PhD degree in Molecular Medicine (curriculum in Molecular Oncology)
European School of Molecular Medicine (SEMM)
University of Milan and University of Naples “Federico II”
Settore disciplinare: BIO/11

Insight from AID-induced DNA damage resolution: cellular context matters

Elisabetta Incorvaia
IFOM, Milan
Matricola n R09399

Supervisor: Svend K. Petersen-Mahrt
IFOM, Milan

Anno accademico 2013-2014

"*Nothing in biology makes sense except in the light of evolution*"

Theodosius Dobzhansky

INDEX

LIST OF ABBREVIATIONS 6

FIGURES INDEX 10

ABSTRACT 13

CHAPTER 1 - INTRODUCTION 14

1.1 - GENOME STABILITY PERTURBATION AND SURVEILLANCE MECHANISMS 14
1.1.1 - DNA spontaneous instability 14
1.1.2 - DNA environmental instability 16
1.1.3 - DNA damage response 17

1.2 - DNA REPAIR MECHANISMS 18
1.2.1 - Base excision repair 19
1.2.2 - Mismatch repair 22
1.2.2.1 - Canonical Mismatch repair 22
1.2.2.2 - Non canonical Mismatch repair 24
1.2.3 - Other repair pathways 25
1.2.4 - Crosstalk between DNA repair pathways 27

1.3 - HOMEOSTATIS: THE BALANCE BETWEEN INJURY AND REPAIR 28

1.4 - ALTERATIONS OF THE HOMEOSTATIS 30
1.4.1 - Alteration of the balance for bad 30
1.4.1.1 - Alteration of the balance leading to cancer 30
1.4.1.2 - Alteration of the balance in inherited diseases 32
1.4.2 - Alteration of the balance for good 32
1.4.2.1 - Alteration of the balance in meiosis 33
1.4.2.2 - Alteration of the balance in *Ig locus* diversification 33

1.5 - THE NATURAL DNA MUTATOR AID 36
1.5.1 - AID in immunity 36
1.5.2 - AID outside of the immune system 39
1.5.2.1 - AID in epigenetic reprogramming 39
1.5.2.2 - AID in cancer 40

1.6 - CURRENT STATE 42

AIM 43

CHAPTER 2 - MATERIALS AND METHODS 44

2.1 - MATERIALS 44
2.1.1 - Buffers and reagents 44
2.1.2 - Plasmids 46
2.1.3 - Oligonucleotides 47
2.1.4 - Antibodies 48
2.1.5 - Chemical inhibitors 48

2.2 - METHODS 49
2.2.1 - Cell culture 49
2.2.2 - Cell staining 50
2.2.3 - In vitro transcription assay 51

2.2.4 - Interphase frog egg extract preparation 51
2.2.5 - Cloning methods 53
2.2.6 - G-AID expression and purification 55
2.2.7 - Western Blot 56
2.2.8 - Coomassie 56
2.2.9 - Gel mobility assay 57
2.2.10 - Rifampicin assay 57
2.2.11 - ssDNA oligonucleotide deamination assay 58
2.2.12 - dsDNA oligonucleotide DNA glycosylases assay 58
2.2.13 - In vitro resolution assay (IVR) 59
2.2.14 - RT-PCR and statistical analysis 59
2.2.15 - Atomic Force Microscopy 60
2.2.16 - In vitro nucleosomes reconstitution 62

CHAPTER 3 - RESULTS 63

3.1 - DEVELOPMENT AND VALIDATION OF A NEW BIOCHEMICAL ASSAY TO STUDY G-AID INDUCED LESIONS RESOLUTION IN FE **63**

3.2 - VALIDATION OF THE IVR ASSAY IN FE **65**
3.2.1 - G-AID protein production optimization leads to the purification of an endonuclease free protein 65
3.2.1 - G-AID protein retains its catalytic activity 68
3.2.1 - G-AID-independent DNA repair activation is excluded from IVR readout 69

3.3 - KINETIC PROFILES OF AID-INDUCED DAMAGE AND FE DEPENDENT REPAIR **70**

3.4 - IVR IS SUITABLE FOR QUANTIFYING DNA REPAIR ACTIVATION IN DIFFERENT CELL LINES **74**

3.5 - ANALYSIS OF SINGLE REPAIR PATHWAYS ACTIVATED BY DIFFERENT CELL LINES UPON AID-INDUCED DAMAGE **76**

3.6 - MMR AND BER PROTEINS ARE DIFFERENTIALLY EXPRESSED IN B CELL EXTRACTS AND NON-B CELL EXTRACTS **85**

3.7 - ANALYSIS OF BER PROTEINS ACTIVITY IN DIFFERENT CELL EXTRACTS **86**
3.7.1 - UNG activity is dominant over the other DNA glycosylases activities 88
3.7.2. - Number of active DNA glycosylases molecules does not directly correlate with BER activity 90

3.8 - TRASCRIPTIONAL ACTIVITY DOES NOT SEEM TO INFLUENCE IVR READOUT **94**

3.9 - CELL CYCLE STATE CONTRIBUTION TO DNA REPAIR PROTEIN AVAILABILITY IN THE RESOLUTION OF AID-INDUCED DAMAGE **95**

3.10 - B CELL EXTRACTS ARE ENRICHED IN HISTONES COMPARE TO NON-B CELL EXTRACTS **97**

3.11 - ANALYSIS OF DNA SUBSTRATES ABLE TO ALTER THE CHROMATIN COMPACTION **99**
3.11.1 - pEI7150 has regions reluctant to form nucleosomes and pEI7151 has regions prone to form nucleosomes 101

3.12 - NUCLEOSOME FORMATION ALTER DNA REPAIR PROTEIN RECRUITMENT UPON AID-INDUCED DAMAGE **105**

3.13 - PROTEIN INTERFERING WITH AID CAN MODIFY ITS ACTIVITY AND INFLUENCE REPAIR PROTEIN RECRUITMENT **109**

CHAPTER 4 - DISCUSSION 111

4.1 - ESTABLISHMENT OF AN *IN VITRO* APPROACH TO STUDY DNA LESION RESOLUTION 111

4.2 - DIFFERENT CELL LINES ACTIVATE DIFFERENT REPAIR SYSTEMS UPON AID-INDUCED DNA LESION 114

4.3 - DNA GLYCOSYLASES DISPLAYED DIFFERENT ACTIVATION KINETICS IN DIFFERENT CELL LINES 116

4.4 - DNA SEQUENCE SPECIFICITY INFLUENCE AID-INDUCE LESION RESOLUTION 117

4.4 - FUTURE PLANS 120

4.5.1 - IVR future plans 120

4.5.2 - Analysis of G-AID efficiency and G-AID function in DNA repair recruitment 121

CHAPTER 5 - REFERENCES 123

AKNOWLEDGEMENTS 132

LIST OF ABBREVIATIONS

4PPs	6–4 photoproducts
8-oxo-G	8-hydroxiguanine
AID	activation-induced cytidine deaminase
AFM	atomic force microscopy
AP	apurinic/apyrimidinic
Aphid	aphidicolin
APE1	AP endonuclease 1
AT	ataxia telangiectasia
ATM	ataxia-telangiectasiamutated
ATR	ataxia-telangiectasia and Rad3-related
BER	base excision repair
CHK1	checkpoint kinase 1
CHK2	checkpoint kinase 2
CNS	central nervous system
CO	crossing-over
Co	concatenated DNA
CPDs	cyclobutane–pyrimidine dimers
CS	Cockayne syndrome
CSR	class-switch recombination
D	diverse segment
dA	adenosine
dC	cytosine
dG	guanine
dT	thymine
dU	uracil
DDR	DNA damage-response
DDT	DNA damage tolerance

DNA-PK	DNA-dependent protein kinase
dRP	*5*'-deoxyribose phosphate
DSBs	double strand breaks
EXO1	exonuclease 1
FEN1	flap endonuclease 1
FE	frog egg extract
G	GAL4 DNA binding domain
GC	germinal centre
GGR	global genome repair
HIV-1	human immunodeficiency virus type 1
HNPCC	hereditary non-polyposis colorectal cancer
HPV	human papillomavirus
HN	high nucleosome forming region
HR	homologous recombination
Ig	immunoglobulin
IgH	heavy chain locus
IgL	light chain locus
IVR	*in vitro* resolution
J	joining segment
L	Linear DNA
Lig1	DNA ligase 1
Lig3α	DNA ligase 3α
LP-BER	long-patch BER
LN	low nucleosome forming region
MBD4	methyl-CpG binding domain protein 4
MMR	mismatch repair
MO	microorganism
MSI	microsatellite instability

mt	mutant
ncMMR	non canonical MMR
N	nicked DNA
NES	nuclear export signal
NER	nucleotide excision repair Pol μ
NHEJ	non-homologous end joining
NLS	nuclear localization signal
Oc	open circular DNA
p53	tumor protein 53
PARP-1	poly (ADP-ribose) polymerase 1
PARP-2	poly (ADP-ribose) polymerase 2
PCNA	proliferating cell nuclear antigen
PGCs	primordial germ cells
PNK	DNA polynucleotide kinase
Polβ	DNA polymerase β
Polδ	DNA polymerase δ
Polε	DNA polymerase ε
Pol-η	DNA polymerase μ
Pol-η	DNA polymerase η
PTM	post-translational modification
RF-C	replication factor C
RNA pol	RNA polymerase
ROS	reactive oxygen species
RPA	replication protein A
RSS	recombination signal sequences
SC	supercoiled
SHM	somatic hypermutation
SMUG1	single-strand selective monofunctional uracil DNA glycosylase

SP-BER	short-patch BER
SSBs	single strand breaks
SSBR	single-strand break repair
ssDNA	single-stranded DNA
STM	SMUG1, TDG, MBD4 glycosylases
TCR	transcription coupled repair
TDG	thymine DNA glycosylase
TLS	translesion synthesis
TTD	trichothiodystrophy
UGI	uracil-DNA glycosylase inhibitor
UNG	uracil DNA glycosylase
UV	ultraviolet
V	variable segment
WT	wild type
XP	xeroderma pigmentosum
XRCC1	X-ray repair cross-complementing protein

FIGURES INDEX

Figure 1 - DNA damage and DNA repair specificity 19

Figure 2 - Schematic illustration of BER subpathways 21

Figure 3 - Schematic representation of MMR variants 23

Figure 4 - *Ig locus* organization and protein structure 34

Figure 5 - Schematic drawing of SHM and CSR 35

Figure 6 - Reaction catalysed by AID 37

Figure 7 - Model for lesion-induced DNA demethylation 40

Figure 8 - Schematic drawing of plasmids 46

Figure 9 - Drawing of basic principles of Atomic Force Microscopy (AFM) 61

Figure 10 - Schematic representation of the IVR assay 64

Figure 11 - Quantitative and qualitative analysis of G-AID purified protein 66

Figure 12 - DNA topology following G-AID induced-damage 67

Figure 13 - G-AID deaminates cytosine to uracil in E. coli and in vitro 68

Figure 14 - Quantification of FE replication dependent repair 69

Figure 15 - Quantification of SSBR in FE 70

Figure 16 - AID-induced lesions repair in FE 71

Figure 17 - G-AID and FE titration in IVR assay 73

Figure 18 - Quantification of SSBR in different cellular extracts 75

Figure 19 - Extracts competence for IVR assay 75

Figure 20 - IVR assay is suitable for studying AID-induced damage resolution in different cell lines 77

Figure 21 - Tabular representation of the strategy used to identify single repair pathways involvement in AID-induced damage repair 79

Figure 22 - Quantification of single repair pathway activation upon AID damage in IVR assay using different cellular extracts 81

Figure 23 - Quantification of single repair pathway activation upon AID damage in IVR assay using FE 82

Figure 24 - Detailed analysis of single repair pathways contribution to maximum IVR activity 83

Figure 25 - BER and MMR protein expression in different cell lines 85

Figure 26 - ds oligomucleotide DNA glycosylases assay is suitable to evaluate DNA glycosylases activity in different cellular extracts 87

Figure 27 - UGI decreases DNA glycosylases activity in ds oligonucleotide DNA glycosylases assay 89

Figure 28 - UGI inhibition on not completely active UNG molecules from Colo704 extract 90

Figure 29 - Quantification of DNA glycosylases activity, concentration and turnover in different cell extracts 91

Figure 30 - Integrated analysis of single repair pathways contribution to maximum IVR activity 93

Figure 31 - Transcriptional competence of extracts 95

Figure 32 - Analysis of nocodazole efficacy and toxicity 96

Figure 33 - Analysis of HelaS3 and Ramos synchronisation 97

Figure 34 - Analysis of histones expression in different cell lines 99

Figure 35 - Chicken variable and constant regions analysis for affinity with nucleosomes formation 100

Figure 36 - Different plasmids topology and nucleosome formation visualization with AFM 102

Figure 37 - LN region topology in the presence of histone octamers 103

Figure 38 - HN region topology in the presence of histone octamers 104

Figure 39 - Plasmids with different affinities for nucleosome formation differentially repaired AID-induced lesion 106

Figure 40 - DNA sequence composition influences both G-AID deamination activity and IVR repair activity 108

Figure 41 - PCNA interfere with G-AID induced DNA resolution 110

Figure 42 - Multiple-model of how cellular microenvironment influences AID-induced NDA damage repair 119

TABLES

Table 1 - Spontaneous DNA damages occurring every day 15

Table 2 - Mammalian DNA glycosylases substrate specificities 20

Table 3 - Exogenous DNA damaging agents associated with increased risk of cancer 30

Table 4 - Inherited mutations in DNA repair genes that increase the risk of cancer 31

Table 5 - Oligonucleotides length, orientation and applications 47

Table 6 - Antibodies, manufacturers and concentration used 48

Table 7 - Chemical inhibitors and manufacturers 48

ABSTRACT

When genome integrity is perturbed, surveillance and repair mechanisms are activated to restore genome integrity through high fidelity DNA repair. However, in some physiological situations, those mechanisms are channeled away from integrity towards mutations and recombinations. During the diversification of the immunoglobulin *locus* in B cells, Activation Induced Deaminase (AID) triggers the physiological introduction of mutations. The current work was based on the observation that upon simple lesion generation, AID-induced deamination of a cytosine to uracil, the resolution by the molecular mechanism of DNA repair can lead to different outcomes. This homeostatic outcome, error-free or error-prone, is governed by specific cellular context and processes associated with DNA. To uncover the regulation of the pathway choice an *in vitro* system, named *in vitro* resolution (IVR), was developed. In the 1st phase of the IVR, AID was targeted to a DNA plasmid for uracil lesion generation. In the 2nd phase, a cellular extract resolved the lesions via Base Excision Repair [BER, divided in short patch (SP)-BER or long patch (LP)-BER] or Mismatch Repair (MMR). The quantitative nature of the IVR provided a novel means to precisely quantitate the contribution of each single DNA repair pathway. This set-up allowed us to evaluate how different cellular environments influenced the choice. Cell origins presented quantitative differences in DNA repair kinetics: a) overall sensitivity, b) non-B cells activating non-canonical MMR first, c) B cells activating SP-BER first, and d) LP-BER is significantly activated only in B cells. To understand the possible molecular mechanisms, we analysed single components known to influence DNA repair, such as transcription, protein availability, and chromatin. Changing the DNA substrate to either prefer or avoid forming nucleosomes, we uncovered significant changes in AID deamination preference and in DNA repair pathway choice. DNA with nucleosome favourable base-stacking preferred LP-BER, while non-nucleosome stacked DNA preferred SP-BER and MMR. Overall our findings provide novel insight into the cellular context that can influence DNA repair. The use of B cells and cancer cell lines can recapitulate in vivo *Ig locus* diversification, and our findings have a direct bearing in understanding mechanisms of tumorigenesis.

CHAPTER 1 - INTRODUCTION

1.1 - GENOME STABILITY PERTURBATION AND SURVEILLANCE MECHANISMS

DNA is a fragile molecule and its stability depends on fine interactions between bases, such as base pairing between complementary strands and base-stacking interactions between adjacent bases. Already in 1953 Watson and Crick proposed that adenine and thymine (dA and dT) are held together by two hydrogen bonds, while guanine and cytosine (dG and dC) are involved in three hydrogen bonding. These bonding are necessary to hold the two strands of DNA together. The spacing between parallel base pairs is stabilized by base-stacking forces: adjacent base pairs attract one another through non-covalent forces, while due to the hydrophobic interactions the polar surfaces of the bases are exposed to the surrounding environment.

In mammalian cells the very long linear DNA molecule is compacted for protection and regulation when it is associated with proteins – forming the nucleosome: 147 base pairs are wrapped around eight histone molecules, forming the basic repeat unit of chromatin. Various interconnected mechanisms regulate DNA accessibility, stability and dynamics, with slight variations in the chemistry of the DNA bases having profound implications on the stability of the DNA duplex.

1.1.1 - DNA spontaneous instability

Although DNA carries the genetic information, it can be considered fragile because it is chemically instable and susceptible to damage. Modifications such as hydrolysis, oxidation and alkylation occur at significant rates *in vivo*, increasing DNA instability (1) (Table1).

DNA has N-glycosyl bonds susceptible to spontaneous hydrolysis. In particular purines are lost more rapidly than pyrimidines (2). Moreover, in single-stranded DNA (ssDNA) cytosine residues are susceptible to hydrolytic deamination, which lead to conversion into uracil (dU). The double helix structure, through specific hydrogen bonding patterns, provides a good protection against this spontaneous reaction (3).

Another chemical feature of DNA that contributes to its instability is the presence of oxygen and nitrogen atoms in its bases, which can be targets of reactive oxygen species (ROS), which are produced during normal cellular metabolism. The major base lesion generated by ROS oxidation is 8-hydroxiguanine (8-oxo-G), which can base-pair with adenine rather than cytosine. Other reactive molecules can modify the double helix structure, one of them is S-adenosylmethione, which can mediate alkylation of guanine and adenine (4).

Spontaneous damages occurring every day in the genome of a cell
20000 ssDNA breaks
10000 depurinations
5000 alkylating lesions
2000 oxidative lesions
600 deaminations
50 sister chromatid cross

Table1 - Spontaneous DNA damages occurring every day

Other cellular processes like lipids peroxidation and cellular stress have a role in increasing the amount of modified DNA bases (5). The replication process itself can be a source of DNA instability, if a nucleotide with an incorrect base is incorporated in the nascent strand. Misincorporations occur every 10^6 to 10^8 nucleotides for replicative polymerases (6). DNA instability can also result from insertion or excision of (retro-) transposons. Furthermore, specific regions in the genome, hotspots, have a higher probability to host instability, because of their sequence composition: including DNA repeats, fragile sites, slow replication zones, G-quadruplexes and telomeres. For example, tandem repeats undergo long expansions during replication, due to their particular sequence composition (7), and fragile sites can be composed of several AT-rich repeats susceptible of breakage (8).

1.1.2 - DNA environmental instability

Besides endogenous instability, environmental agents, such as chemical or physical agents and microorganisms (MO), can induce specific adverse effects on genome stability. Ultraviolet (UV)-B radiation induces abundant mutagenic and cytotoxic DNA lesions, such as cyclobutane–pyrimidine dimers (CPDs) and 6–4 photoproducts (6–4PPs) (9) at dipyrimidine sites, where two pyrimidine bases are juxtaposed in the nucleotide sequence of DNA. Each CPD impairs DNA polymerases progression during replication. Furthermore, UV-A can cause DNA damage indirectly through the production of ROS.

Several DNA aberrations, such as base damages, single strand breaks (SSBs) and double strand breaks (DSBs) can be observed after exposure to ionizing radiations (10). X-rays can ultimately lead to dramatic genetic changes such as deletions, translocations and other chromosomal rearrangements.

Chemical mutagens are classified into different groups based on the mechanism by which they interact with DNA (11). For example, base analogues are chemicals that structurally resemble normal DNA bases, but often pair with the wrong bases, leading to base pair changes in the DNA. Frameshift mutagens can intercalate between bases in the same strand of the DNA, increasing the distance between the bases and producing helix distortions. Other chemicals can directly modify bases or bind covalently to DNA with subsequent addition of some organic group to the bases or to the sugar–phosphate backbone.

Exogenous DNA damaging agents can be found in toxic tobacco residues, air pollution, pesticides, toxic metals and food toxins. Toyooka *et al.* demonstrated that cigarette sidestream smoke can induce DSBs (12). While many other compounds were proven to induce DNA damage directly or indirectly through ROS generation: asbestos, silica, and titanium dioxide (13). Calderón-Garcidueñas *et al.* were able to detect ROS-induced SSBs in single cells of the nasal epithelium of patients exposed to air pollutants (14).

MO infection can also be an important source of DNA damage. *Helicobacter pylori*, through adhesion molecules, directly induces DSBs in infected cells in a time- and dose-dependent manner (15).

Similar, viruses can damage DNA: the human papillomavirus (HPV) E7 protein is able to inactivate pRB, leading to the accumulation of DSBs, and as a consequence genomic rearrangements, chromosomal abnormalities, DNA repair errors, and aneuploidy (16). The same molecular DNA damaging mechanism was demonstrated for human immunodeficiency virus type 1 (HIV-1) (17).

As a result of exposition of DNA to spontaneous and induced damages, every day each of our cells accumulates thousand of lesions. It is then important to distinguish between DNA damage, a physical abnormality in the DNA, and mutation, which is a change in the base sequence of the DNA. Not all DNA damages are converted into mutations. To ensure an accurate transmission of genetic information, cells have evolved a complex system of surveillance and DNA repair mechanisms, which prevent damaged DNA from being converted into inheritable mutations.

