATATGCAAGTACGCCCTTATTGACGCTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
 CATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATAGTCAATGCTATTACCATGGTATGCGGTTTTGGCAGTACATCAATGGGCGTGGTA
 TAGCGGTTTGTACTCAGGGGATTTCCAAGTCTCCACCCATTGACGCTCAATGGGAGTGTGTTTTGGCACAAAATCAACGGGACTTTCAAAAATGTCGTA
 ACAACTCCGCCCATGACGCAAAATGGGCGTAGGCGTACGGTGGGAGGTCTATATAAGCAGAGTGGTTTGTAGTGAACCGTACAGTCCGCTAGCGCTA
 CGGGTCCGCCACCATGGTGAAGGCGAGGAGCTGTTCACCGGGTGGTCCCATCTGGTCAAGCTGGACGGCGACGTAACGGCCACAAGTTTCCAGCG
 TGTCCGGCAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGGCCCTGGCCACCTCGT
 CACCTGACCTACGGCGTGCAGTGTTCAGCCGTACCCGACCATGAAGCAGCAGCACTTCTCAAGTCCGCCATGCCGAAAGCTACGTCCAGGAG
 CGCACCATTCTTCAAGGACGACGGCAACTACAAGACCCGCGCGAGGTGAAGTTCGAGGGCGACACCTGGTGAACCGCATCGAGCTGAAGGGCATCG
 ACTTCAAGGAGGACGGCAACATCTGGGGCACAAGCTGGAGTACAACAGCCACAACGCTATATCATGGCCGACAAGCAGAAAGACGGCATCAA
 GGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGTGTG
 CCCGACAACCACTACCTGAGCACCAGTGAAGCTTGAAGCAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCTGACCGCCGCGGGA
 TCACCTCGGCATGGACGAGTGTACAAGTACTAGCATCTACACATTGATCTTAGCAGAAGCAGGCTGCAGGGTACGGTCCATCCGCTCTCTGGG
 CACAAGGATGGGCGAGCTGCCATCATCTGCTCTCCACCTCCGGGGGAAGCCATGGCTCTGGATCTGCTTCATGAGTGAAGTACTCAGATCTCGAGCT
 CAAGCTTCGAATCTCGACTCGACGGTACCGGGGCGGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCAATTTGTAGAGGTTTTAC
 TTGCTTTAAAAAACCTCCACACCTCCCTTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTGTCAGCTTATAATGGTACAAA
 ATAAAGCAATAGCATCACAATTTCAAAAATAAAGCATTTTTTCTACGCTTCTAGTTGTGGTTTTGTCCAACTCATCAATGTATCTTAAACGCTAAAT
 TGTAAAGGTTAATTTTTGTTAAAACTCGCTTAAATTTTTGTTAAATCAGCTCATTTTTTAAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCA
 AAAGAATAGACCAGATAGGGTTGAGTGTGTTCCAGTTTGGAAACAAGAGTCCACTATTAAGAAGCTGGACTCAAACGTCAAAGGGCGAAAAACCGTCT

ATCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTGGGGTTCGAGGTGCCGTAAGCACTAAATCGGAACCTAAAGGGAGCCCCG
 ATTTAGAGCTTGACGGGAAAGCCGGCAACGTGGCGAGAAAGGAAGGAAGAAAGCGAAAGGAGCGGGCTAGGGCGTGGCAAGTGTAGCGGTACG
 CTGGCGGTAAACACCAACCCCGCCGCTAATGGCCGCTACAGGGCGGTAGTGGCACTTTTCGGGAAATGTGCGGGAAACCTATTTGTTTAT
 TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAGGAAGAGTCTGAGCGGAAAGAAC
 CAGCTGTGGAAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGCAGAAGTATGCAAAAGCATGCATCTCAATTAGTCAGCAACCCGGT
 GTGGAAAGTCCCAGCAGCAGGAGTATGCAAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACCTCCGCCCTCCCGCC
 CCTAACTCCGCCAGTTCGCCCATTTCCGCCCATGGCTGACTAATTTTTTTATTTATGCAGAGGCGAGGGCCCTCGGCTCTGAGCTATTCCAG
 AAGTAGTGAGGAGGCTTTTTGGAGGCTTAGGCTTTGCAAAAGTGCATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCAC
 GCAGTTTCCCGCCGCTTGGTGGAGAGGCTATTCCGGCTATGACTGGGCACAACAGCAATCGGCTGCTGTATGCCCGGTGTTCCGGCTGTACGCGC
 AGGGCGCCGGGTTCTTTTTGTCAAGACCGACCTGTCCGGTCCCTGAATGAAGTCAAGACAGGCGAGCGGGCTATCGTGGCTGGCCACGACGGCGT
 TCCTTGGCAGCTGTGCTCGAGTTGTCACTGAAGCGGGAAGGACTGGCTGCTATTGGGCGAAGTCCGGGGCAGGATCTCTGTCACTCACCTTGTCT
 CCTGCCGAGAAAGTATCCATCATGGCTGATGCTATGCGGGCTGCATACGCTTGATCCGGCTACTGCCATTTCAGCACCAAGCGAAACATCGCATCG
 AGCGAGCAGTACTCGGATGGAAGCCGGTCTTGTGCATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACCTGTTCCCGAGCTCAA
 GCGAAGCATGCCGAGCGGGATCTCATGTGAGTCTTCCGCCACCTAGGGGAGGCTAAGTGGCGAATATCATGGTGGAAAATGGCCGCTTTCTGATTTCAGC
 TGTGGCCGGTGGTGTGGCGGACCGTATCAGGACATAGCGTTGGCTACCGGTGATATTGCTGAAGAGCTTGGCGGCAATGGGCTGACCGCTTCTCG
 TGCTTACGGTATCGCCGCTCCCGATTGCGAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTCTTCTGAGCGGAGCTCGGGGTTGAAAATGACCGAC
 CAAGCGACGCCAACCTGCCATCACGAGATTTGATTCCACCCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGA
 TCCTCCAGCGCGGGATCTCATGTGAGTCTTCCGCCACCTAGGGGAGGCTAAGTGAACACAGGAGGAGCAATACCGGAAGGAAACCGCGCTAT
 GACGGCAATAAAAAGACAGAATAAAAACGACGGTGTGGTCTGTGTTGTTGATAAAGCGGGGTCGGTCCAGGGCTGGCACTGTGATACCCACCG
 AGACCCATTGGGGCAATACGCCCGCTTCTCCTTTTCCCAACCCACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCAGCCAACTCGGGCGG
 CAGGCCCTGCCATAGCTCAGTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTGATAA
 TCTCATGACAAAATCCCTTAACGTGAGTTTTCCCGCTGAGCGTCAAGCCCGTAGAAAAGATCAAAAGGATCTTTTGAGATCCTTTTTCTGCGC
 GTAATCTGCTGTGCAAAACAAAAAACCCCGTACCAGCGGTGGTGTGTTGCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAGTGGCTTCA
 GCAGAGCGCAGATACCAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACATACCTCGCTGTCTAAT
 CCTGTTACAGTGGCTGTGCGAGTGGCGATAAGTCTGTCTTACCGGGTGGACTCAAGACGATAGTTACCGGATAAGGCGACGGCTGGGCTCAAGC
 GGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACGAACTGAGATACCTACAGCGTGAAGTATGAGAAAGCGCCAGCTTCCGAAAGGGA
 GAAAGCGGACAGGATCCGGTAAGCGCAGGGTCCGGAACAGGAGGCGCACGAGGAGCTTCCAGGGGAAACCGCTGGTATCTTTATAGTCTGTCCG
 GTTTCGCCACCTCTGACTTGAGCGTGCATTTTTGTGATGCTGTCAGGGGGCGGAGCCTATGAAAACCGCAGCAACCGCGCTTTTACGGTCTCTG
 GCCTTTGCTGGCCTTTGCTACATGTTCTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCATGCA

pMR_A207P: Length 4880bp

TAGTTAATAAGTAATCAATACGGGGTCAATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCG
 CCAACGACCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAAAGCCCAATAGGGACTTTCCATTGACGTCAATGGTGGAGTATTACGGT
 AAATGCCACTTGGCAGTACATCAAGTGTATCATATGCAAGTACGCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
 CATGACCTTATGGGACTTCTACTTGGCAGTACATCTACGTATTAGTATCGCTATTACCATGGTATGCGGTTTTGGCAGTACATCAATGGCGTGGGA
 TAGCGGTTGACTCAGGGGATTTCCAAGTCTCCACCCATGACGTCAATGGGAGTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTGCTGA
 ACAACTCCGCCCATTGACGCAAAATGGGCGTAGCGGTGTACGGTGGAGGCTTATAAAGCAGAGCTGTTTGTAGTGAACCTCGATCCGATCCGCTA
 CCGGTCGCCACCATGGTGAAGGCGAGGAGCTGTTACCGGGTGGTGGCCATCCTGGTGAAGTGGACGGCAGCTAAACGGCCAAAGTTCAGCG
 TGTCCGGGAGGGCGAGGGCGATGCCACCTACGGCAAGTGAACCTGAAGTTTATCTGACACCCGGCAAGTGGCCGTGGCCACCCCTGTTGAC
 CACCCTGACTCAGCGGTGCGTGTGACCGCTACCCGACCATGAAGCAGCAGCACTTCTCAAGTCCGCCATGCCGAAGGCTACGTCAGGAG
 CGCCACTCTTTTCAAGGACGAGGCAACTCAAGACCCCGCGCGAGGTGAAGTTCGAGGGGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
 ACTTCAAGGAGGACGGCAACCTCTGGGCAACAAGTGGAGTACAACACAGCCACACGCTATATCATGGCCGACAAGCAGAAGAAGGGCATCAA
 GGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCCGCCGACCACTACCAGCAGAACCCCCATCGCCGACGGCCCGTGTGTG
 CCGACAACCACTTCCAGACCCCTCCCGTGAACCTGAACCAAGACCCCAAGGAGGAGCGGATCACATGGTCTGTGGAGTTCGTGAACCGCCGGGGA
 TCACCTCGGCATGGACGAGCTGTACAAGTACTAGCATCTACACATGATCCTAGCAGAAGCAGGCTGACGGGTGACGGTCCATCCGCTCTCTGGG
 CACAAGGATGGGCGAGCTGCCATCATCTGCTCCTCACCCTCCGGGGAAGCCATGGCTCTGGATCTGCTTATGAGTGAAGTACTCAGATCTGAGCT
 CAAGCTTCAATCTCGAGTGCAGGTACCGGGGCGGGGATCACCAGGATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTAC
 TTGCTTTAAAAAACCTCCCAACCTCCCGTGAACCTGAAACATAAAATGCAATTGTTGTTGTTAACTTGTTTATTCAGCTTATATGTTTACAA
 ATAAAGCAATAGCATCACAATTTCAAAATAAAGCATTTTTTCTACGCTTCTAGTTGTGGTGTGTTCCAAACTCATCAATGTATCTTAACGCTAAAT
 TGTAAAGCTTAATTTTTGTAATAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCA
 AAAGAATAGCCGAGATAGGTTGAGTGTGTTCCAGTTTGGAAACAAGAGTCCCACTATAAAGAACGTGGACTCCAACGTCAAAGGGGCAAAAACCGTCT
 ATCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTGGGGTTCGAGGTGCCGTAAGCACTAAATCGGAACCTCAAAGGGAGCCCGC
 ATTTAGAGCTTGACGGGAAAGCCGGCAACGTGGCGAGAAAGGAAGAAAGCGAAAGGAGCGGGCTAGGGCGTGGCAAGTGTAGCGGTACG
 CTGGCGGTAAACACCAACCCCGCCGCTAATGGCCGCTACAGGGCGGTAGTGGCACTTTTCGGGAAATGTGCGGGAAACCCCTATTTGTTTAT
 TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAGGAAGAGTCTGAGCGGAAAGAAC
 CAGCTGTGGAAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGAGGATGCAAAAGCATGCATCTCAATTAGTCAGCAACCATAGT