1.1.3 - DNA damage response

DNA damage-response (DDR) includes a set of DNA repair mechanisms, damage tolerance processes, and cell-cycle checkpoint pathways. To allow complete lesion resolution prior to replication or cell division, DDR is linked to an intricate signal transduction cascade that induces cell cycle arrest, hence providing time for DNA repair (18,19). Different types of DNA damages can activate different DDR signalling cascades. DDR is a multistep process composed of proteins that sense the damage and kinases, such as ataxia-telangiectasiamutated (ATM) (20) and ataxia-telangiectasia and Rad3-related (ATR), that activate effector proteins. For an efficient checkpoint response ATM and ATR kinases phosphorylate a large number of targets that participate in DNA repair, checkpoint signalling, and determine cell fate decision, such as apoptosis or senescence (21). Checkpoint kinase 1 and 2 (CHK1 and CHK2) coordinate the inhibition of cell cycle progression, ensuring the DNA repair machinery enough time to perform efficient repair before replication. Alternatively, when too many DNA injuries are encountered, apoptosis can be triggered by p53 to protect the organism from potential tumorigenic transformations (22). Another DDR kinase is DNA-dependent protein kinase (DNA-PK), which is also

activated by irradiation-induced DSBs and is essential for repair activation or induction of apoptosis (23,24).

DDR kinases control DNA repair at different levels. In particular, they can directly regulate DNA repair enzymes through post-translational modifications (PTM) or modify the chromatin near the lesion, creating the permissive local environment needed for repair (25).

Repair of the DSB, either by homologous recombination (HR) and non-homologous end joining (NHEJ) (discussed in chapter 1.2.4), signals the return of the cell to homeostasis (26): DNA repair proteins and chromatin modellers are removed from the DDR site and the checkpoint response is inactivated.

1.2 - DNA REPAIR MECHANISMS

The functional DDR response can be activated upon lesions that require cell cycle coordination for repair, but when DNA lesions can be rapidly cleared, single repair mechanisms are independently activated. Various DNA alterations (discussed in chapter 1.1) require detection by specific recognition proteins.

In general, although cross-talk exists, each repair mechanism is activated by a particular set of DNA damage (Figure 1). Principal repair mechanisms are: base excision repair (BER), mismatch repair (MMR), nucleotide excision repair (NER), HR and NHEJ. Furthermore, persisting lesions can still be processed to avoid stalled DNA polymerases via DNA damage tolerance mechanisms.

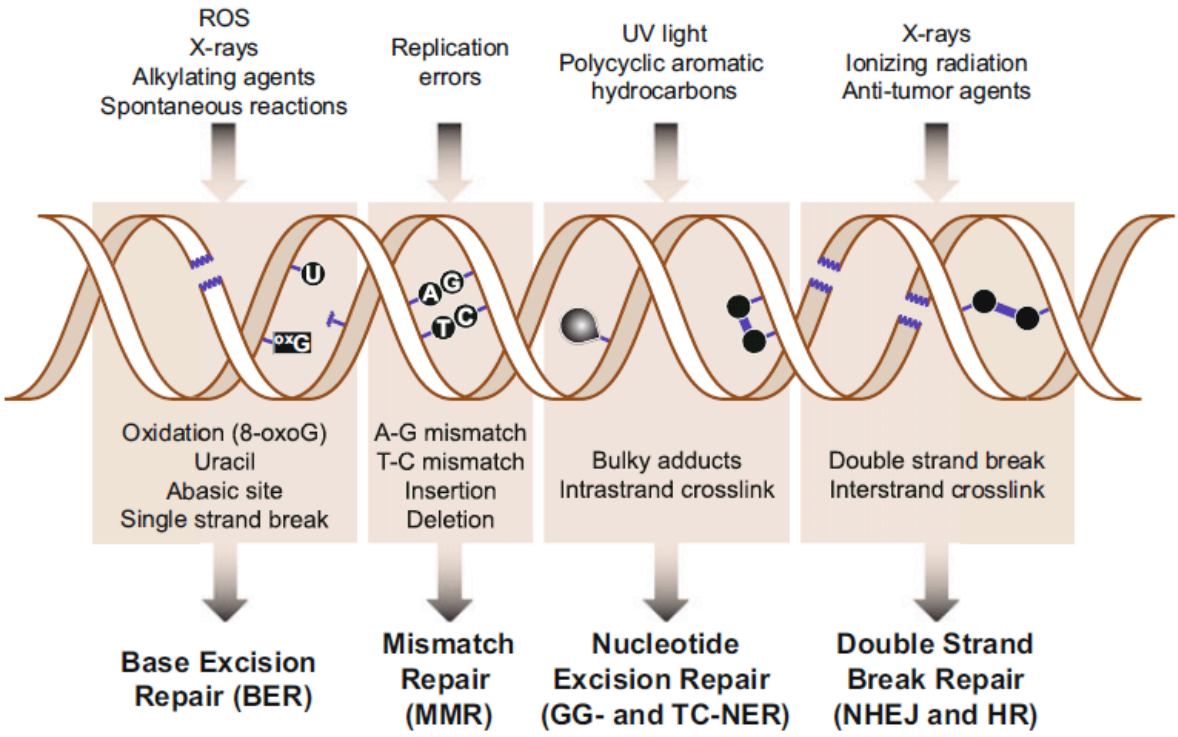


Figure 1 - DNA damage and DNA repair specificity

Endogenous (replication errors and spontaneous reactions) and environmental (ROS, X-rays, UV) DNA damage can result in: modified bases (oxidation or uracil), base pair mismatches, bulky adducts, SSBs and DSBs. Upon sensing DNA damage the cells activate the DNA repair machinery. Each type of lesion is repaired by biochemically distinct and highly specialized repair pathways, such as BER, MMR, NER and DSBR. DNA repair pathways can act independently or co-ordinate to repair DNA lesions.

1.2.1 - Base Excision Repair

BER is highly conserved from bacteria to humans and is active against the vast majority of endogenous DNA damages, including alkylations, oxidations, deaminations, depurinations and SSBs (Figure 1). The initial step in BER is the search and excision of a lesion along DNA molecules and it is carried out by DNA glycosylases. Currently, 11 mammalian glycosylases have been characterized and they differ from one to another with regard to substrate specificity (27). Glycosylases and their substrate specificities are summarized in Table 2.

DNA glycosylase	Major Substrate
UNG	uracil (dU)
SMUG1	dU, 5-hydroxymethyluracil
TGD	thymine (dT), dU, ethenocytosine
MBD4	dT and dU opposite guanine (dG), dT opposite O*6-methyl-guanine*
MUTYH	adenine (dA) opposite 8-oxo-G, 2-hydroxyadenine opposite dG
OGG1	8-oxo-G opposite dC, formamidopyrimidine dG
NTHL1	thymine glycol, formamidopyrimidine dG, *Dihydrouracil*, *5*-hydroxyuracil, *5-hydroxycytosine*
NEIL1	as NTH1 and formamidopyrimidine dA, 8-oxo-G
NEIL2	overlap with NTH1/NEIL1
NEIL3	unknown
MPG	*3*-methyladenine, hypoxanthine, ethenoadenine

Table 2 - Mammalian DNA glycosylases substrate specificities

Modified bases or base adducts that are recognized by specific DNA glycosylases.

DNA glycosylases can be either monofunctional, if they catalyze only the hydrolytic removal of the base to form an apurinic/apyrimidinic (AP) site or bifunctional if they can also cleave the DNA 3' to the base that has been removed (28,29). The first step of processing of the AP site was discovered by Lindhal in 1974: the DNA incision is made by AP endonuclease 1 (APE1), which leaves a sugar attached to the 5' of the nick ((5'-deoxyribose phosphate (dRP)). The resulting 3' hydroxyl is a substrate for DNA polymerase β (Pol β), which has also a lyase activity to remove the dRP. Once Pol β filled the single gap with its DNA synthesis activity, DNA ligase 3α (Lig3α), with the cofactor scaffolding protein X-ray repair cross-complementing protein 1 (XRCC1), can seal the nick (30). This pathway is called short-patch BER (SP-BER) (Figure 2).

An alternative mechanism was described by Dogliotti *et al*. and is termed long-patch BER (LP-BER) (31). LP-BER is activated upon oxidation/reduction of sugar bases, when Pol β cannot remove the dRP (32), thus removal and replacement of 2-10 bp is needed. Other studies suggest that the switch between SP-BER and LP-BER is modulated by the ATP concentration near the AP site (33). In LP-BER the DNA polymerases δ and ε (Pol δ/ε) in combination with proliferating cell nuclear antigen (PCNA) and replication factor C (RF-C) can displace the strand 3' to the nick and synthetize up to 10 nucleotides. Flap endonuclease 1 (FEN1) cuts the displaced strand and finally ligation is mediated by DNA ligase 1(Lig1)(31) (Figure 2).

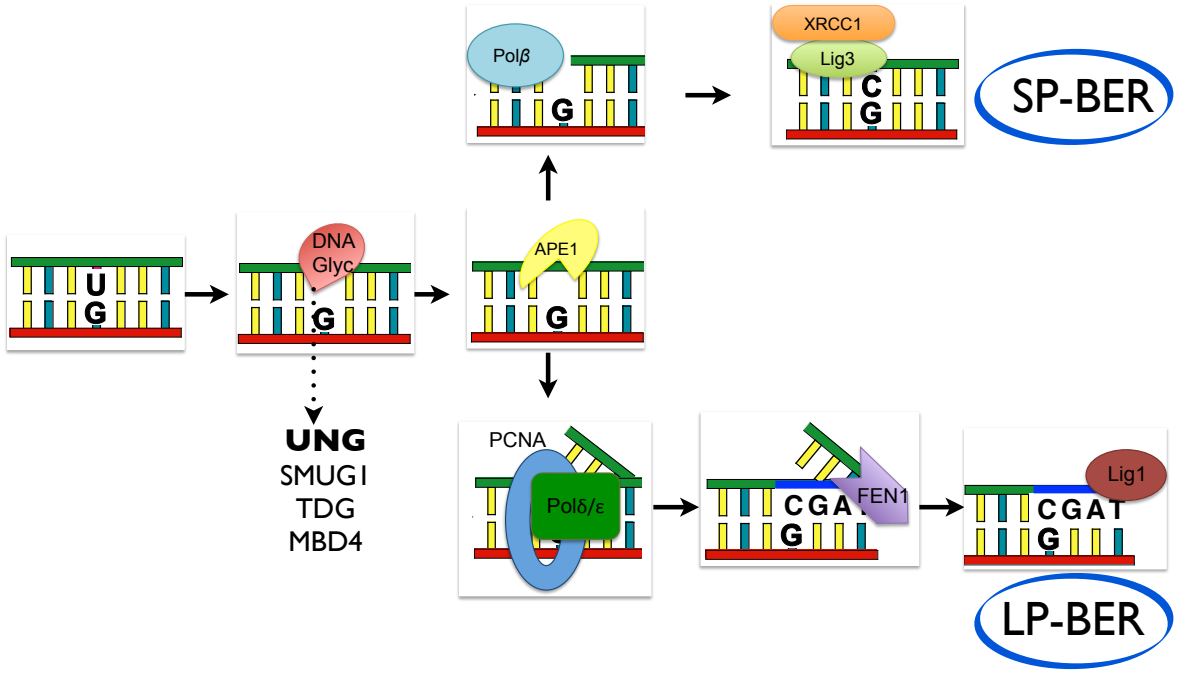


Figure 2 - Schematic illustration of BER subpathways

BER is normally initiated by DNA glycosylases, which specifically recognize and excise the damaged base. BER is composed of two sub-pathways: short-patch BER (top), where only one nucleotide is replaced thanks to the activity of APE1, Pol β and Lig3 and long-patch BER (bottom), where 2-13 nucleotides are replaced. In LP-BER Pol δ/ε are kept on place by PCNA, which promotes the synthesis of several bases (represented as a blue line), while the original bases are cleaved off by FEN1.

Alternatively, SSBs can be repaired by a specialized BER mechanism, named single-strand break repair (SSBR). Poly (ADP-ribose) polymerase 1 and 2 (PARP-1 and PARP-2) are activated upon SSB and synthetize chains of poly (ADP-ribose) on to proteins, including auto-polymersiation. This leads to the recruitment of XRCC1 and Lig3α, where they form a molecular scaffold for subsequent recruitment of DNA polynucleotide kinase (PNK), APE1 and Pol β (34). Damaged 5' or 3' termini are converted by APE1 or PNK, which possess both 5' DNA kinase and 3' DNA phosphatase activities that generate 5' phosphate and 3' hydroxyl DNA ends for DNA synthesis. Pol β enables single nucleotide gap filling, followed by ligation by Lig3α. PARP-1 has been shown to stably interact with specific DNA glycosylases (35,36) suggesting possible intercommunication between different BER steps and pathways.

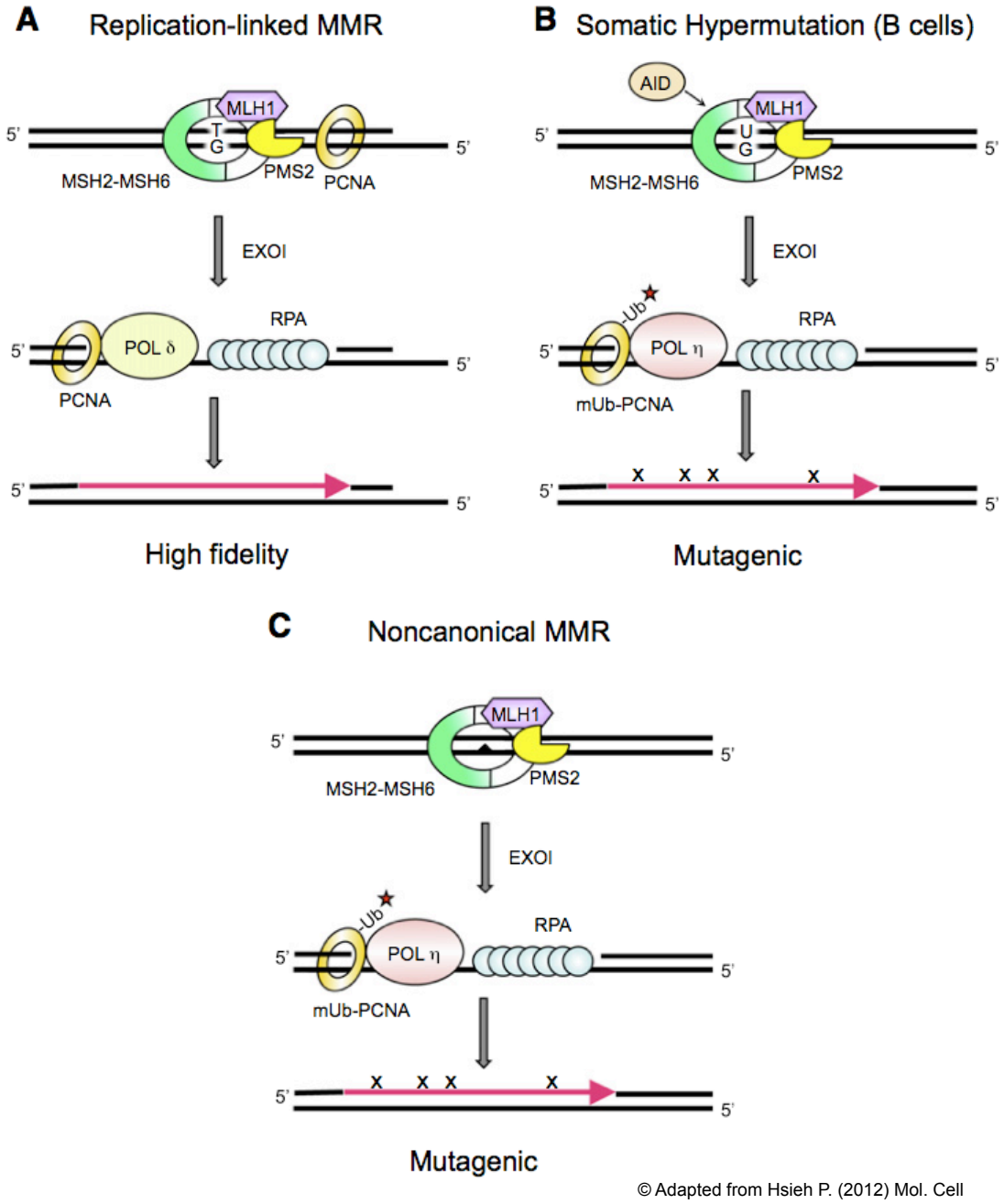
1.2.2 - Mismatch Repair

1.2.2.1 - Canonical Mismatch Repair

MMR corrects DNA oxidation damages as well as mismatches, insertions, and deletions generated during DNA replication (Figure 1). Inactivation of MMR proteins in human cells lead to an increase in the number of spontaneous mutations and is associated with human cancers (37). MMR is highly conserved with orthologous pathways between humans and *E. coli* (38). MMR pathway can be divided into three steps: 1) recognition, 2) excision, where the region surrounding the mismatch is removed, and 3) repair synthesis, where the gap is filled by the DNA polymerases (Figure 3A).

The mismatch recognition proteins are the ATPases hMutSα and hMutSβ, composed of MSH2 forming heterodimers with MSH6 or MSH3, respectively (39). 80 – 90 % of the cellular MSH2 is complexed as hMutSα, which preferentially recognizes base-base mismatches, while hMutSβ preferentially recognizes larger mispairs (small insertion or deletion). The mismatch-recognition proteins recruit MLH1, which in turn can form a heterodimer with PMS2, PMS1, or MLH3 (39). hMutLα, which is composed of MLH1 and PMS2, accounts for 90 % of the MLH1 in human cells (40), and is necessary to activate exonuclease mediated degradation of the damaged strand.

This degradation is performed by exonuclease 1 (EXO1) through its 5’ to 3’ exonuclease activity (41). Degradation is initiated by either a 5’ encision, which is directly preformed by EXO1 in the presence of hMutSα and RPA, or by a 3’ nicking, requiring PCNA and RFC to activate hMutLα endonuclease activity (42). PCNA interacts with hMutLα via a conserved PCNA motif, termed PIP box (43). EXO1 induced resection from the nick to the lesion can extend for up to 2000 bases.



© Adapted from Hsieh P. (2012) Mol. Cell

Figure 3 – Schematic representation of MMR variants

A) Base mismatch replication errors are recognized by the MSH2–MSH6 heterodimer, which recruits the MLH1–PMS2 heterodimer to initiate EXO1-mediated degradation of the damaged strand. Non-ubiquitinated PCNA mediates high-fidelity resynthesis recruiting DNA Pol δ. B) During somatic hypermutation (SHM) in B cells, PCNA is mono-ubiquitinated at lysine 164 and recruits DNA Pol η, which mediate and an error-prone DNA resynthesis, leading to introduction of mutations. C) The same mutagenic pathway has been observed in non-B cells.

Recently, a structural function for EXO1 as been identified, where cells without EXO1 nuclease activity are MMR proficient, in part because the mutant EXO1 is still able to interact with other MMR proteins and coordinate repair (44). As in BER, RPA protects the

ssDNA-gapped region generated during excision. DNA re-synthesis is then catalysed by DNA Pol δ/ε and lastly completed by Lig1 action.

1.2.2.2 - Non Canonical Mismatch Repair

Non-canonical MMR (ncMMR) was described in B cells as a variant of mismatch repair and it is activated by a variety of lesions. ncMMR differs from canonical MMR in its mutagenic outcome. During B cell development, mutations are introduced in the *Ig locus* (discussed in chapter 1.4.2.2): error-prone MMR generates antibody diversity in B cells during the Somatic Hypermutation (SHM) process (Figure 3B). Key roles for MSH2, MSH6, as well as EXO1 have been identified, but ablation of MSH3, MLH1, or PMS2 has only minor effects on the process (45-47). Interestingly, during SHM only 20–30 bp are excided (48). Presumably, recruitment of MSH2 and MSH6 has a role in determining the EXO1 (or some as yet unknown nucleases) excision process (49).

In canonical MMR, a non-ubiquitinated PCNA mediates high-fidelity resynthesis recruiting DNA Pol δ. In SHM and in ncMMR, PCNA is recruited by MutSα / MutLα and is mono-ubiquitinated at lysine 164 (K164) (50,51). This modification induces the recruitment of DNA polymerase η (Pol η), a translesion polymerase that mediates an error-prone DNA synthesis, leading to the incorporation of incorrect nucleotides (49,50). In particular, during SHM MMR is responsible for 80–90 % of dA:dT mutations. Mice with mutation in PCNA K164 don't display A:T mutations in B cells, because they fail to recruit and activate the error-prone Pol η (52).

Importantly, Peña-Diaz *et al*. used an *in vitro* system to monitor repair of dG:dU mismatches and identified the ncMMR pathway exists also in HEK293T cells, a human embryonic kidney-derived cell line, and is activated by non-immune stimuli (53) (Figure 3C). This suggested a more physiological role for ncMMR.

1.2.3 – Other repair pathways

Aside from BER and MMR, other DNA repair pathways can recognize specific DNA lesions or process intermediates of BER and MMR. They include Nucleotide Excision Repair (NER), DSB repair and DNA damage tolerance.

NER is the preferred pathway for removal of bulky DNA lesions formed after exposure to radiation or chemicals (Figure 1). Several human genetic disorders, including Xeroderma pigmentosum (XP), Cockayne syndrome (CS), and Trichothiodystrophy (TTD), are associated with specific NER protein deficiency, underlying the crucial role of those molecules in human physiology. Two sub-pathways of NER are characterized: 1) the transcription coupled repair (TCR), selective for lesions present in the transcribed strand genes (54), and 2) the global genome NER (GGR) that recognize lesions distributed over the rest of the genome.

DSBs can be generated both by exogenous sources, such as ionizing radiation, as well as via endogenous sources. Endogenous DBSs can arise either spontaneously, for example when DNA replication forks encounter unrepaired DNA lesions, which results in fork collapse (55), or DSBs can be programmed, for example during chromosome segregation (56) or Ig diversification (57) (discussed in chapter 1.4.2). DSBs are repaired by NHEJ or HR, two pathways which differ in the synthesis fidelity and in the template requirements.

NHEJ mediates the re-ligation of the broken DNA molecule and does not require a homologous template; therefore it is not restricted to a certain phase of the cell cycle. NHEJ sequential steps are: 1) DNA ends recognition and assembly of the NHEJ complex, 2) protection of DNA ends from non-specific processing, 3) DNA end processing, and 4) ligation of the broken ends. Ligation of the DSB ends promoted by NHEJ is then frequently error-prone, with small insertions, deletions and substitutions at the break site (58).

In contrast to NHEJ, HR is largely error free and is more prevalent after DNA replication, when an identical sister chromatid is available as template for homologous repair (59). The HR reaction includes the following sequential steps: 1) resection of DNA ends, 2) homologous DNA pairing, 3) strand invasion, 4) DNA synthesis and 5) processing of HR intermediates. The HR process (reviewed in (60)) begins with the resection of DSBs to

generate 3′ ssDNA overhangs. This ssDNA is coated by RPA, which is subsequently replaced by Rad51 recombinase to mediate homologous DNA pairing (61). Rad51 triggers the formation of a connection between the invading DNA and the homologous template, until DNA molecules become aligned, producing a joint molecule. Other mediator proteins such as BLM, Rad54, Rad51AP1 and BRCA2 are also involved in this recombination process. DNA is synthesized using the invading 3'-end as a primer, forming a Holliday junction. Resolution of the structure can generate crossover or non-crossover products (discussed in chapter 1.4.2.1) (60).

DNA damage tolerance (DDT) mechanisms do not remove lesions, but are temporary solutions to overcame stalled DNA replication. It is carried out by a set of either error-prone or error-free processes. Translesion synthesis (TLS) allows the replication machinery to bypass DNA lesions using a low-fidelity DNA polymerase (62). TLS polymerases such as Pol η, Pol κ, Pol ι and Rev1 belong to the Y-family polymerases. These polymerases can synthesise directly across DNA lesions, but lack the proofreading activity, are non-processive, and contain active sites that can accommodate aberrant bases and base pair mismatches (63). The process of TLS requires the exchange of the replicative polymerase with TLS polymerases. TLS patch is then extended from 5 to 60 nucleotides, depending on the lesion and polymerase involved (64). Ultimately, a second switch brings the replicative polymerase to the DNA template and processive DNA replication is restored (65).

Another DDT mechanism is called template switching, as it is carried out replacing the DNA containing the damage with the intact template strand of the sister chromatid, and as such this process is error-free (66). The key regulatory mechanism controlling DNA damage tolerance is post-translational modification of PCNA. DNA damage induced monoubiquitination of K164 by RAD6-RAD18 induces the switch from the replicative to a TLS polymerase (67,68). Polyubiquitination of K63 in PCNA instead promotes template switching (69).

1.2.4 - Crosstalk between DNA repair pathways

Crosstalk among repair pathways plays an important role in regulating how DNA lesions are repaired. As a consequence of evolution, some of the DNA repair mechanisms have overlapping specificities, giving rise to the need to coordinate their activities, and the interaction between two or more repair pathways can lead to competition or cooperation.

Competition is more likely to appear at the first stages of repair, since different repair proteins can recognize the same lesion. A deaminated 5-methyl cytosine base paired with guanine (a dT:dG mismatch) can be recognized by both a specific glycosylases in BER and by MutS ATPases in MMR. hMYH, an adenine DNA glycosylase responsible for the removal of the lesion 8-oxoG via LP-BER, can interact with hMSH6 in the hMSH2/hMSH6 heterodimer (70). While methyl-CpG binding domain protein 4 (MBD4), a DNA glycosylase that recognizes dT:dGand dG:dU mismatches was reported to interact with the MMR protein MLH1(71). It is then necessary to ensure that the most appropriate response is activated in order to repair the lesion and to reduce replication errors that may occur during the repair processes. Solutions to the competition may be related to a cell cycle regulation of the amount of proteins expressed, PTM and degradation, which may favour one repair pathway over others (discussed in chapter 1.6).

Interactions between different pathways are not necessarily competitive, they can also be cooperative: uracil DNA glycosylase (UNG) and hMutS are reported to cooperatively generate *Ig locus* diversification (45) (discussed in chapter 1.4.2.2). Cooperation may be a consequence of a direct protein-protein interaction or may be mediated by DNA. Binding of one protein to DNA can accentuate the DNA flexibility and recruit other repair proteins. The sliding clamp protein trimer PCNA, besides functioning as an auxiliary factor for the DNA polymerases and thus being present in every repair pathway that require a re-synthesis step, has also been reported to interact with specific DNA repair proteins. In NER, PCNA binds endonucleases XP-G and XP-A to recruit them to the lesion (72). In MMR, PCNA interacts with MSH2, MSH3, and MSH6 and enhances their binding specificity for the mismatch (73,74). While the LP-BER pathway for repairing AP sites

depends on PCNA interaction with RFC, Pol δ, FEN1 and Lig1 (75). In addition, APE and UNG interact with PCNA, supporting the idea that this protein is involved not only in re-synthesis, but also in different steps of BER (76). PCNA can thus coordinate numerous proteins involved in different repair pathways, possibly through specific binding affinities, sub-cellular localization and PTM (e.g. PCNA ubiquitination *status*).

Moreover, once the specific repair pathway is engaged, intermediate products generated during the processing of the lesion can be recognized as substrates for other repair pathways. For instance, SSBs generated during BER can be recognized and further processed by other repair pathways (77).

This complex interplay requires coordination in time and space of different functions to create a dynamic assembly of multiple factors that ensure the maintenance of genome stability and cellular homeostasis.

1.3 - HOMEOSTASIS: THE BALANCE BETWEEN INJURY AND REPAIR

In 1929 Walter Cannon described homeostasis (*homeo* means "the same" and *stasis* means "standing") as “the coordinated physiological reactions which maintain most of the steady states in the body” (78). When environmental changes alter the cell equilibrium, coordination between different networks is required for maintaining internal stability. When an alteration occurs at the DNA level, homeostasis is maintained by a fine cooperation between DNA repair mechanisms, cell cycle progression, DNA replication, transcription, chromatin remodelling and apoptosis.

When it is not possible to restore the equilibrium inside the cell, other responses such as apoptosis can be triggered. Excessive damage beyond a critical threshold activates apoptosis in a MMR-dependent fashion (79). Indeed, inactivating mutations of *MSH2* and *MLH1* are associated with hereditary non-polyposis colorectal cancer (HNPCC, also known as Lynch Syndrome), and leads both to an increase in the rate of spontaneous mutation at microsatellite, and to an inhibition of apoptosis (80). Similarly, the ATM kinase can trigger DNA repair or apoptosis (81). Because of their involvement in DNA repair,

these proteins can sense how efficiently lesions are being repaired, and establish whether repair functions or apoptosis signalling functions are needed to maintain tissue homeostasis.

Beside apoptosis, other mechanisms are involved in homeostasis maintenance: persistent association of MMR proteins with its lesions can activate cell-cycle arrest (82), while a less toxic damage can be tolerated until the next cell cycle round (83).

Cellular response can also differ: thymocytes are more prone to engage apoptosis in response to DNA damage, while primary fibroblasts appear to be more refractory, this is possibly related to different basal patterns of expression of pro- and anti-apoptotic regulators (84) or to differential DNA repair pathways activation.

Chromatin remodelling can also be involved in the maintenance of the homeostasis: histones modifications that regulate protein binding to DNA, incorporation of histones variants, or changes in nucleosome positions can promote DNA lesion processing by creating a more permissive environment for DNA repair protein accessibility (25). On the other hand, histone H1 phosphorylation is linked to apoptosis induction, through chromatin changes that facilitate nucleases digestion (85). Depending on cellular context histones can have different effects on the homeostasis maintenance.

Homeostasis can also be defined as the result of a balance between evolution and survival. Cells survival is linked to a continuous and accurate repair of the genome. On the other hand, inaccurate repair leads to different mutational rates and support evolutional segregation. In *Mycobacterium tuberculosis,* the acquisition of antibiotic multi-resistance is due to the accumulation of mutations in genes targeted by the antibiotics, thus conferring a selective advantage. Without errors in DNA repair mechanisms, evolution cannot occur. However, too many errors can lead to severe problems, including cancer (discussed in chapter 1.4). A fine balance has evolved to allow enough genetic variation for adaptation and evolution to occur.

1.4 - ALTERATIONS OF THE HOMEOSTASIS

1.4.1- Alteration of the balance for bad

Given that every day DNA is under the constant attack of exogenous and endogenous damaging agents, it is reasonable to think that alterations in the DNA repair machinery can have a strong impact on genome stability. Consequences of DNA repair deficiency include severe pathological states, such as cancer and hereditary diseases.

1.4.1.1 - Alteration of the balance leading to cancer

Genome instability and/or DNA mutation are hallmarks of all human cancers (86). When changes in the DNA sequence are left uncorrected, because of defective or error-prone DNA repair mechanisms, cells can accumulate mutations that contribute to oncogenesis.

The relevance of DNA damage and repair to the generation of cancer is particularly evident since exogenous DNA damaging agent (discussed in chapter 1.1.2) can induce cancer. UV radiation and IR, besides modifying DNA, can induce specific cancer. Not all cells are equally sensitive to the damage, as radiation affects them in different ways - leukemia, a type of cancer that arises in the bone marrow, is the most common radiation-induced cancer (87). Tobacco smoke is a complex mixture of different chemicals, with acrolein rapidly penetrating the cell membrane and binding to dG, thereby generating DNA-adducts that can crosslink with proteins or other DNA bases. By this mechanism, acrolein can inhibit DNA repair proteins scanning the lesion. Acrolein was shown to be the causative agent of lung and bladder cancer, in human and rat models (88). Major exogenous DNA damaging agents associated with human cancer are listed in Table 3.

Exogenous DNA damaging agents and risk of cancer		
Exogenous DNA damaging agent	**Cancer with increased risk**	**Ref**
Helicobacter pylori infection	Stomach	(89)
HPV and nitric oxide from tobacco smoke	Cervix/Uterus	(90)
UV	Melanoma	(91)
X ray	Leukemia	(87)

Table 3 – Exogenous DNA damaging agents associated with increased risk of cancer.

Besides exogenous agents, inherited polymorphisms or mutations in DNA repair genes may increase the predisposition to develop cancer (Table 4). Usually one defective copy of the gene is inherited via a parental germ cell, while the other copy is inactivated later in the somatic cell.

Inherited mutations in DNA repair genes and risk of cancer			
Mutated Protein	**Repair pathways affected**	**Cancers with increased risk**	**Ref**
BRCA1, BRCA2	HR	Breast, ovarian	(92)
ATM	HR, NHEJ	Leukemia, lymphoma, breast	(93)
ATR	HR	Oropharyngeal	(94)
MRE11	HR, NHEJ	Breast	(95)
BLM	HR	Leukemia, lymphoma, colon, breast, skin, esophagus, stomach, tonsil, larynx, lung, uterus	(96)
WRN	HR, NHEJ, LP-BER	Soft tissue sarcoma, colorectal, skin, thyroid, pancreatic	(97)
XPC, XPE	NER	Skin	(98)
XPA, XPB, XPD, XPF, XPG	NER	Skin, central nervous system	(98)
MSH2, MSH6, MLH1, PMS1-2	MMR	Colorectal, endometrial, ovarian	(99)

Table 4 - Inherited mutations in DNA repair genes that increase the risk of cancer.
Each repair pathway has its organ tropism. DNA repair pathway deficiency leads to specific phenotypical manifestation even if deletion occurred at the germinal level.

One of the best documented examples of the association between inherited mutation in DNA repair genes and cancer is HNPCC, in which mutations in MMR genes give rise to microsatellite instability (MSI) (100). Germline alteration in any of five human MMR genes (*MSH2*, *MLH1*, *MSH6*, *PMS2* and *PMS1*) can result in HNPCC (100). MMR deficiency leads to an increase in the mutation rate, with frameshift mutations generating multiple repetitions of short sequence unit (one to six nucleotides long), called microsatellites. If mutations occur in exonic coding sequences of tumour suppressor genes, tumour development is favoured.

Mutations in DNA repair genes have tissue specificity: mutations in MMR lead to colorectal cancer, whereas mutations in proteins involved in SSBR impair the non-dividing neurological cells (101). Tissue specificity is likely due to the concurrence of different factors, such as the proliferation rates, the exposure to mutagens, or the expression levels of genes that contains short repetitive sequences.

Proteins in the DNA repair pathways can also contribute to carcinogenesis when their ability to initiate apoptosis in response to DNA damage is impaired (102), with a decrease in apoptosis playing an important role in tumorigenesis. MSH2 deficiency induce accumulation of genomic mutations and impairment of apoptosis signalling, thus resulting in a fast tumorigenic phenotype; while Lin *et al*. studied an *MSH2* mutant mice model, unable to activate MMR, but still able to perform apoptosis and resulting in slow tumorigenic phenotype (103).

1.4.1.2 - Alteration of the balance in inherited diseases

Besides cancer, DNA repair-deficiencies can be associated with several rare hereditary diseases such as neurodegenerative disorders and immune-deficiencies. While, phenotypes arising from DNA repair-deficiencies include sensitivity of the skin to sunlight.

In the central nervous system (CNS), due to high glucose metabolism, neurons have a pronounced oxygen metabolism. NER and BER deficiencies can lead to an increase in oxidative DNA damage leading to neurodegenerative disorders, e.g. Xeroderma pigmentosum (XP) and (Cockayne syndrome) CS (104). Ataxia telangiectasia (AT) can arise when ATM is not activated in response to DSBs, thus resulting in reduction of repair and prevention of apoptosis. This is particularly deleterious in brain, where accumulation of damaged cells leads to neurodegeneration (105), while in the immune system AT leads to immunodeficiency, due to defects in B cell maturation.

1.4.2 - Alteration of the balance for good

Evolution can be based on mutations introduced by errors in DNA repair mechanisms. If the mutational process is controlled and programmed, genetic variation, and thus the alteration of the homeostasis, can have a positive impact on the organism. It is known that specific DNA sites are particularly prone to mutation (hot spots), thus providing a substrate for mutability. Moreover, to confer a selective advantage, a mutation should be positively selected, reducing the probability of a deleterious impact. Meiotic recombination and *Ig locus* diversification are examples of mechanisms that confer a selectively advantageous

variability, with HR playing an important role in increasing genetic variability by allowing gene recombination.

1.4.2.1 - Alteration of the balance in meiosis

During meiosis, a reproductive cell containing a diploid number of chromosomes generates four haploid cells, each containing a single set of chromosomes. Each of the four resulting cells differs from the other three because of genetic recombination. Reciprocal genetic exchange of chromosomal fragments was observed for the first time in *Drosophila* in 1931 (106). During prophase of the first meiotic division, when homologue chromosomes are aligned, programmed DSBs are introduced throughout the genome by the meiotic topoisomerase-like protein SPO11. Analogous to somatic HR, the 5' ends are resected to generate 3' single-stranded tails (61). After the scanning of the homologous duplex, strand invasion leads to the formation of a D-loop structure. D-loops can be resolved through crossing-over (CO) mechanism, with reciprocal exchange of large DNA sequences, or through a non-CO mechanism, which involves a unidirectional transfer of short DNA sequences (or gene conversion). Most recombination events are concentrated on defined hotspot sequences, thus only *loci* flanking an hot spot undergo recombination events (107). Hence, programmed DSBs formation is regulated both temporally (with respect to meiotic progression) and spatially. Genetic diversity is further enhanced during the first meiotic division, when each chromosome homologue randomly segregates.

1.4.2.2 - Alteration of the balance in *Ig locus* diversification

Every day our antibodies successfully recognize millions of different antigens, including pathogens, with each immunoglobulin (Ig), specifically recognizing an epitope on the surface of an antigen. To obtain this diversity during development, the Ig-producing cells or B-lymphocytes undergo a physiological process of targeted and regulated alterations of the *Ig locus* (termed *Ig locus* diversification).

Igs are composed of four polypeptide chains, two heavy chains and two light chains linked via disulphide bridges (Figure 4). There is a single heavy chain *locus* (IgH) and two light

chain *loci* (Igκ and Igλ) in the human genome. The IgH sub-class (μ, δ, γ, α, ε) define the effector function of the antibodies, called IgM, IgD, IgG, IgA and IgE, respectively.

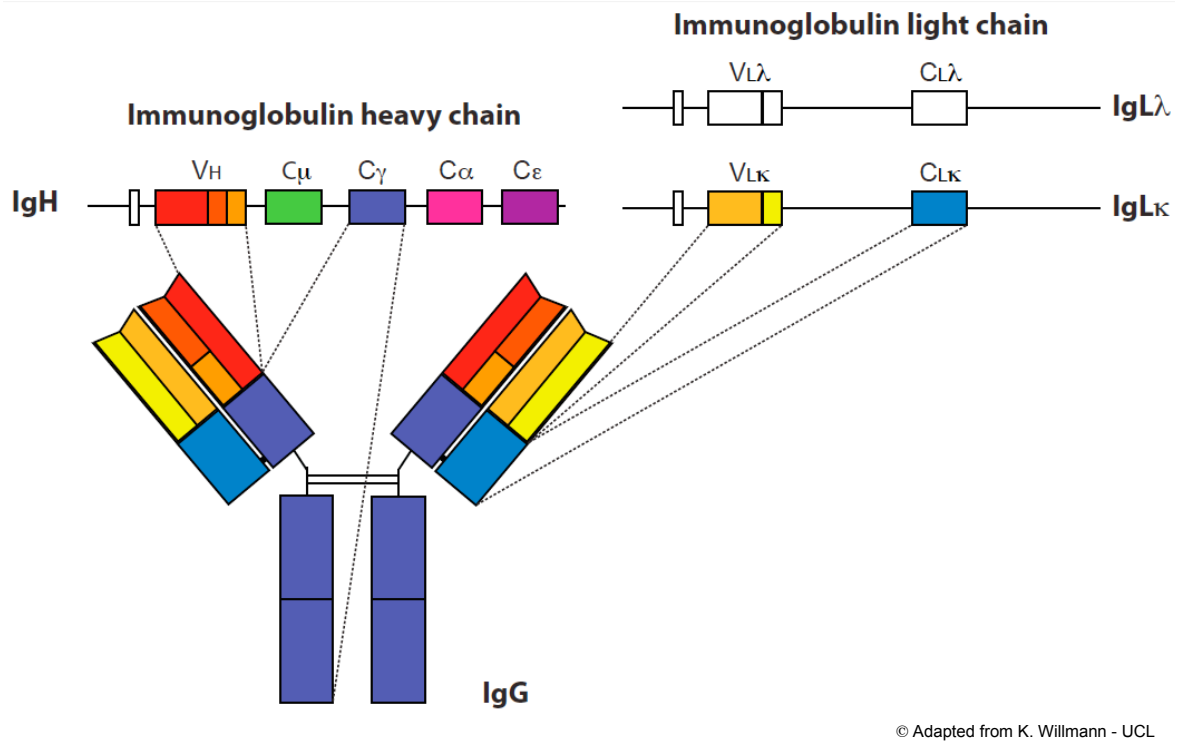


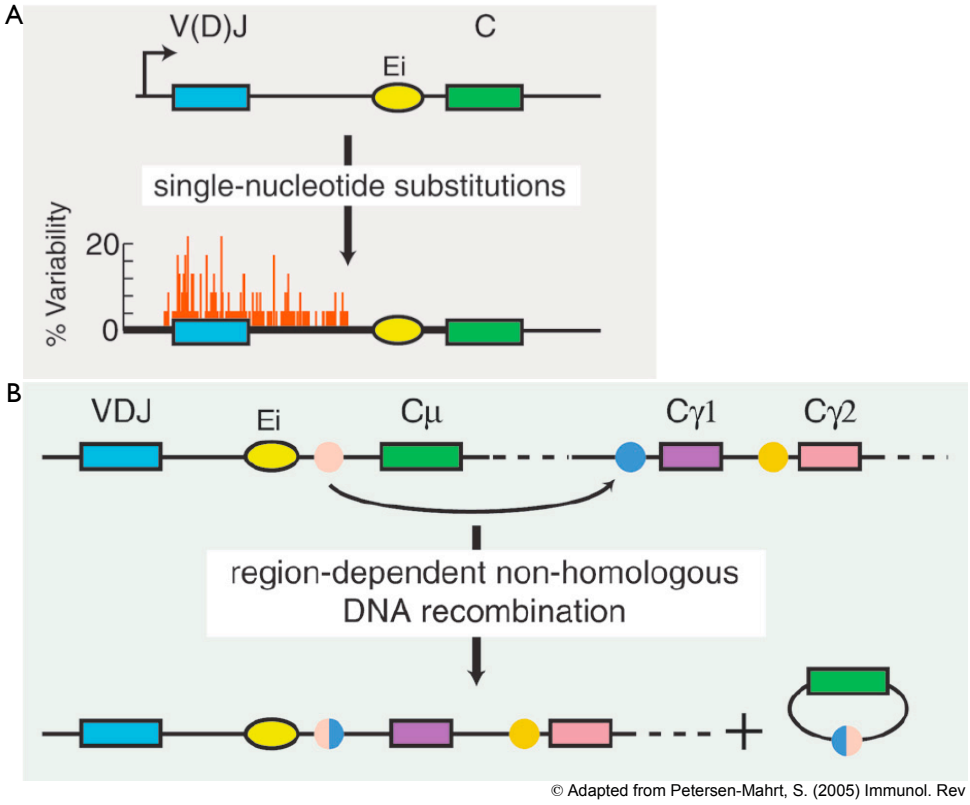
Figure 4 - Ig locus organization and protein structure

Heavy-chain (IgH) variable (V) regions are constituted of three gene segments: V,(D) and J (red to orange VH). The constant (C) region consists of one of several several segments (μ, δ, γ, α, ε). Light-chain (IgL) V-region genes are constituted of two segments V and J (orange and yellow) and form the light chain peptide with a C segment. The antibody is composed of two heavy chains (dark blue) and two light chains (light blue). Black bars mark disulphide bridges

The development of B cells from progenitor cells to functionally mature B cells takes place in the bone marrow. Here, B cells undergo successive programmed rearrangements, first in IgH and then in IgL *locus*. Variable (V), diversity (D) and joining (J) gene segments in the IgH *locus* join randomly to assemble the pre-B cell receptor (108). V(D)J recombination is catalysed by the proteins RAG1 and RAG2, which introduce DSBs at recombination signal sequences (RSSs) that flank each V, D and J gene segment. These DSBs are substrate for TdT polymerase, which adds random nucleotide to the DNA ends, and subsequently the segments are joined through NHEJ, thus the repair is imprecise (109). Overall diversity depends on recombinatorial shuffling of of V, D, and J segments, and on the imprecise joining of those segments. After an analogous VJ rearrangement of the κ or

λ light chain, a complete IgM is expressed on the surface of B cells. However, 55 to 75 % of the B cells emerging from the bone marrow are self-reactive (110) and undergo apoptosis. The surviving B cells leave the bone marrow for the periphery, where after antigen encounter, they undergo further steps of diversification.