 GTGAAAGTCCCAGGCTCCCCAGCAGGAGAAGTATGCAAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCATCCCGCC
 CCTAACTCCGCCAGTTCGCCCATTTCCGCCCATGGCTGACTAATTTTTTTATTTATGCAGAGGCGGAGGGCCCTCGGCTCTGAGCTATTCCAG
 AAGTAGTGAGGAGGCTTTTTGGAGGCTAGGCTTTTTGCAAAAGTGCATCAAGAGACAGGATGAGGATCGTTTTGCATGATTGAACAAGATGGATTGCA
 GCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCCGGCTATGACTGGGCACAACAGACAATCGGCTGCTGTGATCGCGCGTGTCCGGCTGTACGCG
 AGGGCGCCGGGTTCTTTTTGTCAAGACCGACCTGTCCGGTCCCTGAATGAAGTCAAGAGCAGGAGGCGGGCTATCGTGGCTGGCCACGACGGGGT
 TCCTTGGCAGCTGTGCTCGAGTTGTCACTGAAGCGGGAAGGACTGGCTGCTATTGGGCGAAGTCCGGGGCAGGATCTCTGTCACTCACCTTGTCT
 CCTGCCGAAAGTATCCATCATGGCTGATGCTATGCGGGCTGCATACGCTTGATCCGGCTACTGCCATTTCAGCACCAAGCGAAACATCGCATCG
 AGCGAGCAGTACTCGGATGGAAGCCGGTCTTGTGCATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACCTGTTCCCGAGCTCAA
 GCGAGCATGCCGAGGCGAGGATCTCGTGTGACCCATGGCGATGCTGCTTGGCGAATATCATGGTGGAAAATGGCCGCTTTCTGATTATCATCGAC
 GTGGCCGGTGGTGTGGCGGACCGTATCAGGACATAGCCTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCAATGGGCTGACCGCTTCTCG
 TGCTTTACGGTATCGCCGCTCCCGATTGCGAGCGCATCGCCTTCTACGCTTCTGACGAGTCTTCTGAGCGGAGCTTGGGGTTCGAAATGACCGAC
 CAAGCGACGCCAACCTGCCATCACGAGATTTGATTCCACCCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGA
 TCCTCCAGCGCGGGATCTCATGTGAGTCTTCCGCCACCTAGGGGAGGCTAAGTGGTTCATAAAGCGGGGTCGGTCCAGGGCTGGCACTGTGATACCCACCG
 AGACCCATTGGGGCAATACGCCCGCTTCTCCTTTTCCCAACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCAGCCAACTCGGGCTGGGGCG
 CAGGCCCTGCCATAGCTCAGTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTGATAA
 TCTCATGACAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAAGCCCGTAGAAAAGATCAAAAGGATCTTTTGAGATCCTTTTTTTCTGCGC
 GTAATCTGCTGTGCAAAACAAAAAACCCCGTACCAGCGGTGGTGTGTTGTTGCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAGTGGCTTCA
 GCAGAGCGCAGATACCAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACATACCTCGCTGTCTAAT
 CCTGTTACAGTGGCTGTGCGAGTGGCGATAAGTCTGTCTTACCGGGTGGACTCAAGACGATAGTTACCGGATAAGGCGACGGCTGGGCTGAACG
 GGGGGTTCGTGCACAGCCAGCTTGGAGCGAACGACCTACACGAACTGAGATACCTACAGCGTGAAGTATGAGAAAGCGCCAGCTTCCGAAAGGGA

Appendix II: Sequences of plasmid vectors in FASTA format

```
GAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCTGTCCGG  
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTG  
GCCTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCA
```

Appendix III: Publications in peer-reviewed journals which included data from this work.

Kitsera, N., Rodriguez-Alvarez, M., Emmert, S., Carell, T., & Khobta, A. (2019). Nucleotide excision repair of abasic DNA lesions. *Nucleic Acids Research*, *47*(16), 8537–8547. <https://doi.org/10.1093/nar/gkz558>

Rodriguez-Alvarez, M., Kim, D., & Khobta, A. (2020). EGFP reporters for direct and sensitive detection of mutagenic bypass of DNA lesions. *Biomolecules*, *10*(6), 1–19. <https://doi.org/10.3390/biom10060902>

Index of figures

Figure 1-1: Chemical structure of abasic DNA lesions.	8
Figure 1-2: Abasic site formation and repair through mammalian short- and long-patch BER pathways.	10
Figure 1-3: DTT regulation by post-translational modifications of PCNA	13
Figure 1-4: Schematic representation of EGFP reporter vector.	19
Figure 2-1: Cloning procedure followed in this report.	31
Figure 2-2: Indexes combinations for second PCR amplification.	43
Figure 2-3: Electropherogram profiles of two representative samples obtained by the Agilent Bioanalyzer	44
Figure 2-4: MiSeq sequencing by synthesis from Illumina.	45
Figure 3-1: Schematic representation of pZAJ_5C vector encoding for EGFP protein.	47
Figure 3-2: Schematic representation of the novel screening strategy followed in this section.	50
Figure 3-3: Verification of nicking reaction of the pZAJ_5c at both Nb.Bpu10I sites and further generation of a gap in the template strand (TS).	51
Figure 3-4: Verification of the insertion of 25 synthetic oligonucleotides into the template strand gapped pZAJ_5c vector.	52
Figure 3-5: Phenotypic screening of mismatches with the indicated nucleotide substitutions in the TS to identify potential point mutations that would lead to loss of EGFP fluorescence.	54
Figure 3-6: Sanger sequencing of cloned vectors.	55
Figure 3-7: Phenotypic characterisation of the specified EGFP mutant vectors.	56
Figure 3-8: Verification of insertion of the synthetic oligonucleotides carrying all possible single mismatches into four different mutant reporters.	57
Figure 3-9: Characterisation of all possible nucleotide substitutions at the mutated position of pZAJ_Q205*, pMR_Q205P, pMR_S206Y and pMR_A207P.	58
Figure 3-10: Generation and characterisation of pMR_S206* mutant reporter.	60
Figure 3-11: Schematic representation of pZAJ_Q205* and pMR_A207P reporters encoding for the non-fluorescent EGFP mutants.	62
Figure 3-12: Scheme of preparation of constructs for TM assay.	62
Figure 3-13: Abasic DNA lesions and linkages used to analyze TM.	63
Figure 3-14: Scheme of the pZAJ_Q205* vector used for transcriptional mutagenesis (TM) assays.	64
Figure 3-15: Preparation of constructs with THF and S-THF modifications in position 613 of the template strand of the pZAJ_Q205* mutant.	65
Figure 3-16: APE1 digestion assays of ts.613A, THF and S-THF constructs derived from pZAJ_Q205* reporter.	66
Figure 3-17: Transcriptional mutagenesis at BER-Resistant AP sites (S-THF) reactivates EGFP fluorescence in position ts.613 in HeLa cells.	67
Figure 3-18: Scheme of the mutant pMR_A207P used for transcriptional mutagenesis (TM) assays.	68
Figure 3-19: Preparation of constructs with THF and S-THS in position ts.619 of the pMR_A207P reporter.	69
Figure 3-20: Transcriptional mutagenesis at BER-Resistant AP sites (S-THF) reactivates EGFP fluorescence in position ts.619 in HeLa cells.	70
Figure 3-21: Transcriptional mutagenesis triggered by BER-resistant AP sites in NER deficient cell lines.	72