In the germinal centres (GC) of lymphoid organs, the DNA in the Ig loci of B cells undergo two important processes known as somatic hypermutation (SHM) and class-switch recombination (CSR) (reviewed in (111) – Figure 5).



© Adapted from Petersen-Mahrt, S. (2005) Immunol. Rev

Figure 5 - Schematic drawing of SHM and CSR

A) SHM introduces point mutations in V genes. Mutations are shown as red vertical lines in the V segment. B) CSR occurs between switch regions located upstream of the different C region (pink and light blue circles). Rearrangement to a different C region results in isotype switching. The intervening DNA sequence is released as a circular episome.

SHM introduces point mutations into the V(D)J exon of IgH and IgL, resulting in differentially mutated variable regions unique for each B cell, thereby creating a novel B cell repertoire (112) (Figure 5A). Only B cells expressing antibodies with higher affinity for specific antigens are clonally expanded. Alternatively, B cells can also undergo CSR,

where the initially expressed IgH constant region (μ) is exchanged with one of the alternative downstream regions (δ, γ, α, ε), thus changing the Ig isotype from IgM to IgD, IgG, IgA or IgE. The choice of the particular isotype being selected is according to the extracellular signals received by cytokines (113) (Figure 5B). CSR and SHM are initiated by activation-induced deaminase (AID), a member of the AID/APOBEC family of cytosine deaminases (114). This enzyme introduces DNA lesions by deaminating dC into dU in DNA (Figure 6 - discussed in chapter 1.5).

It is clear that introduction of controlled and beneficial mutations has a positive impact on the selection and evolution of antibodies, that enhance the recognition of antigens with high efficiency. Moreover, individuals without AID or with a mutated version of the gene develop hyper IgM syndrome and a mild immunodeficiency (115). Thus perturbation of the homeostasis is necessary for survival and evolution.

1.5 - THE NATURAL DNA MUTATOR AID

AID is the protein that best recapitulates the necessity of a balance in DNA stability to exist between survival and evolution. The consequences of AID activity, both in physiological and pathological situations, will be discussed below more in details.

1.5.1 - AID in immunity

AID is a 24 kDa enzyme discovered in 1999, while screening for genes that are upregulated upon CSR in CH12 mouse B lymphoma cells (114). Its DNA deaminase activity was elucidated *in vitro* and *in vivo* years later (116-118). The DNA substrates that are recognized by AID are well defined: including dC, 5FdC and 5mdC (119,120). *In vitro*, AID deaminase activity was shown to be distributive, rather than processive (121), meaning that rather than moving along the DNA towards the next target, it dissociates from its substrate after each deamination and reassociate with the template at a different site. Another group proposed a jumping and sliding model: AID was hypothesized to bind ssDNA and move bidirectionally, moving or jumping to a distal part of the molecule without

disassociating (122). Catalytic activity is harboured within the cytosine deaminase domain, which includes a zinc co-ordinating motif and histidine 56 (H56) and glutamic acid 58 (E58). The C-terminal portion of AID contains a nuclear export signal (NES), and was shown to be essential for CSR (123). The N-terminus contains a nuclear localization signal (NLS) and is essential for SHM (124). AID requires ssDNA as a substrate, which *in vivo,* during *Ig locus* diversification, seems to be provided by the concomitant transcription elongation process (51,125-128). Moreover, in hypermutated Ig genes, dC is more frequently targeted in the context of WRC (W= dA or dT; R = dA or dG), a sequence preference that was also observed *in vitro* experiments (120,129). Once dC, in the preferred context of WRC is accessible, AID deaminates dC to dU, leading to the formation of a dU:dG mismatch in dsDNA (Figure 6).

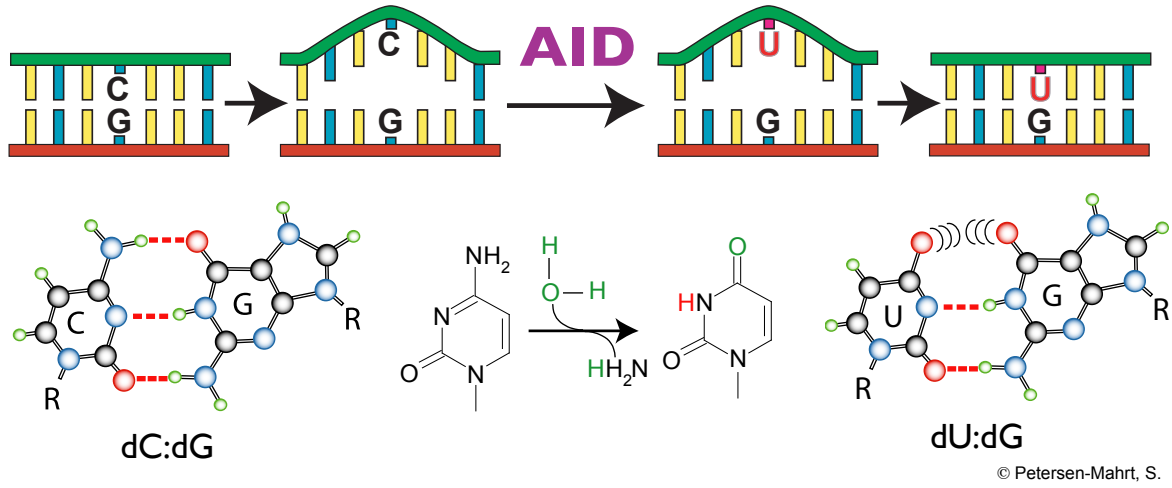


Figure 6 - Reaction catalysed by AID

The catalytic activity of AID in ssDNA produces a uracil from a cytosine by hydrolytic deamination. AID deaminates the HN3 at the C4 position in the pyrimidine ring. This results in a dU:dG mismatch in dsDNA.

If DNA replication occurs before mismatches recognition, a transition from dC to dT will take place. A mechanism through which most of the mutations are generated in phase 1a of SHM (112). Otherwise, in phase 1b of SHM, dU can be recognized and excised by UNG glycosylase and partially processed by the BER pathway (130); if replication occurs after the AP site generation, the polymerase can incorporate any base and transversion (interchange of a purine with a pyrimidine base, e.g. dC to dA) or transition (interchange within purine or within pyrimidine bases, e.g. dC to dT) mutations can occur. Replication

past AP site requires the polymerase to incorporate a base without a template; this could involve either normal replicative polymerases or translesion polymerases (131,132). This mutagenic process was confirmed by the fact that although deamination takes place at dC, mutations at dC/dG and dA/dT basepairs are observed with equal frequency *in vivo* and transitions are slightly more frequent than transversions (133), with *UNG* deficient mice displaying altered SHM (130). SMUG1 glycosylase can also remove dU from both dU:dG mismatch and dU:dA pairs, but has a very low turnover number compared to UNG(134), and has been shown not to partake in SHM (135).

Furthermore, in phase 2 of SHM, dU:dG mismatches or the AP site can be recognized by other repair pathways. When the MMR machinery is recruited (45,47), mutations at A/T near the initiating U/G lesion are generated. Indeed, deficiency in MSH6 (136), DNA Pol η (132), or EXO1 (46) leads to a reduction in the accumulation of mutations at AT pairs in Ig genes. In $MSH2^{-/-}$ residual mutations at AT pairs are observed, probably because LP-BER is involved through PCNA interaction in error prone polymerases recruitment.

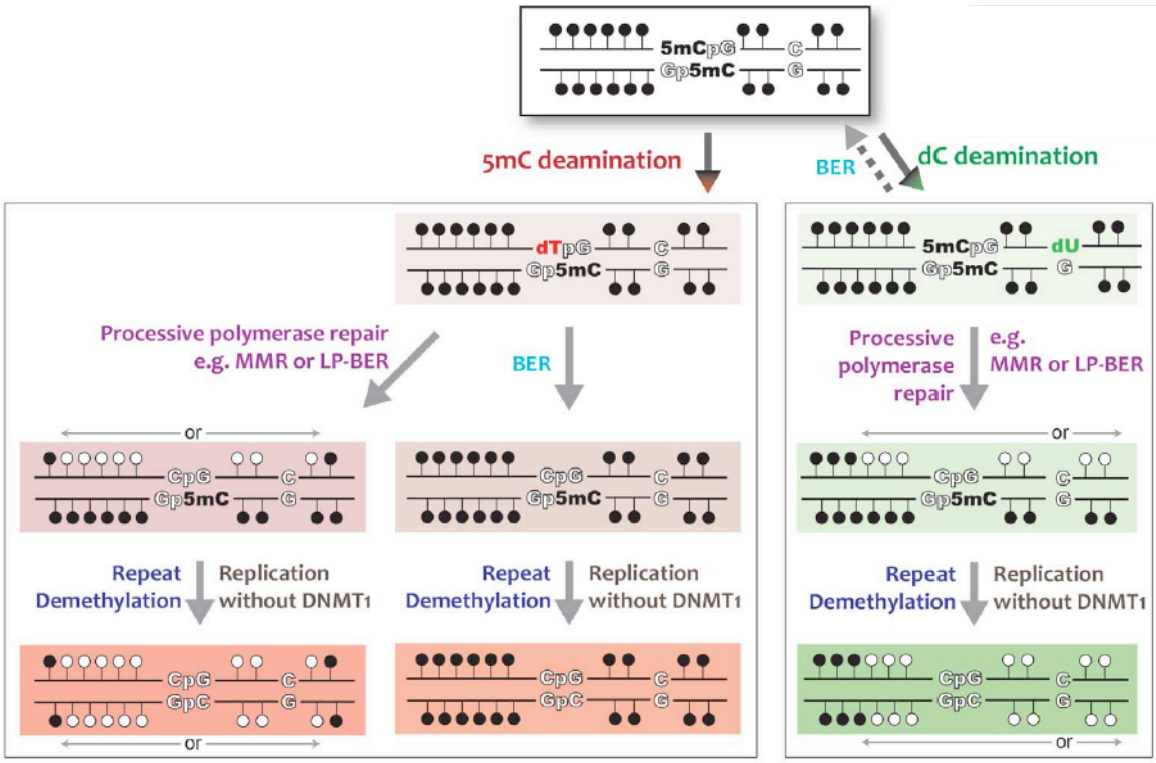
Moreover, while $UNG^{-/-}/SMUG^{-/-}$ mice are viable and fertile, $UNG^{-/-}/SMUG^{-/-}/MSH2^{-/-}$ triple knock-outs have a shortened lifespan, mostly because of cancer development (137). This data demonstrate that MMR may serve as an ultimate backup in repair of dU lesions.

The model of error-prone ncMMR at *Ig locus* (53) was discussed in chapter 1.2.2.2. However, recently it has been demonstrated that UNG and MSH2 contribute both to error-free and error-prone repair (138), possibly via different accessibility of polymerases to the damaged *locus*. Further studies showed that in the presence of the hMutSα complex dU:dG mismatches in the *Ig locus* are repaired with high fidelity, while in the absence of the MSH2/MSH6 heterodimer error-prone repair is favoured (139). This balance between error-free and error-prone repair may be a protective mechanism, to avoid excessive *Ig locus* mutagenesis and subsequent malignant transformation of B cells. This also implies that SHM is an active mutagenic process, rather than a failure in the maintenances of the fidelity of DNA replication.

1.5.2 - AID outside of the immune system

1.5.2.1 - AID in epigenetic reprogramming

The discoveries that induction of AID expression can be stimulated by estrogen (140) and that AID associates with proteins that are not part of B cell development, suggested for AID to have a function outside the immune system. Indeed, AID was found to be expressed not only in immune tissues, but also in oocytes and primordial germ cells (PGCs) (120). In PGCs, AID was linked to epigenetic reprogramming and active DNA demethylation, because of its ability to deaminate 5mdC (120). Deamination of 5mdC by AID generates dT, which in the context of dsDNA results in a dT:dG mismatch. This lesion could be recognized by TDG and MBD4 glycosylases, which can excise the dT and generate an AP site that is repaired with a non-methylated dC. Indeed, *TDG*$^{-/-}$ mice exhibit impairment in global methylation and embryonic lethality (141); while in *AID*$^{-/-}$ mice the content of 5mC in PGCs is increased by 30 % (142). dT:dG lesions and AP sites can also be recognized by proteins involved in MMR, translesion synthesis, or NHEJ. The consequence of recognition of dT:dG by non-BER pathway is the discovery that processive DNA demethylation can be induced by AID (143). Mice expressing a G-AID fusion protein were bred with mice harbouring the GAL4 binding site (UAS) inside the methylated H19 *locus*. Subsequent analysis of the methylation status of the DNA surrounding the UAS region was performed. Significant processive demethylation dependent on G-AID catalytic activity was observed (143). Since AID was shown to be non-processive (121), the processive demethylation can only be the result of the activated repair mechanisms (Figure 7). While the SP-BER pathway will produce a single demethylation event, the processive repair replaces long stretches of all bases, including those surrounding the 5mC, leading to multiple demethylation events. Furthermore, demethylation can also be achieved indirectly, with AID targeting a non-modified dC: activation of the same processive repair mechanisms will lead to extensive removal of surrounding bases, including the methylated ones (143). This model reduces the number of DNA damages initially required and explains the efficiency of DNA demethylation.



© Adapted from Franchini D.M. *et al*. (2014) PLoS One

Figure 7 - Model for lesion-induced DNA demethylation

AID can target both normal cytosine (green panel) and methylated cytosine (red panel). Different lesions are generated, (dU and dT:dG mismatch) and depending on the repair mechanism recruited, resolution can be processive or not. Independently from the targeted base, if processive polymerases dependent repair is recruited, several bases surrounding the lesion (including the methylated cytosines) are removed and replaced with normal bases (white circles), leading to DNA demethylation.

This fully support the notion that downstream lesion resolution and activation of DNA repair pathways determine the outcome of AID activity. As DNA methylation has important regulatory functions in development and in the control of gene expression, DNA demethylation is a necessary counterbalance for epigenetic marks on DNA.

1.5.2.2 - AID in cancer

AID is a natural DNA mutator, designed to introduce genomic instability into a defined *locus*, thus misregulation of its activity can easily lead to severe pathological outcomes, including cancer. Alteration of AID function can occur both in B cells, while AID is carrying out its natural function leading to B cell malignancies, and when AID is activated outside of the immune system. As a result of aberrant CSR, many B cell malignancies bear

chromosomal translocations that join the highly transcribed *Ig locus* with portions of proto-oncogenes (144). Burkitt's lymphoma is characterized by translocation between the *c-myc* oncogene and the IgH *locus*. AID expression was found to increase *c-myc/IgH* translocations, and AID was shown to be essential for the DSBs generation in *c-myc* that result in those translocations (145). In addition to *c-myc*, a subset of other Ig and non-Ig genes undergo aberrant hypermutation or translocation in B cells, including *BCL6*, *FAS* and *PAX5* (146). Moreover, AID is also required for tumour progression in different leukaemia's, such as chronic myeloid leukaemia (147) and chronic lymphocytic leukemia (148). Here, the acquired mutations resemble those introduced by SHM, as they are predominantly in the WRC hotspots and are transition mutations.

Constitutive expression of AID in mice leads to various tumour formation, such as lung T-cell (149), liver (150) and gastric cancer (151). Furthermore, *Helicobacter pylori* infection has been shown to trigger NFκB-dependent upregulation of AID, which lead to gastric cancer (151). Diversification of tumour formation let us hypothesize that there is a specific cell susceptibility to AID-induced lesions. Moreover, bioinformatics analysis of the dC to dT transitions occurring in *p53* and *APC* genes (in different human cancers) reveal similarities between the mutation spectrum of those genes and AID/APOBEC preferential targets (152). Moreover, recently, the TC sequence context of the mutations generated by APOBEC3A and APOBEC3B was related to signatures of localized hypermutation, in human breast cancers (153). This possible tumorigenic mechanism is defined by the outcome of AID lesion resolution, with inefficient repair and fixation of a mutation.

Alteration of the balance between survival and evolution, with AID overexpression or AID depletion can severely affect our health, ranging from autoimmunity (review in (154)) to tumour formation and, in the worst scenario, lethality. Natural AID lesions can lead to the activation of repair pathways, recombination, or mutation, each in its physiological context.

1.6 - Current State

As previously discussed, downstream lesion resolution and DNA repair specific activation determine the outcome of AID activity. Regulation of AID damage repair can occur at different levels: depending on where the damage is located, the cellular microenvironment, or chromatin interference a different outcome can be expected.

- dU:dG mismatches in non-Ig *loci* usually lead to proper DNA repair and reversion to dC, with the spontaneous deaminations that occur every day in every cell being efficiently repaired.
- Aside from the location of the AID-induced damage (e.g. *Ig locus*), the cellular milieu (e.g. B cells) plays an important role in lesion processing. This was shown when AID-induced dU:dG mismatches in the post-replicative meiotic nuclei cause an induction of meiotic recombination (155), while the same damage in replicative germ cells is repaired. Indicating that cells have variable sensitivity for DNA damage depending on the presence or absence of a single repair pathway, leading to efficient or inefficient lesion processing.
- It is known that the chromatin structure influences both the recruitment of DNA repair mechanisms and the accessibility to the lesion. It has been reported that hMutSα can lead to fluctuation in the nucleosome positioning, with the rate of disassembly being enhanced by histone H3 acetylation (156). It is possible that chromatin remodelling interact with MMR protein, coordinating repair with other metabolic processes, such as DNA replication and transcription.
- Moreover, cell cycle checkpoints are activated upon DDR, and DNA repair protein expression in different phases of the cell cycle may have a role in directing AID-damage resolution. AID localizes at the IgH *locus* during the G1 and early S phase in splenic activated B cells (157), hence AID could coordinate UNG and MMR proteins that are present in the G1 phase of the cell cycle (158). However, it is not clear, if error-prone repair occur preferentially during replication, a time when DNA is more accessible. Post-translational modification, such as phosphorylation and ubiquitination may modulate the DNA damage response in a spatial and temporal manner.

AIM

DNA damage is a prerequisite for physiological processes during meiotic recombination, *Ig locus* diversification, and epigenetic reprogramming. On the other hand, DNA damage is a major threat for a cell, as it can lead to genomic instability, including DSB-induced translocation or mutation. In each case the DNA damage response plays a crucial role in the outcome, but the mechanisms that govern error-free *versus* mutagenic processing of uracil remain obscure. Several molecular pathways can contribute to uracil lesion resolution, such as cell environment, location of the damage, sequence specificity, chromatin conformation, specific signal transduction, protein interacting with repair pathways, rate of mutations, cell replication state and many others.

The aim of my PhD thesis was to investigate the cited molecular pathways that could explain the cellular differences that, upon AID lesion resolution, lead to repair, recombination, or mutation.

To this end, I decided to use AID as a tool and study AID for its physiological functions. As a tool, DNA deaminases provide two important aspects not previously utilised in the DNA repair field: 1) DNA deaminases are the only fully physiological lesion generating proteins known, 2) they produce only one type of lesion in non-modified DNA deaminated cytosine bases. These specific points allow us to control DNA lesion generation and resolution, and provide novel insight into the Ig dependent AID-lesion resolution. The aim was to directly target a plasmid DNA with AID to introduce dU:dG mismatches and determine how different cell extracts in different conditions respond to the damage leading to DNA repair. The working hypothesis is that the distribution of mutation *versus* repair is determined by a balance between recruitment of high fidelity *versus* error prone repair systems. This balance can be moved toward repair or introduction of mutation depending on cell context, interacting factors, access of polymerases to the target locus, DNA methylation, nucleosomes formation, and other factors yet not identified.

CHAPTER 2 – MATERIALS AND METHODS

2.1 - MATERIALS

All media or buffers marked with "*" were prepared by IFOM kitchen

2.1.1 - Buffers and reagents

2 M NaCl buffer: 2 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM DTT

1 M NaCl buffer: 1 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM DTT

0.65 M NaCl buffer: 0.65 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM DTT

0.2 M NaCl buffer: 0.2 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM DTT

CALBIOCHEM® EDTA-free protease Inhibitors*

Coomassie Brilliant Blue solution*: 50 % methanol, 10 % glacial acetic acid, 40 % water, 0.25 % coomassie blue R-250

De-staining solution*: 25 % methanol, 7.5 % glacial acetic acid, 87.5 % water

Dejellying Buffer: 10 mM Tris-HCl pH 8.5, 110 mM NaCl, 5 mM DTT

Elution Buffer: 20 mM MES pH 6.0, 700 mM NaCl, 150 mM KCl, 2.5 mM TECP, 1.6 mM CHAPS, 300 mM L-arginine HCl, 5 % glycerol, 200 mM imidazole; filtered with a 0.2 µm filtering device and adjusted at pH 6.0

Extraction Buffer: 20 mM MES pH 6.0, 700 mM NaCl, 150 mM KCl, 2.5 mM TECP, 1.6 mM CHAPS, 300 mM L-arginine HCl, 5 % glycerol, 10 mM imidazole. Filtered with a 0.2 µm filtering device and adjusted at pH 6.0. Freshly added 50 mg/L RNAse A and CALBIOCHEM® EDTA-free protease inhibitors (1:500)

Fushin Formamide: 0.1 % fushin dye (Sigma) 10 mM Tris-HCl pH 8 in formamide

Hypotonic Buffer: 10 mM Tris-HCl pH 7.5, 3 mM $CaCl_2$, 0.32 M sucrose, EDTA-free protease inhibitors (1:500)

IVR buffer 5X: 50 mM NaCl, 200 mM Tris-HCl pH 8.0, 400 mM KCl. Freshly added 0.005 mM DTT and 5 µg/µl RNAse A

Laemmli buffer: 10 % SDS, 50 % glycerol, 0.1 % bromophenol blue, 50 mM DTT, 150 mM Tris-HCl pH 6.8.

Lysis Buffer: 0.1 M Tris-HCl pH 8.0, 5 mM EDTA, 200 mM NaCl, 0.2 % SDS. Freshly added 0.1 mg/ml of proteinase K

MMR Buffer: 10 mM Hepes-KOH pH 7.5, 200 mM NaCl, 1 mM KCl, 0.5 mM $MgSO_4$, 1 mM $CaCl_2$, 0.05 mM EDTA.

PBS 1x: 10 mM phosphate, 137 mM NaCl, 2.7 mM KCl pH 7.

R Buffer 5x: 250 mM NaCl, 15 mM $MgCl_2$, 200 mM Tris-HCl pH 8, 200 mM KCl, 5 mM DTT, 10% glycerol

Running Buffer*: 25 mM Tris-HCl pH 8.6, 190 mM glycine, 0.1 % SDS

S-Buffer: 50 mM Hepes-KOH pH 7.5, 50 mM KCl, 2.5 mM $MgCl_2$, 250 mM sucrose. Freshly added 2 mM β-mercaptoethanol and 15 µl/mL leupeptin.

TBE* 1x: 89 mM Tris-HCl pH 8, 89 mM boric acid, 2 mM EDTA

TBS 1x: 50 mM Tris-HCl pH 7.5, 150 mM NaCl

TBST 1x: 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 % Tween 20

TE*: 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA

TE-0: 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 0.01 % Tween-20

TE-50: 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 0.01 % Tween-20, 50 mM NaCl

TE-100: 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 0.01 % Tween-20, 100 mM NaCl

TE-500: 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 0.01 % Tween-20, 500 mM NaCl

TE-1000: 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 0.01 % Tween-20, 1 M NaCl

TEN Buffer: 10mM EDTA, 50 mM Tris-HCl pH 7.5, 1M NaCl

Transfer Buffer*: 25 mM Tris-HCl, 192 mM glycine, 20 % methanol, pH 8.3

Washing Buffer: 20 mM MES pH 6.0, 700 mM NaCl, 150 mM KCl, 2.5 mM TECP, 1.6 mM CHAPS, 300 mM L-arginine HCl, 5 % glycerol, 10 mM imidazole; filtered with a 0.2 µm filtering device and adjusted at pH 6.0

2.1.2 - Plasmids

Plasmids used in this thesis are schematically depicted in Figure 8. pSPM589 (159) was constructed by S.Petersen-Mahrt (the DNA-binding domain of GAL4 was inserted into the *NcoI* restriction site of AID-His in pSPM75 - a pET30 derived vector), pDMF1620 [REF of IVR paper] was constructed by D.M. Franchini (the catalytic inactive GAL4-AID C87R was created by site-directed mutagenesis from pSPM589), and pGL4.31 was purchased from Promega (UK). pEI7150 and pEI7151 were cloned substituting 2 kB of pGL4.31 with 1.1 kB of PCR amplified low/high nucleosome forming regions (LN/HN) from chicken kappa and lambda Ig *locus*. For the rifampicin assay pETduet (Addgene) and pDS4501 (expressing hAID WT) were used.

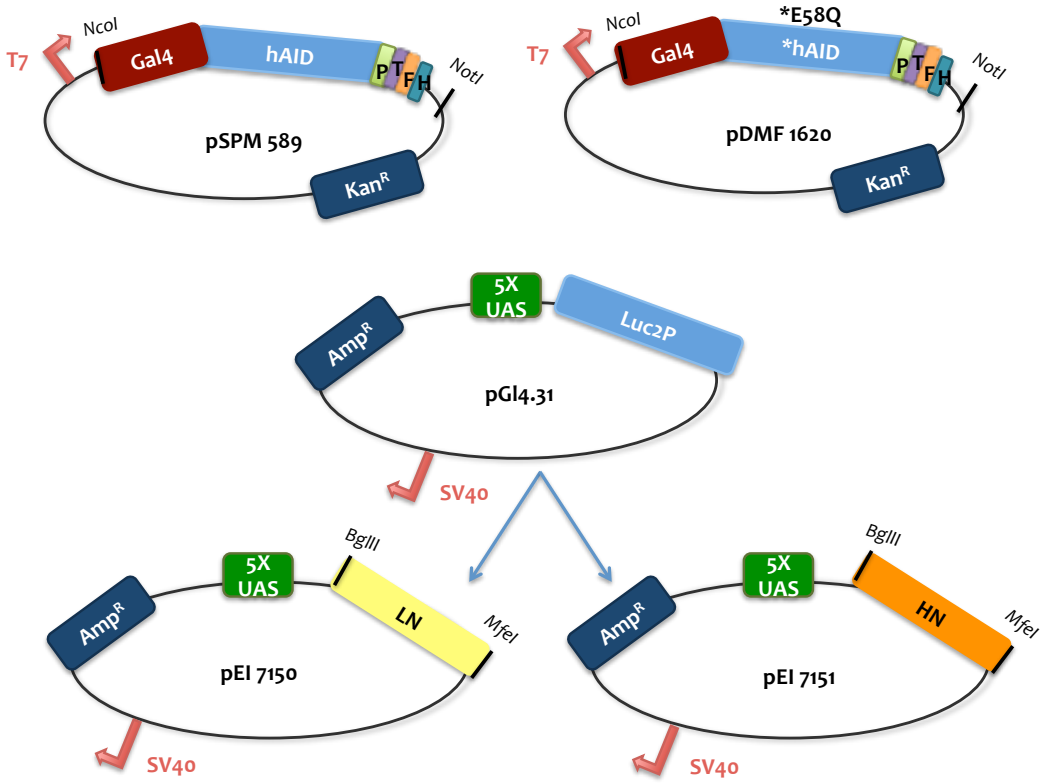


Figure 8 – Schematic drawing of plasmids

pSPM589 was cloned adding the cDNA sequence for human AID (hAID) to the DNA binding domain of GAL4 (GAL4); additional tags were added for specific isolation: P = protein C tag, T = TEV sequence to cut off tags from the protein, F = Flag tag, H = His-tag. KanR = Kanamycin resistance cassette. pDMF1620 is identical to pSPM589 with exception of a missense mutation E58Q. pEI7150 and pEI7151 were cloned substituting 2 kB of pGL4.31 with 1.1 kB of PCR amplified low/high nucleosome forming regions (LN/HN) from chicken kappa and lambda locus.

2.1.3 - Oligonucleotides

Oligonucleotides used in this thesis are listed in Table 4

NAME	OLIGONUCLEOTIDE SEQUENCE	LENGTH	ORIENTATION	APPLICATION
1512	GGCCTAACTGCCCGGTAC	18	F	qPCR (IVR with pGL4.31)
1518	GTCCACCTCGATATGTGC	18	R	qPCR (IVR with pGL4.31)
7290	CCCCCACCCCCTTTTATAG	19	R	qPCR (IVR with pEI7151 and pEI7150)
163	bioATTATTATTATTAGCTATTTATTTATTTATTTATTTATTTfluor	40	F	ssDNA oligonucleotide deamination ASSAY
164	bioATTATTATTATTAGUTATTTATTTATTTATTTATTTATTTfluor	40	F	dsDNA oligonucleotide glycosylases ASSAY
166	AAATAAATAAATAAATAAATAATAGCTAATAATAATAAT	39	R	dsDNA oligonucleotide glycosylases ASSAY
7175	AAACAATTGCAGTCCTCCCCAGCAGTAAG	29	F	Construction pEI7151
7176	ATACGGAGCTTAAGTTATTAGACGTCCAATGCAAAAACCTCCCACT	46	R	Construction pEI7151
7195	TTGCATTGGACGTCTAATAACTTAAGGCTCCGTATCACTGTGTGGTATAT	50	F	Construction pEI7151
7196	GAAAGATCTATTTCACCCCCAAGGACACT	29	R	Construction pEI7151
7179	AAACAATTGGCTGAGAGGAATATTTTGTCA	30	F	Construction pEI7150
7180	CTCAACCCCTTAAGTTATTAGACGTCTCGGGCTCTTTCTACCTCAA	46	R	Construction pEI7150
7181	GAGCCCGAGACGTCTAATAACTTAAGAAATTGAGTGCTTTTTCTTTGAA	49	F	Construction pEI7150
7182	GAAAGATCTAAACACATCCACCCATCCAT	29	R	Construction pEI7150
7100	GCTTTAATGCGGTAGTTTATCA	22	F	cDNA amplification for transcriptional activity assay
7101	CGTAGTCGATAGTGGCTCCAA	21	R	cDNA amplification for transcriptional activity assay

Table 4 – oligonucleotides length, orientation and applications.

2.1.4 - Antibodies

Antibodies used for Western Blot are indicated in table 5.

ANTIBODY	MANUFACTURER	CONCENTRATION / DILUTION
Mouse monoclonal anti-HIS HRP conjugated	sc-8036, Santa Cruz	1:1000
Mouse monoclonal anti-AID STR9	Bernardo Reina-S-Martin	1:10000
Rabbit polyclonal anti-MBD4	ab12187, Abcam	0.5-1 ug/ml
Rabbit polyclonal anti-UNG	ab123668, Abcam	2 ug/ml
Rabbit Polyclonal anti-EXO1	AP20554PU-N, Acris Antibodies GmbH	1-5 µg/mL
Rabbit polyclonal anti-H3	ab1791, Abcam	1:10000
Mouse monoclonal anti-alpha Tubulin	Sigma*	1:5000
anti-PhosphoSerine	ab17464,abcam	1:1000

Table 5 – Antibodies, manufacturers and concentration used.

*was supplied by IFOM kitchen.

2.1.5 - Chemical Inhibitors

Chemical inhibitors used for cell cycle synchronization (Nocodazole) and inhibition of specific DNA repair pathways are listed in table 6

INHIBITORS	MANUFACTURER
Nocodazole	M1404-2MG, Sigma
Uracil Glycosylase Inhibitor (UGI)	M0281L, Euroclone
T2 amino alcohol PCNA Inhibitor (T2AA)	Naoaki Fujii
Aphidicolin	A0781-1MG, Sigma

Table 6 – Chemical inhibitors and manufacturers.

2.2 - METHODS

2.2.1 - Cell culture

Cell thawing

A cryotube with cells was thawed quickly in a waterbath at 37 °C and immediately transferred to a 15 mL Falcon Tube with 10 mL warm culture medium. Cells were collected by centrifugation at 1200 rpm for 5' and the supernatant was aspired. The cell pellet was resuspended, accordingly to cells density, in fresh media and transferred in culture device.

Cell freezing

Cells were centrifuged at 1200 rpm for 5' and the supernatant aspired. Cells were gently resuspended in 1 mL of freezing medium (10 % DMSO and 20 % FCS in culture medium) and transferred in a cryotube, which was placed in a NALGINI Freezing Container containing isopropanol for 72 h at -80 °C. Vials were then transferred into liquid nitrogen tank.

Culturing cells in suspension

Cells were monitored on a daily base analysing morphology, cell density and changing in the colour of the culture medium. Live and death cells were counted using Tryphan Blue solution and hemocytometer. Cells were cultured at 37 °C with 5 % CO_2 and medium was always warmed at 37 °C prior to adding it to cells.

Ramos (human Burkitt's lymphoma) cell line was maintained at 2 x 10^5 in RPMI 1640 supplemented with 10 % FCS from South America (SA), 2 mM L-Glutamine, 10 mM HEPES pH 7.5, 1 mM Sodium pyruvate and 5 % Penicillin/Streptomycin.

BL2 (human Burkitt's lymphoma clone 2) cell line was maintained at 2 x 10^5 in RPMI 1640 supplemented with 10 % FCS SA, 2 mM L-Glutamine and 5 % Penicillin/Streptomycin.

HelaS3 (human cervix epithelial carcinoma) and Colo704 (human ovarian adenocarcinoma) cell lines were maintained at 4 x 10^5 in DMEM supplemented with 10 % FCS SA, 2 mM L-Glutamine and 5 % Penicillin/Streptomycin.

Chicken bursal derived DT40 cell line was maintained at 5 x 10^5 in DMEM/F-12 Glutamax™ supplemented with 10 % FCS SA, 5 % Chicken Serum (heat inactivated 30' at 55 °C), 5 % Penicillin/Streptomycin and 50 µM β-Mercaptoethanol.

Cell synchronisation

Ramos and HelaS3 cell lines were grown at 2 x 10^5 at different concentrations of Nocodazole (0-100 ng /mL) and cell number was counted at different time points (0-24h). Nocodazole 1mg/mL was diluted in 20 % DMSO and 80 % growing medium. After testing, Final synchronization occurred with 40 ng/mL Nocodazole 14h or 17 h (for Ramos and HelaS3 respectively).

Cell extract Preparation

Cells were grown to reach density of 1-5 x 10^8 and were collected by centrifugation in 500 mL bottles at 1200 rpm for 10'. Cell pellet was washed with 50 mL ice-cold 1x PBS and after centrifugation the packed cell volume (PCV) was measured. After measurement, Hypotonic Lysis Buffer (5 times PCV) was added to the pellet, which was properly resuspended. Pellet were incubated 12' on ice and at 4 °C to allow cells to swell. After 12', 10 % Triton X-100 in Hypotonic Lysis Buffer was added to the pellet dropwise to a final concentration of 0.3 %. The suspension was mixed and incubated for 3' on ice at 4 °C. After 3' the suspension was centrifuged at 2500 rpm and supernatant, named cellular extract, was collected and was stored in liquid nitrogen.

2.2.2 - Cell staining

Propidium Iodide (PI) and Caspase 3 staining: 10^6 cells were harvested and centrifuged at 3000 rpm for 5'. Cells were washed once in 1mL of cold PBS before incubation for 20' on ice with formaldehyde 1 % in 1x PBS. Cells were then washed in PBS 1 % BSA and resuspended in 250 µL of PBS and 750 µl of cold ethanol (added dropwise and slowly while vortexing, to avoid aggregate formation). Cells were left in fixative 30' on ice, washed in 1mL PBS 1 % BSA and permeabilized with 100 µL of TritonX 0.1 % in PBS 10' at RT. After another wash in PBS 1% BSA the cells were resuspended in 500 µL of Goat Serum 10 % in PBS. After 30' cells were centrifuged 3000 rpm 5' and the pellet was

resuspended in 100 µL antibody anti caspase3 diluted 1:50 in PBS 1 % BSA and incubated for 1h at RT. Cells were then washed in PBS 1 % BSA and incubated with 100 µl anti-rabbit FITC diluted 1:50 in PBS 1 % BSA 1h at RT protected from light. Cells were finally washed in PBS 1 % BSA and resuspended in 500 µL PI (2.5 µg/mL) (diluted in PBS with 250 µg/mL RNAse) and kept at 4 °C until FACS analysis. FACS was performed by IFOM imaging unit.

2.2.3 - In vitro transcription assay

The in vitro transcription assay was performed using HeLaScribe® Nuclear Extract in vitro Transcription System (Promega), but instead of using a radioactive detection system we performed RT-qPCR. Hela Nuclear Extract Positive Control DNA (containing a CMV promoter) was added to a reaction mix containing: 1x transcription buffer, 2 mM $MgCl_2$, 0.4 mM rNTPs and 35 µg of cellular extracts. The mix was incubated at 30 °C for 60' and the reaction was stopped by adding 175 µL of Stop Solution. RNA was extracted with 200 µL of phenol:chloroform:isoamyl alcohol (25:24:1) and RNA was precipitated with 700 µL of 100 % ethanol. After centrifugation at 14000g for 10' pellet was resuspended in 10 µL of Nuclease-Free water. DNA was removed using RQ1 RNase-Free DNase treatment (Promega) following the manufacturer's instruction. Reverse transcription of extracted RNA was performed with SuperScript® VILO cDNA Synthesis Kit and Master Mix (Invitrogen), following manufacturer's instructions. Controls with no Superscript enzyme were performed in parallel and cDNA was analysed with RT-PCR, using primers indicated in table 4.

2.2.4 - Interphase frog egg extract preparation

From Xenopus eggs is possible to isolate a clear cytosol, which contains both soluble proteins of the disassembled nuclei as well as soluble cytoplasmic proteins.

Xenopus eggs are isolated from hormonally primed female frogs. Eggs are arrested in interphase with cycloheximide, which, acting at the ribosome level, prevents new protein synthesis. In particular Cyclin B in not synthetized and mitosis is inhibited.

Collected eggs are treated with Dejellying Buffer to remove the gelatinous coat around the eggs and washed three times with MMR wash buffer, by swirling gently and pouring out the wash buffer. Eggs are then activated for 5' with calcium ionophore at 1 µg/ml, which leads the exit from M phase and re-entry into interphase, to have a synchronous activated eggs extract. Eggs are washed three times with MMR wash buffer and 3 times with ice-cold S-buffer (supplemented with Leupeptin 15 µl/mL) for lysis and to stop activation. Packed eggs are then centrifuged at 13200 rpm for 12' at 4 °C. Eggs will separate into three different layers: a top bright yellow layer that contains lipids, a middle golden layer containing cytoplasm and membranes and a bottom dark layer containing pigment and yolk granules. Cytoplasmic fraction is collected with a small Pasteur pipette and cytochalasin B (4 µg/mL) is added in order to prevent actin polymerization and to dissociate the membranes. The fraction is then transferred in ultracentrifuge tubes (Beckman 349622) and centrifuged at 70000 rpm 12' at 4 °C (TLA100.3 rotor). Membranes and cytosol are separated from the bottom mitochondria layer, which is discarded. Cyclohexemide (40µg/mL) is added to inhibit protein synthesis and ATP (30 mM creatin phosphate and 150 µg of creatin phosphor kinase are added to have a nuclear competent extract). With 5 % glycerol addition FE is frozen drop wise in LN.

Genomic DNA extraction

DT40 CL18- cell pellet was washed in PBS 1x and resuspended in Lysis Buffer (with SDS and Proteinase K). After 2h at 56 °C phenol/chloroform extraction was performed. Residual phenol was removed with 2 volumes of diethyl ether and DNA was precipitated with 2 volumes of isopropanol and 750 µl of cold 70 % ethanol. DNA was resuspended in 10 mM Tris-HCl pH 8.

2.2.5 - Cloning methods

General PCR

PCR from chicken genomic DNA was performed in a total volume of 25 µl. 50 ng of DNA were amplified in a reaction buffer composed of 1x Pfu Buffer, 0.2 mM dNTPs, 0.4 pMol / µL of primers and 1.25 Units of Pfu polymerase. The amplification program was as follow: 94 °C for 30", 10 cycles of 94 °C for 30", decreasing touchdown from 65 °C by 1 °C every cycle and 72 °C for 2'. Then 25 cycles at 94 °C for 30", gradient from 58 °C to 68 °C for 1' and 72 °C for 2' was applied. PCR was carried out using an Eppendorf Mastercycler gradient PCR machine.

PCR Gel purification

PCR products were loaded onto agarose gels and after electrophoresis bands were excised and purified with QIAquick gel extraction kit (Qiagen), accordingly to the manufacturer's instructions. Elution from columns was performed in 50 µL of TE buffer.

Topo TA Cloning

PCR fragments were cloned into TOPO vector using the TOPO TA Kit dual cloning promoter (Invitrogen), according to manufacturer's protocol. Transformed *E. coli* cells were plated on LB agar plates with ampicillin (50 µL of 50 mg/mL) and X-gal (80 µL of 20 mg/mL) and grown at 37 °C overnight.