Figure 3-22: Transcriptional mutagenesis triggered by BER-resistant AP sites in NER deficient cell lines.....	73
Figure 3-23: Preparation and characterisation of constructs with variations of uracil (U) modification in position ts.613 of the pZAJ_Q205* mutant reporter.	74
Figure 3-24: Analysis of the sensitivity of various types of AP lesions towards Endo IV, Endo III and Fpg activities.....	76
Figure 3-25: Preparation of constructs with natural AP sites in position ts.613 of the pZAJ_Q205* mutant vector.....	77
Figure 3-26: Relative EGFP expression triggered by BER-resistant natural AP sites constructs in NER proficient (MRC5 and XP-A+XPA) and NER deficient (XP-A) cells.78	78
Figure 3-27: Transcriptional mutagenesis triggered by BER-resistant natural AP sites is undetectable in all cell lines tested.	79
Figure 3-28: Scheme of the pZAJ_5c reporter vector.	79
Figure 3-29: Preparation of constructs with variations of U in position ts.613 of the pZAJ_5c vector.	80
Figure 3-30: Verification of the presence of natural AP sites in position ts.613 of the pZAJ_5c reporter..	80
Figure 3-31: EGFP expression of human cells transfected with ts.613dG/AP/sAP/sAPs opposite to C constructs.....	81
Figure 3-32: Transcriptional mutagenesis triggered by uracil (ts.613U) in NER deficient (XP-A) and proficient (XP-A/+XPA) cell lines..	82
Figure 3-33: Transcriptional mutagenesis triggered by ribonucleotide abasic DNA lesions in NER deficient (XP-A) and proficient (XP-A/+XPA) cell lines.	84
Figure 3-34: Scheme depicting the preparation of constructs for detection of mutagenic TLS.....	85
Figure 3-35: Generation of constructs with site-specific THF AP sites in the coding strand of the pZAJ_Q205* mutant vector..	87
Figure 3-36: Analytical ligation of constructs with modifications in position 613 of the non-template strand (NTS) opposite to an 18-nucleotide gap into the template strand.	88
Figure 3-37: Analyses of APE1 activity towards AP lesion (THF) in ss-DNA and ds-DNA.....	88
Figure 3-38: Reporter reactivation assay for direct detection of mutagenic TLS using pZAJ_Q205* as a model vector..	89
Figure 3-39: Detection of mutagenic TLS on the gapped pZAJ_Q205* template containing THF and S-THF at position c.613 transfected to HeLa cells.....	90
Figure 3-40: Detection of mutagenic TLS on the DNA template containing THF AP lesion in MRC5, XP-A, and XP-A/+XPA cell lines.....	91
Figure 3-41: Preparation of constructs with uracil in position c.613 of the pZAJ_Q205* reporter opposite to a 18-nt gap.....	93
Figure 3-42: Detection of mutagenic TLS on DNA templates containing THF and U in HeLa and UNGsh cell-lines by FACS analysis..	94
Figure 3-43: Preparation of constructs with U and FU opposite to dA in double-stranded DNA.	95
Figure 3-44: Detection of mutagenic TLS on the DNA template containing U and FU lesions in the HeLa cell line.....	96
Figure 3-45: Overview scheme of all mutant vectors used for translesion synthesis (TLS) assays.....	97
Figure 3-46: Generation of constructs with site-specific THF AP sites in the coding strand of all mutant vectors.....	98