Colony PCR

A single bacterial colony was resuspended in 100 µL of LB medium. 1 µl of a 1:100 dilution of the bacterial suspension was added to a 24 µL PCR reaction that included the primers, 2.5 µl of 10x Taq buffer, dNTPs at 0.4 pMol/µL and 1.25 Units of Go Taq polymerase (Promega). The reaction was initially denatured at 95 ºC for 30", followed by 10 cycles of denaturation at 94 ºC for 30", annealing at 55 ºC for 30", extension at 72 ºC for 30" before a final extension at 72 ºC for 52' to generate the PCR product. The acquired PCR product was run on a 1 % agarose gel for verification purposes.

Fragment Preparation

Samples were prepared in a total reaction volume of 30 µL by mixing 1 µg of DNA with NEB buffer compatible with the enzyme used and 10 Units of restriction enzyme. Samples were then incubated at the recommended temperature for 2h. After incubation the enzymes were heat inactivated and the DNA was treated with 5 Units of Alkaline Phosphatase (CIP, Neb) at 37 °C for 1h. Digested fragments were purified on agarose gels.

DNA Ligation

Ligation reaction was performed in 20 µL by mixing 2 µL of 10x T4 DNA ligase Buffer, approximately 30 ng of digested vector backbone and DNA insert in molar excess of 3-10 times compared to backbone. 400 Units of T4 DNA ligase (NEB) were added to the reaction and samples were incubated for 1h at RT.

Competent cells preparation

C41 cells, *E.Coli* cells derived from BL21(DE3) strain, were freshly plated onto a LB plate and grown ON at 37 °C. A single colony was inoculated in a 3 mL starter culture of 2xTY. After 1h at 37 °C the culture was diluted to 100mL with fresh 2xTY and grown until OD600 reached 0.35. Cultures were then cooled down in ice cold water for 20'. Cells were harvested by centrifugation at 4000 rpm for 15' at 4 °C. Pellet was washed with 10 mL of icecold 100 mM $MgCl_2$ and cells were collected by centrifugation at 3000 rpm for 15'. Cells were resuspended in 20 mL of 100 mM $CaCl_2$ and left on ice for at least 20'. After final centrifugation at 2100 rpm for 15', cells were resuspended in 200 µL of ice cold 100 mM $CaCl_2$ and were ready for transformation.

Bacterial transformation

E. coli cells were transformed with pGL4.31 as follow: 10 ng of plasmid were gently mixed with 50 µL of Subcloning Efficiency™ DH5α™ Competent Cells (Invitrogen). Bacteria were incubated on ice for 20' before heat shock at 42 °C for 20" and subsequent cold shock for 2' on ice. Bacteria were incubated at 37 °C 1h with 900 µL of S.O.C. warm medium. 200 µL of the culture suspension was plated in amp/LB plates and incubated at 37 °C O.N. When C41 fresh competent cells were used for transformation (with pSPM589

and pDMF1620 plasmids), bacteria were mixed with plasmid DNA and incubated on ice 20' before heat shock at 42 °C for 2'. After growth at 37 °C 1h bacteria were plated on kan/LB plates. One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen) were transformed with pEI7150 and pEI7151 plasmids following manufacturer's instructions.

Mini and Maxi-preparation of plasmid DNA

3 ml of LB medium with antibiotics in a 5 ml polystyrene round-bottom tube (BD FalconTM) was inoculated with a single transformed *E. coli* colony. The culture was incubated overnight, shaking at 37 ºC. For Mini-preparation 2 ml of bacteria culture was harvested by centrifugation at 10,000 rpm for 3'. Isolation of plasmid DNA was then carried out using Wizard Plus SV Miniprep DNA purification system (Promega) according to the manufacturer's instructions. DNA Mini preparation of pGL4.31 was performed with the same kit at 4 °C to maximize supercoiled plasmid elution (160).

For Maxi-preparation 2 mL of bacteria culture was diluted in 200 mL LB medium and grown for 12-16 h at 37 °C shaking. Bacterial cells were harvested by centrifugation at 6,000 x g in a Beckam JA-14 rotor for 15' at 4 ºC and DNA was purified using QIAGEN Plasmid Maxi Kit (Qiagen) according to the manufacturer's instructions.

DNA Sequencing

DNA sequencing reaction was performed by Cogentech sequencing unit.

2.2.6 - G-AID expression and purification

Recombinant GAL4-AID, wild-type (WT) and mutant (Mt), were expressed and purified from *E. coli* C41 cells. Cells were transformed with pSPM589 and pDMF1620 (described in Figure 9), were grown until they reached OD600=0.6. GAL4-AID expression was induced by addition of 0.4 mM IPTG in the presence of 10 μM $ZnCl_2$ for 14 h at 16 °C. After the overnight culture, the bacterial pellet was washed in ice cold water and resuspended in 20 mL (per 1 L of culture) of extraction buffer. The cell lysate corresponding to 1 L culture was sonicated for 12', duty cycle 5 in a Branson sonicator, at maximum power. The sonicated suspension was centrifuged at 165,000 g for 90' at 4 °C, and the supernatant was collected and filtered in a 0.2 μm filtering device. The protein

solution was then incubated with 0.5 mL (per 1 L of culture) of Talon Metal affinity resin (slurry, 50 %) (Clontech) for 90' at 4 °C. The resin was washed in Bio-Rad Gravity columns (Bio-Rad) by addign 5x bed volumes of extraction buffer and proteins were eluted in 2.5x bed volumes of extraction buffer containing 200 mM imidazole. After elution, proteins were concentrated in VIVASPIN 500 columns (Sartorius), with a Molecular Weight Cut-Off (MWCO) of 10,000 Daltons (as GAL4-AID is 45,000 Daltons).

2.2.7 - Western Blot

Protein samples were mixed with laemmli Buffer to a final concentration of 1x laemmli, denatured at 95 °C for 5' and loaded on a 12.5 % Poly-Acrylamide gel. Electrophoresis was performed for 90' at 120 Volts in Running Buffer. After electrophoresis, proteins were transferred from the gel to a PVDF membrane (Westran 0.2 µm) for 2 h in Transfer Buffer at 100 V at 4 °C. After the wet transfer, the membrane was blocked with 5 % milk in TBST for 1 h at RT to saturate the non-protein containing areas and to prevent unspecific binding of the antibodies. Once the membrane was blocked, it was incubated with specific antibody (indicated in table 5) 2h at RT. After 3 washes for 10' in 1x TBST vigorously shaking, the membrane was incubated with the specific secondary antibody for 45' at RT. After 3 washes in 1x TBST, signals were detected with ECL Luminata at ChemiDoc™ MP System (Bio-Rad).

2.2.8 - Coomassie

Protein samples were mixed with laemmli buffer to a final concentration of 1x laemmli, denatured at 95 °C for 5' and loaded on a 12.5 % Poly-Acrylamide gel. Electrophoresis was performed for 90' at 120 Volts in Running Buffer. To visualize protein in Poly-Acrylamide gels, coomassie Brilliant Blue Staining was performed. The gel was incubated with coomassie Brilliant Blue Staining Solution for 4 h to ON at RT upon mild agitation. Gel was destained in De-staining Solution, which was exchanged until the excess dye was removed from the gel. Images were collected using the ChemiDoc™ MP System (Bio-Rad).

2.2.9 - Gel mobility assay

300 ng of plasmid DNA was either mixed with water (untreated) or treated with GAL4-AID or with restriction enzymes and incubated for 30' at 37 °C. After treatment, to stop the reaction and to digest added enzymes 15 µl of lysis buffer were added and incubation was performed at 56 °C for 2 h. The topology of the plasmid was then assayed on a 0.8 % agarose gel. After allowing migration of the DNA at 5–10 V/cm for 16 h at 4 °C the gel was soaked in 1x TBE containing SYBRH Safe DNA gel stain (Invitrogen) for 30'. Plasmids were visualised by analysing the gen with a ChemiDoc™ MP System (Bio-Rad).

2.2.10 – Rifampicin assay

BW310 E.Coli strain (lacking UNG) were transformed with plasmids pDS4501 and pSPM589 (expressing G-AID). Plates were grown at 37 °C for 16-18h. 8 clones per transformant (with same dimension) were collected and diluted in 1.5 ml of 2xTY in 96 well dish. After vortexing the plate for 5 min at 1300 rpm, from each 1.5 mL of culture 1 µL was transferred and diluted 1:1000 in 1.5 ml of 2xYT, supplemented with the inducing factor (0.3 mM IPTG), the co-factor for AID activity (10 µM $ZnCl_2$) and the specific antibiotic for selection (100 µg/ml carbencillin). Cultures were grown O.N at 37 °C upon constant vortexing at 1300 rpm. For putative mutation observation 200 µL of randomly chosen 8 cultures of each transformation were plated onto rifampicin LB plates. For a preparation of 150 rifampicin plates 0.4 g of rifampicin was dissolved in 9 ml DMSO and added to 5 liters of LOW salt LB. Plates containing rifampicin were protected from light during storage. For viability observation 6 cultures of each transformation were diluted $1/10^6$ and 50 µl of each diluted culture were plated on ampicillin LB plates. The frequency of deamination affects the number of cells acquiring resistance. Mutation frequency was determined by counting the number of rifampicin resistant colonies vs the number of viable colonies.

2.2.11 - ssDNA oligonucleotide deamination assay

The GAL4-AID ssDNA oligonucleotide deamination assay was performed as follows: 2.5 pmol of the biotin-tagged and fluorescein-tagged oligonucleotide 163 (table 4) was mixed with 10 µg of RNase A in 10 µl reaction buffer R, denatured for 3' at 90 °C. Reaction were stopped by transferring tubes in ice-water. Oligonucleotides were incubated for 15' at 37 °C with 10 - 50 ng of recombinant protein. Reactions were stopped by addition of 100 µl of H_2O. 8 µl of streptavidin magnetic beads (Dynal M270, Invitrogen) were added to 1 mL of TE-1000 and incubated with the oligonucleotides for 15'. Beads were collected, washed twice in TE-1000 preheated to 70 °C, once in TE at RT, and resuspended in Uracil-DNA Glycosylase (UDG, New England Biolabs) reaction mix, prepared according to manufacturer protocol, and incubated for 1 h at 37 °C. Cleavage reactions were stopped by addition of 20 µl 0.4 % fushin in formamide and denaturation at 90 °C for 3'. After quenching on ice, samples were resolved on 17.5 % TBE PAGE urea gels at 200 V and visualised using a Typhoon scanner (GE Healthcare) for fluorescence imaging (Filter: 526 SP (532 nm), Laser: Blue 488 nm).

2.2.12 - dsDNA oligonucleotide DNA glycosylases assay

The ds DNA oligonucleotide glycosylases assay was performed with oligonucleotides 164 and 166 (table 4). The oligonucleotides were mixed in 50 mM NaCl, Tris-HCl pH 8.0, denatured for 5' at 90 °C, and annealed by slowly cooling to RT (30'). ds oligonucleotides (containing the mismatch dU:dG) were purified using streptavidin magnetic beads and incubated with cellular extracts for 15' at 37 °C. To enhance strand cleavage NaOH (0.3 M) was added. Samples were analysed as above. To inhibit UNG activity present in the cell extracts UGI was added to the extracts 10' (on ice) prior to the incubation with the oligonucleotides. Samples were resolved 17.5 % TBEPAGE urea gels as previously described.

2.2.13 - In vitro resolution assay (IVR)

The 'in vitro resolution assay' (IVR) was developed in the Petersen-Mahrt laboratory (159) to quantify GAL4-AID activity and repair achieved upon recruitment of DNA repair proteins. IVR can be divided into two main reactions:

1) Deamination reaction: 300 ng of supercoiled pGL4.31, 10 - 50 ng of GAL4-AID were mixed in IVR buffer 1X and incubated for 30' at 37 °C. During this phase, GAL4-AID can deaminate pGL4.31.

2) Repair reaction: 0 to 40 µg of cellular extract was added to the deamination mix in the presence of 0.2 µg/µl aphidicolin, 0.05 mM dNTP-bioC (without dCTP and 0.05 mM biotinylated-dCTP (Invitrogen) or 0.05 mM dNTP-A (without dATP and 0.05 mM biotinylated-dATP (Invitrogen), and incubated for 30' at 37 °C. When specified, UNG inhibitor (UGI - New England Biolabs) or PCNA inhibitor T2AA (T2AA is the inhibitor of the PIPbox peptide/PCNA interaction – (161)) were added to the extracts for 10' on ice prior to the addition to the repair reaction.

After the repair reaction, free Bio-dNTPs were removed with the Qiaquick PCR purification kit (Qiagen) and plasmids were eluted in 100 µl of elution buffer. From this eluate 2 µl (further diluted 10 fold) was used as 'input'. The remaining eluted plasmid was subjected to Dynabeads M-270 streptavidin magnetic bead (Invitrogen) purification in TE – 1000. Mixtures were gently rotated for 15' at RT and beads were collected by incubation for 2' on a DynaMag-2 magnet (Invitrogen). The following washes of the beads were applied: TE – 500 at RT, TE – 100 at 65 °C, TE – 50 at 65 °C and TE -0 at RT. Beads were finally resuspended in 100 ml TE (without Tween-20) and subjected to quantitative real-time PCR.

2.2.14 - RT-PCR and statistical analysis

The RT-PCR reaction was performed in a final volume of 10 µl, containing 2 µl of the bound DNA bead mixture, 5 µl of the LightCycler 480 SYBR Green I master (Roche Applied Sciences) and specific primers (table 4). The reaction was monitored in a LightCycler 480 Real-Time PCR System (Roche Applied Sciences); with the 'input' DNA

analysed in parallel as reference. Ct values for the biotinylated-DNA were correlated to the Ct values for the input DNA. The IVR results were presented either as relative quantification (fold-change) or absolute (% of input). For fold-change, all samples were compared to their input (defined as the plasmid DNA that was present in the reaction prior to the streptavidin isolation) and then the FE alone sample (or another samples specified) was used as reference and set to one. The fold change was calculated according to the ΔΔ-Ct algorithm, represented as: Fold Change = $2^{\{-[(Xo-Xi)-(Ro-Ri)]\}}$, where X - sample; R - reference extract alone; o - output (post bead isolation); i - input (pre-bead isolation). Alternatively, in the % of input analysis the Ct qPCR values of input and output were converted to an absolute amount of DNA based on a standard curve (performed in parallel to the analysis), with the amount of isolated biotinylated-plasmid being expressed as a percentage of the initial amount of plasmid (input).

2.2.15 - Atomic Force Microscopy

Atomic Force Microscope (AFM) was used to image single molecules at nanoscale level, with high resolution. The heart of the AFM is the cantilever, a plate spring fixed at one end; while at the other end it supports a sharp tip. The pointed tip is brought into contact with the sample and moved across the sample surface line by line (Figure 9A). AFM images are obtained by measurement of the force created between a sharp tip and the surface of the sample (162). In particular, attractive or repulsive forces between the sample and the tip are converted into a deflection of the cantilever, which is detected by a laser beam reflection from the back of the cantilever onto a photodetector (Figure 9A). Deflection of the cantilever results in a change of the laser reflection angle and a change in the voltage signal.

AFM visualization was initially performed in collaboration with Alessandro Podestà, Università degli studi di Milano using a Bruker Nanoscope. Further images are obtained from JPK Nanowizard 3 AFM at IFOM. Using a tapping mode (Bruker) or AC mode (JPK) the tip was intermittently contacting the surface with constant oscillation, without being trapped by adhesive forces coming from sample (Figure 9B).

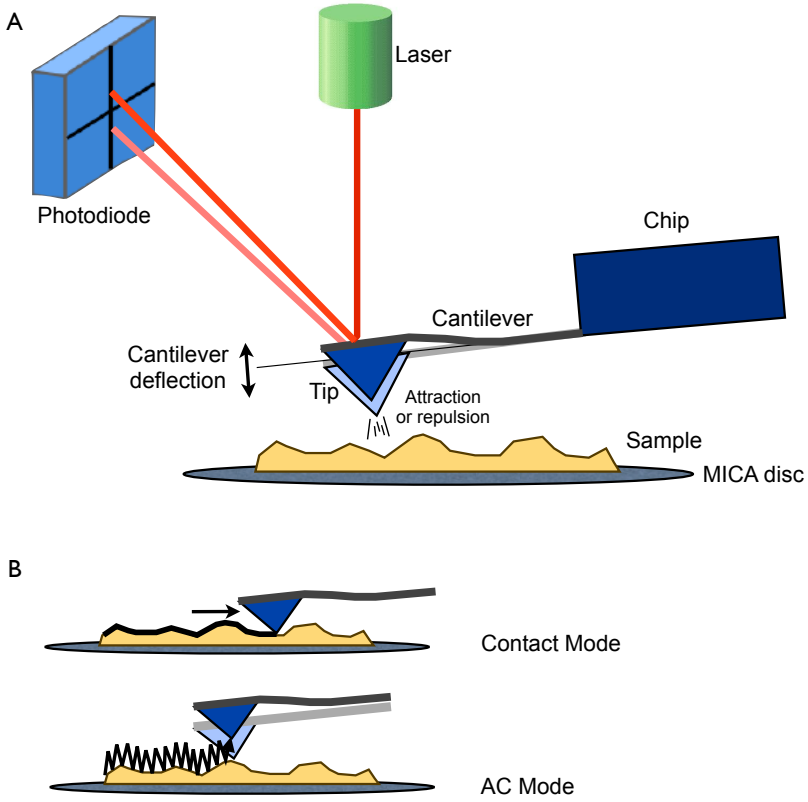


Figure 9 – Drawing of basic principles of Atomic Force Microscopy (AFM).

A) A cantilever, with a very small tip (or probe), moves along the surface and experiences atomic forces. The tip is on the end of a cantilever, which deflects when the tip encounters features on the sample surface. An optical laser beam senses this deflection. The end of the cantilever is reflecting the laser onto a segmented photodiode that magnifies small cantilever deflections into large changes in the relative intensity of the laser light. **B)** AFM imaging in contact mode the tip continuously sense the sample surface; soft samples can thus been damaged or distorted. AFM imaging in tapping mode or AC mode minimizes this problem by having the tip oscillate over the sample, making only brief intermittent contacts. This mode also provides additional information about the sample surface that corresponds to the height image.

AFM samples preparation

DNA was diluted in mM $MgCl_2$ and 10 mM HEPES or in 1 mM $NiCl_2$ and 10 mM HEPES. the standard concentration of DNA used was 1 ng/µl. 20 µL of the DNA solution was applied on the MICA surface, incubated for 5' and MICA surface was washed with 3x with 500 mL of MilliQ grade H_2O. Finally the MICA was dried with N_2 flow for 5', with and angle of 75 ° relative to the MICA. After DNA deposition, proteins were added to DNA, in the same buffer. MICA disc was washed and dried as before.

When the MICA disc was coated with poly-L-ornithine (SIGMA) to create a positive surface for DNA binding, 20 µL of a 0.01 % poly-L-ornithine solution was spread on the

MICA surface. After 5' incubation the MICA surface was washed three times with MilliQ grade H_2O.

2.2.15.1 - In Vitro nucleosome reconstitution

160 fmols of DNA were incubated with different concentrations of recombinant human histone octamer (Tebu-Bio S.R.L.) (0 to 280 fmols) in 20 µL of 2M NaCl buffer for 30 min at RT. A subsequent dialysis reaction was performed reducing the NaCl concentration from 2M to 1M (2h at RT), to 0.65M (2h at RT) to 0.2M (O.N at 4 °C). The final solution (fully reconstituted nucleosomes) in 0.2 M NaCl buffer was loaded onto MICA polyornithine coated surface and imaged in AC mode in AFM.

CHAPTER 3 - RESULTS

3.1 - DEVELOPMENT AND VALIDATION OF A NEW BIOCHEMICAL ASSAY TO STUDY G-AID INDUCED LESIONS RESOLUTION IN FROG EGG EXTRACT (FE)

Previous work in our lab defined the biochemical characterisation of *in vitro* AID enzymatic activity (119). Those results provided the kinetics of AID-induced lesion generation, but less was known about the extent of DNA repair of those uracils. We decided to use the characterized AID damaging agent as a tool to determine the contribution of various DNA repair pathways to lesion resolution in a physiological context.

A new biochemical assay, named in vitro resolution (IVR) was designed to investigate the molecular mechanisms of damage resolution. The method is composed of two phases: 1), AID was targeted to a ssDNA substrate and 2) a cell extract was used to repair the AID-induced lesions (Figure 10 and (159)). To target AID to a DNA substrate *in vitro*, a fusion protein was generated. A bacterial expression vector (pSPM589) was cloned fusing the cDNA sequence for human AID (AID) to the DNA binding domain of GAL4 (G) generating G-AID. G-AID is targeted to a supercoiled (SC) plasmid harbouring the GAL4 binding sites (UAS). The SC topology of the plasmid provides both the dsDNA for GAL4 binding and the ssDNA required for AID activity. After G-AID creates dU lesions, in the second phase of IVR the lesions are repaired by cell extracts competent for DNA repair. To establish the IVR, we utilised *Xenopus leavis* egg extract (FE). This type of extract has been used in the past as a model extract to study the control of cell cycle progression, apoptosis, nuclear formation, nucleocytoplasmic transport, spindle microtubule dynamics, chromatin structure, and importantly DNA repair (163). DNA topoisomerases in the FE relax the plasmid forming a dU:dG mismatch, providing a substrate for multiple DNA repair pathways. During repair of the uracil lesion a biotinylated-dCTP (bio-dC) can be incorporated into the plasmid, providing a tag for isolation and quantification. Importantly, DNA repair pathways that excise several bases surrounding the lesion can also incorporate other dNTPs; hence addition of biotinylated-dATP (bio-dA) can be used to

identify specific DNA repair pathways subtypes. Biotinylated plasmids are isolated via streptavidin beads, and the amount of deaminated and repaired plasmid is determined by quantitative real-time PCR. Ct values for the biotinylated-DNA were correlated to the Ct values for the input DNA (total plasmids prior to bead isolation). IVR results can be represented as relative (fold change) or absolute (% of input) quantifications.

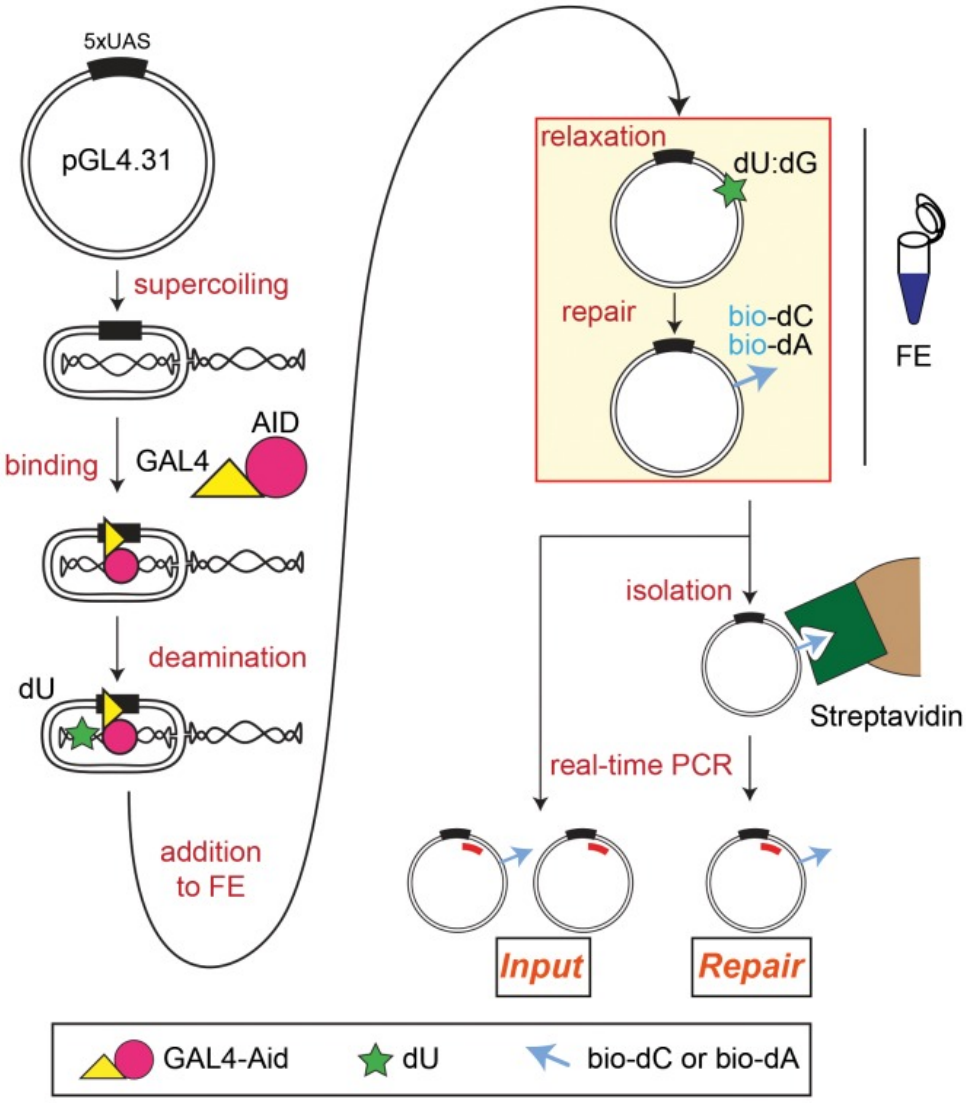


***Figure 10* – Schematic representation of the IVR assay**

The supercoiled pGL4.31 (containing 5x UAS sequences for GAL4 binding) was incubated with G-AID protein (yellow triangle and red circle). During incubation at 37°C for 30 min G-AID randomly deaminated dC creating dU lesions in ssDNA (green star). The repair phase (yellow box) was carried out by the addition of FE in the presence of biotinylated dCTP (bio-dC) or biotinylated dATP (bio-dA) - (blue arrow), along with normal dNTPs. After DNA repair of the dU lesion the biotinylated-tagged DNA was isolated via magnetic streptavidin beads. Deaminated and repaired plasmids were analysed through quantitative real-time PCR (red bar), and compared to input values (total amount of plasmids not streptavidin isolated).

3.2 - VALIDATION OF THE IVR ASSAY IN FE

IVR is a sensitive method; we ensured that DNA lesions occurring during IVR were exclusively due to G-AID induced uracils. Each component used in IVR assay was evaluated for non-specific DNA damaging activity. In particular, we checked G-AID protein after purification form *E. coli* and if the extracts were retaining nicking activity. As a control, to ensure G-AID catalytic activity was the only requirement needed to activate repair mechanisms, a G-AID mutant protein (G-AID mt) was purified at the same time. The catalytic mutant was generated by substituting the negative charged glutamic acid (E58) in the active site of hAID with a glutamine (Q58) (pDMF1620). This mutant is unable to provide the required proton donor in the active site (164)

To validate the assay we checked: 1) G-AID purity with Western blot and coomassie, 2) G-AID nicking activity on SC plasmids, 3) G-AID catalytic activity and 4) extracts nicking activity on SC plasmids.

3.2.1 - G-AID protein production optimization leads to the purification of an endonuclease free protein

In order to study DNA lesion generation and resolution of the DNA damaging protein AID *in vitro*, we needed to isolate AID from *E. coli*. The His-tagged G-AID fusion proteins were isolated from bacteria lysates using TALON beads. Eluates were quantified and analysed for the presence of co-purified contaminant proteins (Figure 11). Full length G-AID (FL G-AID) and two break-down products, due to sensitivity to proteolytic cleavage between GAL4 and AID, were visible both in Western blot (Figure 11A – arrows) and coomassie analysis (Figure 11B - arrows). Quantitative analysis revealed that 280 ng/µL of FL G-AID were purified (Figure 11C); in total 28 µg of pure FL protein were extracted from 2 L of bacterial culture. However, coomassie analysis revealed additional bands (Figure 11B). Through a mass spectrometry approach, proteins associated with AID (purified from *E. coli* with the same approach used for G-AID) were identified as 60 kDa chaperonin (or

GroEL), elongation factors, and ribosomal units required for prokaryotic translation (data not published – Gopinath Rangam). The more prominent band in coomassie can be identified as GroEL.

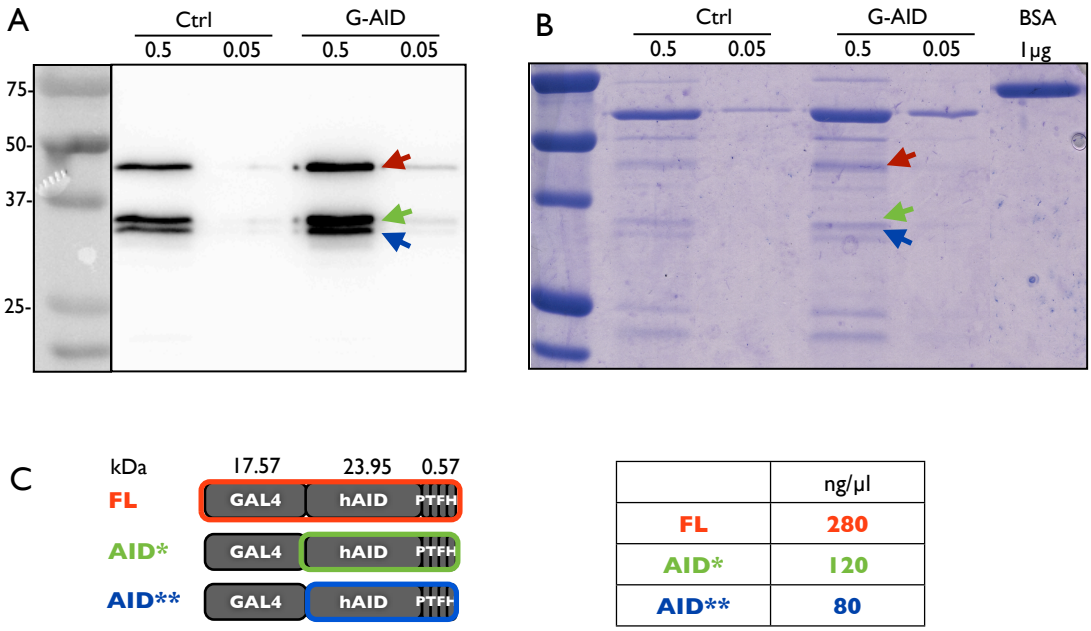


Figure 11 – Quantitative and qualitative analysis of G-AID purified protein

A) G-AID protein is detected with anti-His antibody. 0.5 and 0.05 µL of concentrated protein were loaded on the gel. FL protein is visible both in Western Blot and in coomassie analysis (red arrow). Two other specific products are recognized by anti-His antibody and correspond to breakdown products (green and blue arrows). **B)** In the coomassie staining additional bands appeared. Based on previous experiments, the more prominent band at 60 kDa was identified as the 60 kDa chaperonin. **C)** FL protein (red) and breakdown products (green and blue) were characterized for dimension and quantified. FL protein harbour human AID fused to the DNA binding domain of GAL4 (GAL4). Additional tags are: protein C tag (P), TEV sequence to cut off tags to the protein (T), Flag tag (F) and His-tag (H). FL protein is 42kDa and was quantified from coomassie gel using BSA as standard. FL protein is 280 ng/µL.

To analyse whether the co-purified factors exhibit potential nuclease activity that could interfere with the IVR assays, the G-AID eluate was incubated with supercoiled pGL4.31, a plasmid containing 5x G-DNA-binding sites (UAS). Topology and integrity of the plasmid were monitored with a gel mobility assay (Figure 12). The untreated plasmid displayed three topological isoform: concatenated, nicked (or open circular) and supercoiled (Figure 12 - (-) samples). When plasmids were incubated for 30 min with G-AID, topology was unvaried with the exception of a faint linearized band, due to contaminant nuclease activity (Figure 12A - green arrow). To separate G-AID from co-purified nucleases, the

eluate was fractionated and evaluated for linearization (Figure 12B). DSBs appeared only when plasmids were incubated with the first eluate fraction (Figure 12B - orange arrow), indicating that nucleases elute early from the TALON beads. In all subsequent G-AID preparations the first eluate fraction was excluded from the final concentrated product. After this optimization G-AID WT and G-AID mt proteins displayed the same topological profile as the mock-incubated plasmid, with no nucleases contamination (Figure 12C). To maximize the SC topology, needed for the IVR assay, the plasmid isolation was performed at 4°C (160). Under this condition more than 90 % of the plasmids were in SC conformation (Figure 12C).

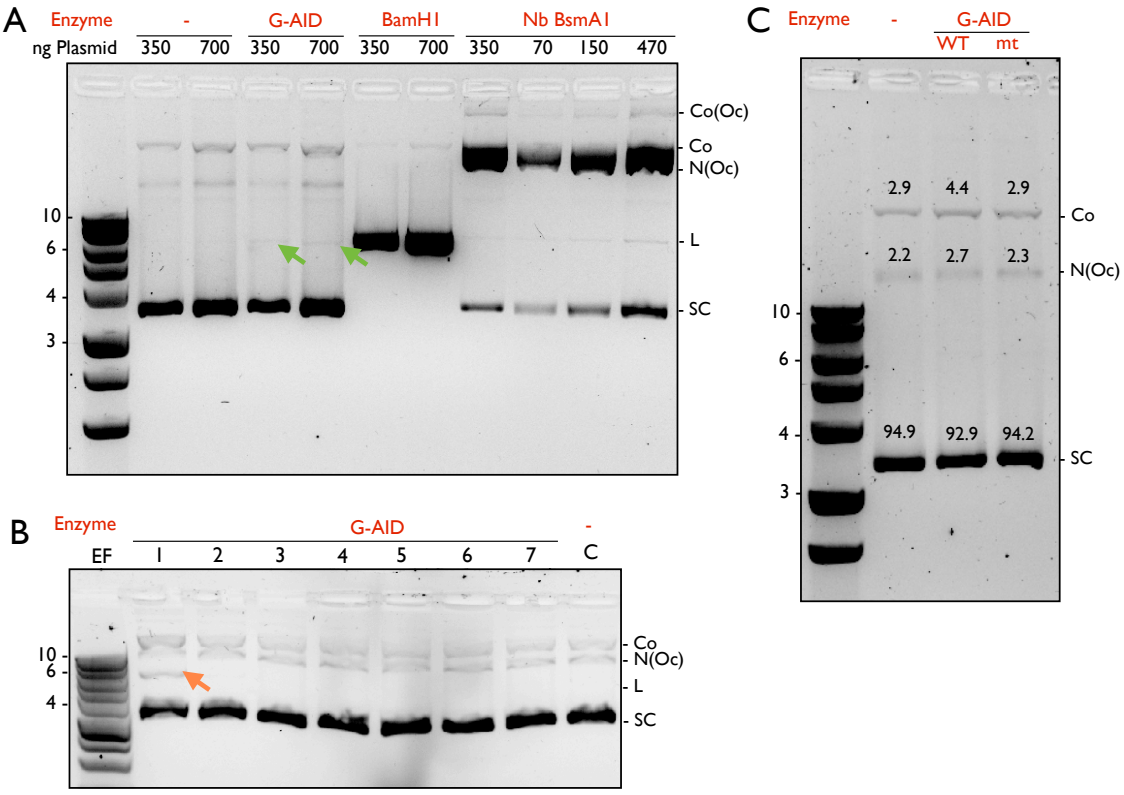


Figure 12 – DNA topology following G-AID induced-damage

A) Various forms of pGL4.31 were analysed on a 0.8 % agarose gel run at 5 – 10 V/cm for 16 h at 4°C. Supercoiled DNA (SC) migrated faster than linear (L – 6 kb, digested 2h at 37°C with BamH1), while nicked/open circular (N/Oc – digested 2h at 65°C with Nb BsmA1) or concatenated DNA (Co) migrated slowest. When G-AID was incubated with the plasmid a faint linear band appeared (green arrow). **B)** Different elution fractions of G-AID (lane 1 to 7) were incubated with pGL4.31 and analysed on agarose gel: in lane 1 linear DNA was visible (orange arrow). **C)** G-AID WT or mt concentrated proteins (without the first elution fraction) were incubated with pGL4.31 and analysed on agarose gel. % of plasmid topologies (calculated with Chemidoc™ Image Lab™ Software) are indicated in the Figure.

3.2.2 - G-AID protein retains its catalytic activity

In order to use G-AID as a natural DNA mutator, we checked its activity both *in vivo* and *in vitro*. *E. coli* cells lacking *UNG* activity (BW310 strain) and thus unable to repair AID lesions, were transformed with an expression vector containing either AID or G-AID cDNA (pDS4501 or pSPM589, respectively). After induction of protein expression, only cells harbouring a deaminated and mutated genome can acquire resistance to rifampicin antibiotic. Mutation frequency is a direct measure of AID deamination activity and was determined by establishing the number of resistant colonies *versus* the viable colonies (165). As expected, AID was able to mutate *E. coli* genome, compared to cells transfected with a mock plasmid (pETduet). G-AID, despite the structural implication of the fusion of AID with GAL4, displayed mutagenic activity *in vivo*, reduced only by 2.5 fold compared to the unmodified protein (Figure 13A).

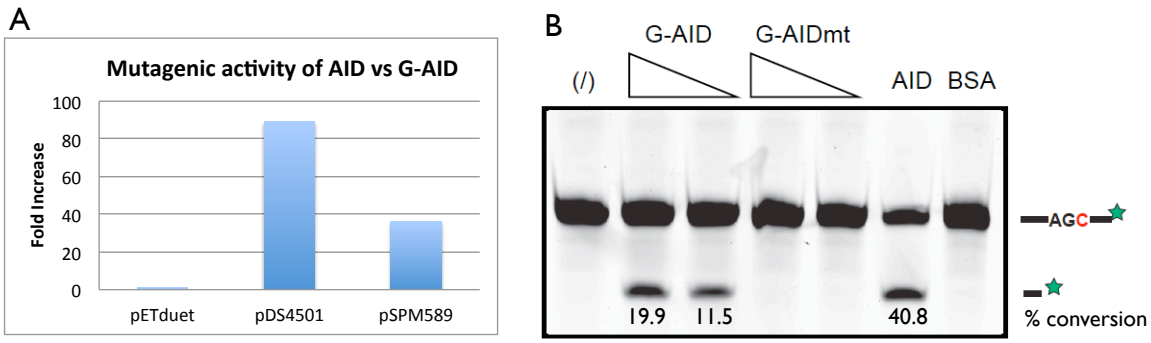


Figure 13 - G-AID deaminates cytosine to uracil in E. coli and in vitro

A) *E. coli* strain *UNG* deficient (BW310) was transformed with AID or G-AID expressing plasmid (pDS4501 or pSPM589, respectively) and pETduet as control. Median number of colonies grown on rifampicin was divided by average number of colonies grown on ampicillin plates. Results from *E. coli* transformed with pETduet were set to 1. **B)** ssDNA oligonucleotide deamination assay was performed using a 40mer oligonucleotide (spm163) containing a single cytosine. Two concentrations (20 ng and 200 ng) of G-AID and G-AID mt, as well as untagged AID (200 ng) and BSA (1 µg) were incubated with 2.5 pmols of oligonucleotides for 30 min at 37°C. Image-J program was used to calculate bands relative area and density. Further calculations were performed to remove background specific for each single band. % of conversion was calculated as: intensity lower band*100/[intensity upper band + intensity lower band]. 40 % of the oligonucleotides were deaminated by AID, while G-AID displayed activity *versus* 20 % of the substrates.

The *E. coli* results were confirmed *in vitro* with a ssDNA oligonucleotide deamination assay (165). Only if AID deaminates the dC in the WRC motif is UNG able to remove the lesion and APE or NaOH to cleave the oligo. Hence, oligo cleavage is proportional to deamination activity. From the assay it is clear than G-AID WT retained the deamination activity *in vitro*, while G-AID mt was not catalytically active (Figure 13B). The % of conversion confirmed the *E. coli* results: G-AID was roughly 2 fold less active than AID.

3.2.3 - G-AID-independent DNA repair activation is excluded from IVR readout

After confirming that G-AID is not contaminated by nucleases and retains its catalytic activity, we checked whether G-AID activity is the only factor able to induce DNA repair activation and hence bio-dC incorporation. *Xenopus* cell-free extracts support efficient DNA replication of added DNA substrates (166,167). In the IVR system, replicative DNA polymerases could induce replication of the SC plasmid and hence bio-dC incorporation independently from AID-induced damaged. Replication was inhibited by aphidicolin in the extract, and is added before the repair phase. Indeed, in the presence of aphidicolin the number of plasmids incorporating bio-dC was reduced by over 40 fold (Figure 14). Unless indicated, subsequent IVR experiments were performed in presence of aphidicolin to exclude DNA repair independent incorporation of bio-dC.

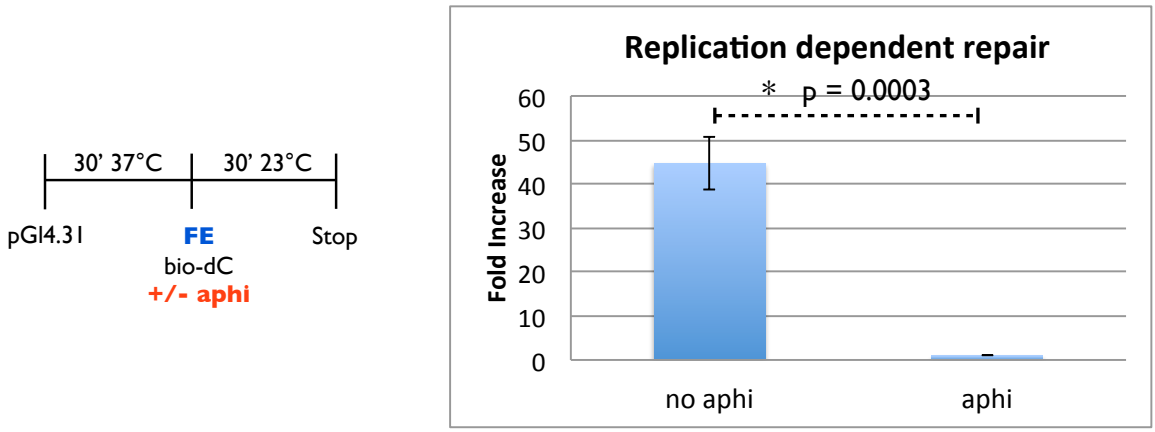


Figure 14 – Quantification of FE replication dependent repair

IVR was performed without G-AID protein. Repair was performed by FE (with bio-dC) in the presence or absence of aphidicolin. Samples treated with aphidicolin were set to one. Aphidicolin treatment reduced the IVR by 40 fold. Error bars indicate SD (n = 3). Time line of experiment is

shown next to the graph. Statistical analysis (t-test) was performed on differences of indicated fold increase. * p value is indicated in the figure.