Index of figures

Figure 3-47: Detection of mutagenic TLS on the DNA template containing THF in 5 different positions.....	99
Figure 3-48: Detection of mutagenic TLS on the DNA template containing THF in 5 different positions in HAP cells.....	101
Figure 3-49: Generation of constructs with site-specific uracil sites in the coding strand of all mutant vectors.....	102
Figure 3-50: Detection of mutagenic TLS on the DNA template containing U in 5 different positions in HAP cells.....	103
Figure 3-51: Detection of mutagenic TLS on the DNA template containing synthetic and physiological AP lesions in HAP cell line.	105
Figure 3-52: Detection of mutagenic TLS over synthetic and natural AP sites using pZAJ_Q205* mutant in HAP and Rad18KO cells.	106
Figure 3-53: Detection of mutagenic TLS over synthetic and natural AP sites using pMR_A207P mutant in HAP and Rad18KO cells..	107
Figure 3-54: Detection of mutagenic TLS over synthetic AP sites using pMR_Q205* mutant in HAP and Y-family polymerases KOs cells.	109
Figure 3-55: Experimental approach to create RNA sequencing library.	110
Figure 3-56: 1% agarose RNA electrophoresis gel of RNA extracted from HAP cell-line transfected with c.613T/THF/U/FU constructs.....	111
Figure 3-57: Representative analytical PCR of cDNA diluted samples from c.613T/THF/U/FU constructs transfected into HAP cells.	112
Figure 3-58: Scheme and validation of 1 st PCR amplification..	112
Figure 3-59: RNA sequencing results of ts.619G/THF/S-THF constructs transfected in XP-A cells.	115
Figure 3-60: S-THF libraries showing % of nucleotide substitution and deletions from the nucleotide 611-632..	116
Figure 3-61: RNA sequencing results of ts.617S-THF constructs transfected in XP-A cell-line..	118
Figure 3-62: RNA sequencing results of T/THF/U/FU constructs for TLS assay transfected into HAP cell-line..	121
Figure 3-63: RNA sequencing analysis of c.617THF construct in the double mutated plasmid pMR_S206*..	122
Figure 3-64: c.617 THF libraries showing % of nucleotide substitution and deletions from the nucleotide 601-629..	123
Figure 3-65: Design of constructs with double mutation for detection of mutagenic TLS over natural AP sites.....	125
Figure 3-66: FACS analysis and RNA sequencing results of c.612G-c.613T/U/FU constructs for TLS assay transfected into HAP cell-line.	126
Figure 4-1: Scheme of the enhanced green fluorescent protein (EGFP) Q205* vector and its derived constructs for transcriptional mutagenesis (TM) and mutagenic translesion DNA synthesis (TLS) assays.	131
Figure 4-2: Model of transcriptional outcomes of abasic DNA lesions placed in the template strand of a gene..	137
Figure 4-3: Model illustrating the bypass of abasic DNA lesions during DNA synthesis in human cells.....	144

CURRICULUM VITAE

CURRICULUM VITAE**Personal Data**

Name	Marta Rodríguez Alvarez
Date and place of birth	27 th of May, 1990 in Oviedo (Spain)
Nationality	Spanish

Research Work

October 2016 – Present	PhD Molecular Genetics at Institute of Toxicology of Mainz
March 2016 – July 2016	Research Assistant at the Institute of Molecular Biology (IMB) of Mainz
Dec 2014 – Sept 2015	Internship Molecular Genetics in CABIMER Institute, Seville

Academic Background

2014-2015	MSc Molecular Genetics and Biotechnology at Sevilla University
2008-2014	BSc Biology at University of Oviedo

Presentations at national and international conferences

Date	Conference	Presentation	Additional notes
September 2017 <i>Cologne</i>	German-French DNA Repair Meeting	Poster	
September 2018 <i>Karlsruhe</i>	DGDR DNA Repair meeting	Poster	Poster Award
September 2018 <i>Mainz</i>	UCT Science Day	Poster	Poster Award
September 2019 <i>Mainz</i>	UCT Science Day	Talk	Talk Award
November 2020 <i>Online</i>	German-French DNA Repair Meeting	Poster	

DECLARATION OF AUTHORSHIP

I hereby confirm that I have authored this Ph.D. thesis independently and without the use of anything other than the indicated resources.

Mainz, July 20, 2021

Marta Rodríguez Alvarez