Plasmid integrity was also monitored as SSBs or DSBs generation can lead to DNA repair activation (AID independent). Indeed, when nicking enzymes were used to create a damaged plasmid, a 4-fold increase in repair was registered, compared to non-treated plasmid (Figure 15). When SSBs were sealed by addition of T4DNA ligase, repair pathways activation was reduced to the levels of the unmodified substrate (Figure 15), indicating that the 4-fold increase was exclusively dependent on nicking activity.

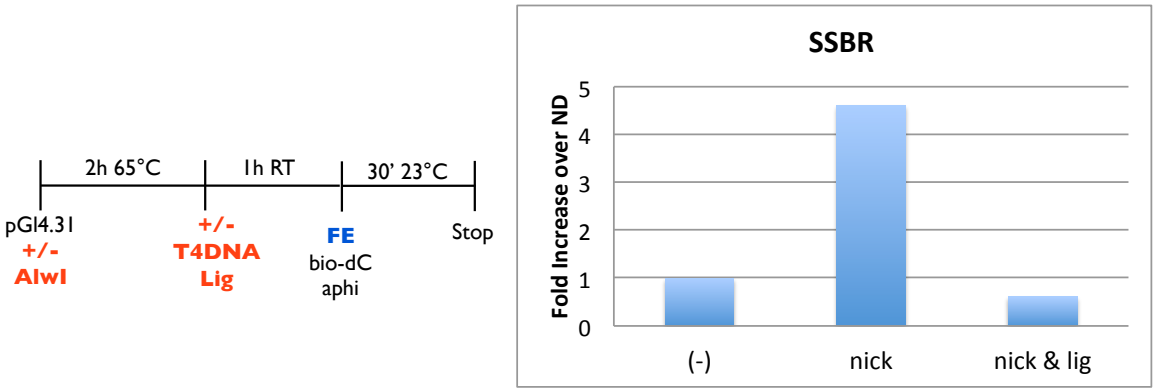


Figure 15 – Quantification of SSBR in FE

IVR was performed using a modified substrate. Plasmids were either digested or not with AlwI nicking enzyme. After enzyme heat inactivation, plasmid was treated or not with T4DNA ligase. Purified plasmids were then repaired by FE (with bio-dC) in the presence of aphidicolin. Results from untreated plasmids were set to one. Nicking treatment increased the IVR by 4 fold. Time line of experiment is shown next to the graph.

Integrity of the plasmid was checked also with gel mobility assay (Figure 12C). Quantification of the ratios between supercoiled, nicked, and concatenated plasmids topologies showed that optimal G-AID preparation did not impact original plasmid topology (Figure 12C).

3.3 – KINETIC PROFILES OF AID-INDUCED DAMAGE AND FE DEPENDENT REPAIR

Once we could control the IVR readout, we performed the assay in the presence of G-AID and different controls (Figure 16). Only when G-AID was incubated with SC plasmid and then added to FE could we detect a significant biotinylation of plasmid. Using only the

GAL4-DNA binding domain (G-DBD) or the catalytic mutant of G-AID we did not induce any bio-dC incorporation, indicating that AID catalytic activity is required to initiate DNA damage and resolution in the FE. Interestingly, when untagged AID was used in IVR, less DNA repair was observed compared to G-AID, while from *in vitro* oligonucleotide assay AID was more active than G-AID (Figure 13B). GAL4 binding domain helped AID activity on the SC substrate during IVR assay.

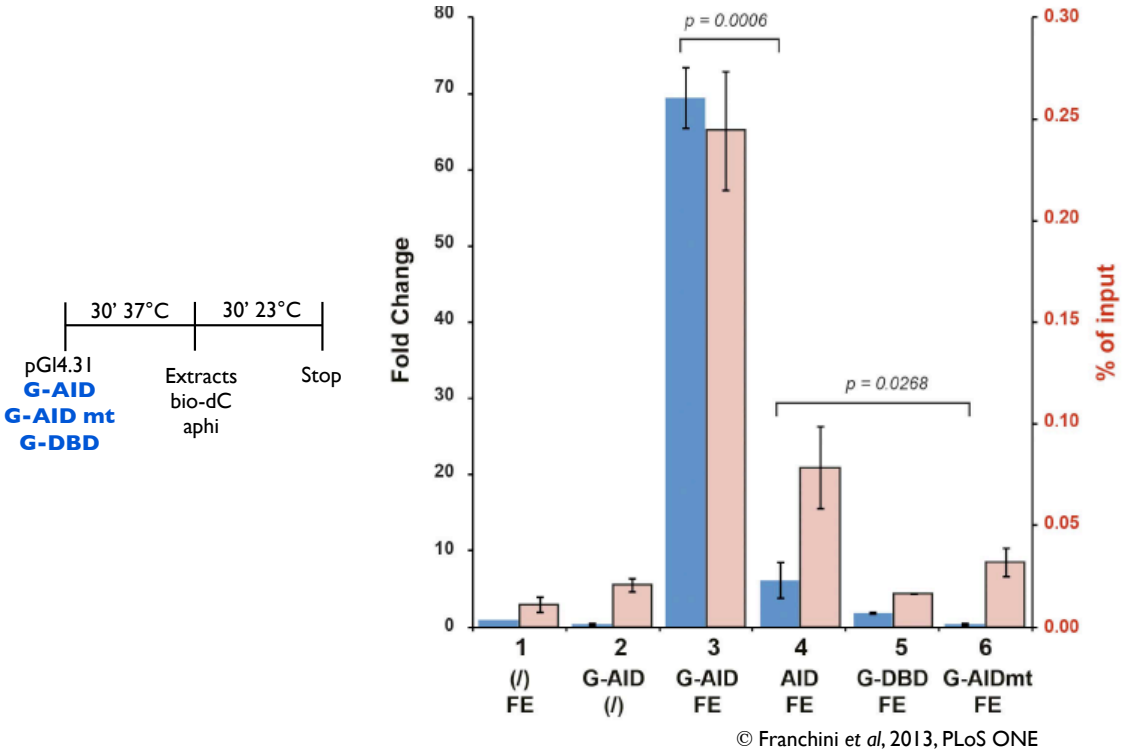


***Figure 16* – AID-induced lesions repair in FE**

AID-induced lesions are repaired in FE. SC plasmids were treated (or not - bar 1) with the G-AID WT or mt proteins (the mutant used in this assay is C87R (164)) and then incubated in FE (or not - bar 2), isolated, and quantified by qPCR. The blue bars represent the ratio (fold change) of the amount of recovered plasmids from reactions carried out in the presence of G-AID (bar 3), untagged AID (AID; bar 4), GAL4 DNA binding domain (G-DBD; bar 5), or mutant G-AID C87R (G-AID mt; bar 6) versus levels of plasmids recovered from reactions that did not contain G-AID (FE alone, set to 1; bar 1). Pink bars represent the absolute recovery of each treated sample in relation to its input. Error bars indicate SD (n = 3). Statistical analysis (t-test) was performed on differences of indicated fold change (brackets), with p values shown. Time line of the experiment is shown above the graph.

Once the IVR system relied on G-AID induced DNA repair activity only, we optimized the concentration of G-AID protein and the concentration of FE to maximize their efficiency

(Figure 17). Different concentrations of G-AID WT or mt were used in the IVR assay to establish the kinetic profile of DNA damage-induced repair activation (Figure 17A-B). In Figure 17A, both G-AID WT (filled bars) and G-AID mt (empty bars) results are shown, to empathize that at higher concentration G-AID mt is not able to initiate DNA lesions and to provide a visual record to the background activity of the IVR system. In Figure 17B and in all subsequent results G-AID mt readout was subtracted from G-AID WT results.

Moreover, we can demonstrate that repair proteins are responsible for IVR readout by using UGI, a small phage peptide known to specifically inhibit UNG enzymes (168). Since UNG is a DNA glycosylase able to recognize uracil in DNA and activate BER pathway (29), its inhibition should result in a reduction in repair activation upon AID-damage. Indeed, when UGI was added to FE activity was reduced by 56 % to 88 % (depending on G-AID concentration) (Figure 17B).

Using the concentration of G-AID that gives maximum IVR activity (0.66 pmols – Figure 17B), we performed FE titration (Figure 17C-D). After a linear increase in activity with increasing amount of FE, 50 µg of extract were less efficient. While using even higher concentrations of FE, we detected again a linear correlation until it plateaued-out near 200 µg (Figure 17C). We did not pursue this observation, and hypothesised that this trend could be the result of the interplay between two different repair pathways, whose activation may require a threshold amount of active repair protein.

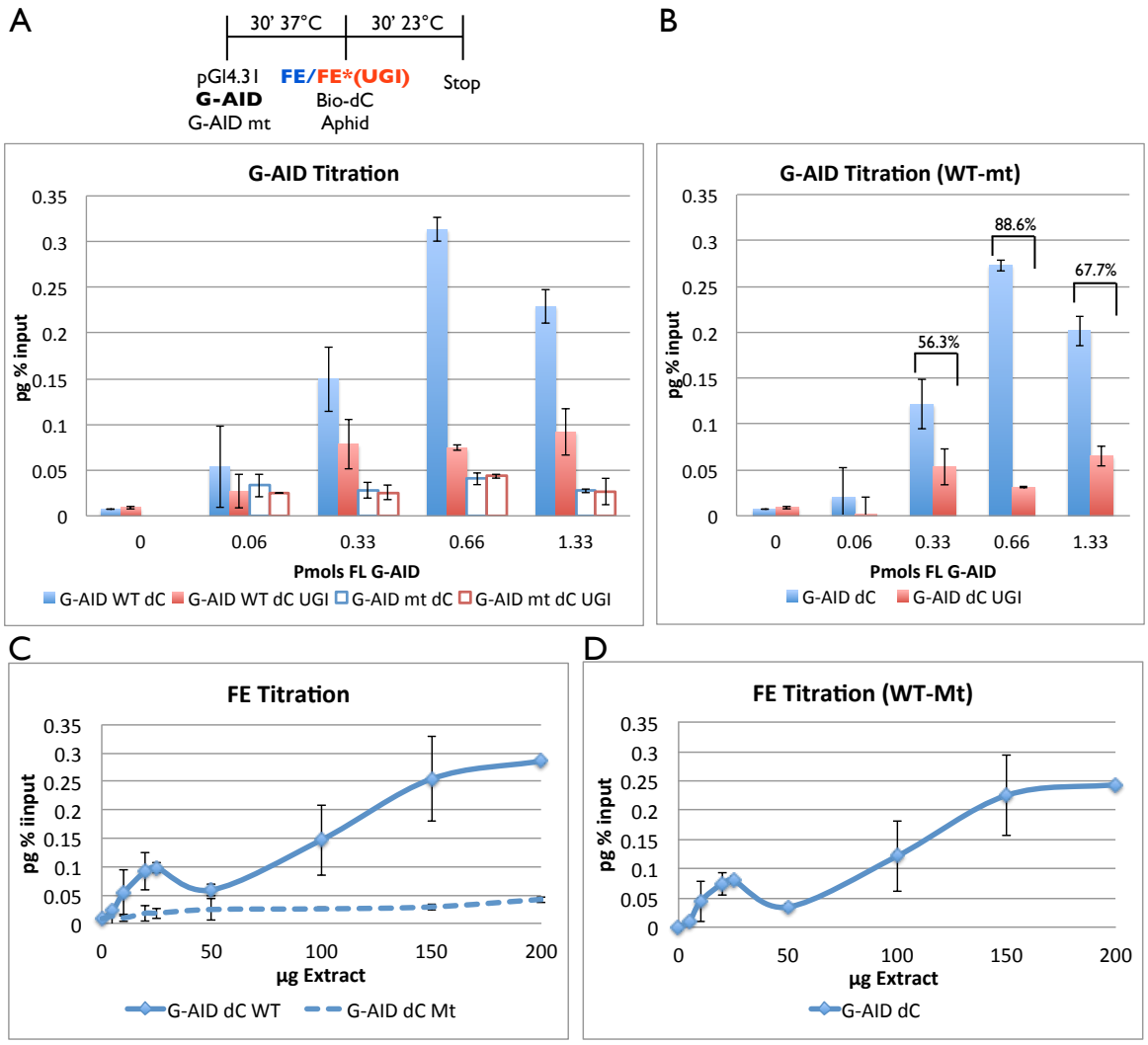


Figure 17 – G-AID and FE titration in IVR assay

A) SC plasmids were treated with increasing concentrations of G-AID WT (filled bars) or G-AID mt (empty bars) and then incubated with FE, in the presence of aphidicolin (Aphid). The blue bars represent the absolute recovery of each treated sample in relation to its input. 0.66 pmols of G-AID deaminated 0.3 % of the total plasmids. Red bars represent the repair pathways activated when 1 unit of UGI is incubated with 150 µg of FE 10 min on ice prior repair reaction. **B)** G-AID mt results were subtracted from G-AID WT results. From 55 % to 88 % of the total repair activity was inhibited by UGI, depending on G-AID concentration used. Error bars indicate SD (n = 3). Time line of the experiment is shown above the graph. **C)** SC plasmids were treated with 0.66 pmols of G-AID WT or mt and then incubated with different amounts of FE. **D)** G-AID mt results were subtracted from G-AID WT results. Error bars indicate SD (n = 3). Time line of the experiment is shown above the graph.

3.4 - IVR IS SUITABLE FOR QUANTIFYING DNA REPAIR ACTIVATION IN DIFFERENT CELL LINES

To obtain a more cancer and B cell line relevant biochemical analysis and to link biochemical data with biological information, I established the IVR using human cell lines. Using different cell lines (from different tissues) we can correlate functional aspects of DNA damage repair with physiological aspects of different cellular milieu. This allows us to determine if a cell's microenvironment can influence AID-induced lesions processing and analyse the coordination between different DNA repair networks required for maintaining stability. We performed the IVR resolution phase using extracts from: human B cell lines (Ramos and BL2) and human epithelial cell lines (HelaS3 [cervical] and Colo704 [ovarian]). Those cell lines were selected because of their ability to grow fast and in suspension, allowing for a more reproducible extract preparation.

Our emphasis was to compare B cells environment with non-B cells environment and to characterize single repair pathways activation after DNA damage by AID. For example, Ramos is a Burkitt's lymphoma cell line that constitutively express AID and undergoes SHM (169), thus we were expecting a cellular environment prepared to respond differently to the AID-induced damage.

Before using cellular extracts in IVR assay, we checked that the extracts preparation protocol leads to the purification of active DNA repair proteins. We started the analysis by evaluating SSBR capacity of each extract (Figure 18). SC plasmids were treated with nicking enzymes and then used in the second phase of IVR. Repair activity detected in different extracts when plasmids were nicked was 4 to 7 fold higher (Figure 18 - turquoise bars) compared to not treated plasmids (set to 1). When CIP treatment was performed to reduce the number of self-ligations and thus to increase the number of damaged plasmids, we scored an increase in repair activity (Figure 18 - blue bars). On the contrary, when FE was used to repair a plasmid nicked and de-phosphorylated, a reduction was registered. This can be due to specific repair proteins turnover in different extracts.

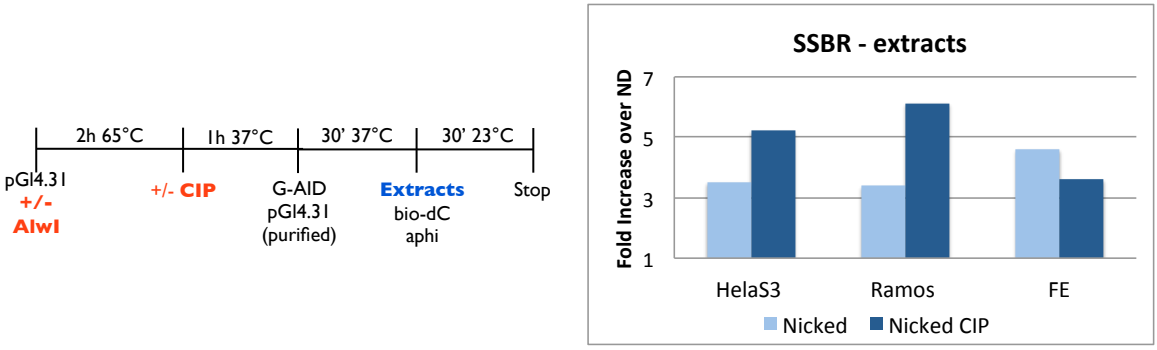


Figure 18 – Quantification of SSBR in different cellular extracts

IVR was performed using a modified substrate. Plasmids were either digested or not with AlwI nicking enzyme. After enzyme heat inactivation, plasmid was treated or not with CIP enzyme. Purified plasmids were then repaired by FE, Ramos and HelaS3 extracts (with bio-dC) in the presence of aphidicolin (aphi). Results from untreated plasmids (ND = not digested) were set to one. Time line of experiment is shown next to the graph.

After confirming that cell extracts are able to perform DNA repair, we evaluated extracts competency in the IVR assay (Figure 19).

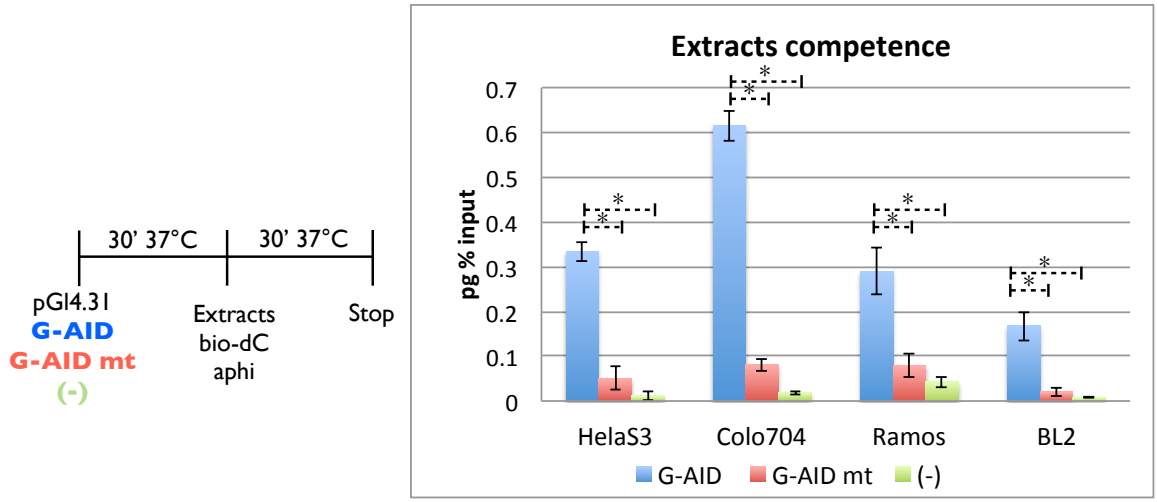


Figure 19 – Extracts competence for IVR assay

SC plasmids were treated with 0.66 pmols of G-AID, G-AID mt or (-) and then incubated with different cellular extracts. The blue bars represent the absolute recovery of each treated sample in relation to its input. Red bars represent the repair pathways activated by G-AID mt and green bars represent extracts-induced repair activity (without the addition of DNA damaging proteins). Error bars indicate SD (n = 3). Time line of the experiment is shown next to the graph. Statistical analysis (t-test) was performed on differences of indicated pg % input; * indicates p values < 0.05.

In the absence of G-AID the extracts were not providing nicking activity, as no DNA repair machinery was activated (Figure 19 - green bars). Different cell lines displayed significant repair activity only if G-AID was added to the SC plasmid, indicating that, similar to the FE,

G-AID catalytic activity is required to initiate DNA repair (Figure 19 - blue bars). As expected, different cell lines displayed different repair activity.

3.5 - ANALYSIS OF SINGLE REPAIR PATHWAYS ACTIVATED BY DIFFERENT CELL LINES UPON AID-INDUCED DAMAGE

To distinguish which repair pathways were responsible of differential AID-induced damage repair in different extracts, a more detailed analysis was performed with IVR assay.

In particular, dU lesions can be recognized by two pathways: BER and MMR. BER can be subdivided into SP-BER and LP-BER, specialized in replacing only uracil or up to 10 bases surrounding uracil, respectively. Furthermore, four different glycosylases can initiate BER: UNG and STM (SMUG1, TDG or MBD4). MMR can be classified into clMMR and ncMMR, if error-free or error-prone polymerases are recruited, respectively. Depending on the acting repair pathway different nucleotides can be incorporated, for example dA will be incorporated only if long patches of DNA are removed and replaced.

As stated, IVR was performed in the presence of aphidicolin, to inhibit replication-dependent biotin incorporation. However, aphidicolin is also inhibiting Pol δ/ε, which can be part of LP-BER and clMMR (170,171). LP-BER can also occur with Pol β, a polymerase involved in SP-BER (172). ncMMR is an alternative to clMMR and utilizes other error prone polymerases (53). The repair pathways utilizing bio-dC incorporation in the presence of aphidicolin are: SP-BER, LP-BER (Pol β dependent), and MMR (Pol η dependent). Although still in the presence of aphidicolin, detailed analysis of single DNA repair pathway usage was performed using different concentrations of cellular extracts (x axis - Figure 20): A) bio-dC incorporation (Figure 20A), B) bio-dC incorporation and extract treatment with UGI (Figure 20B) or 3) bio-dA incorporation (Figure 20C). Single repair pathways that are activated in each condition are indicated below each chart (Figure 20): A) bio-dC is incorporated by both BER and MMR, B) if UGI is used, bio-dC will be incorporated by MMR and BER-STM dependent and 3) bio-dA will be incorporated by MMR and LP-BER pathways.

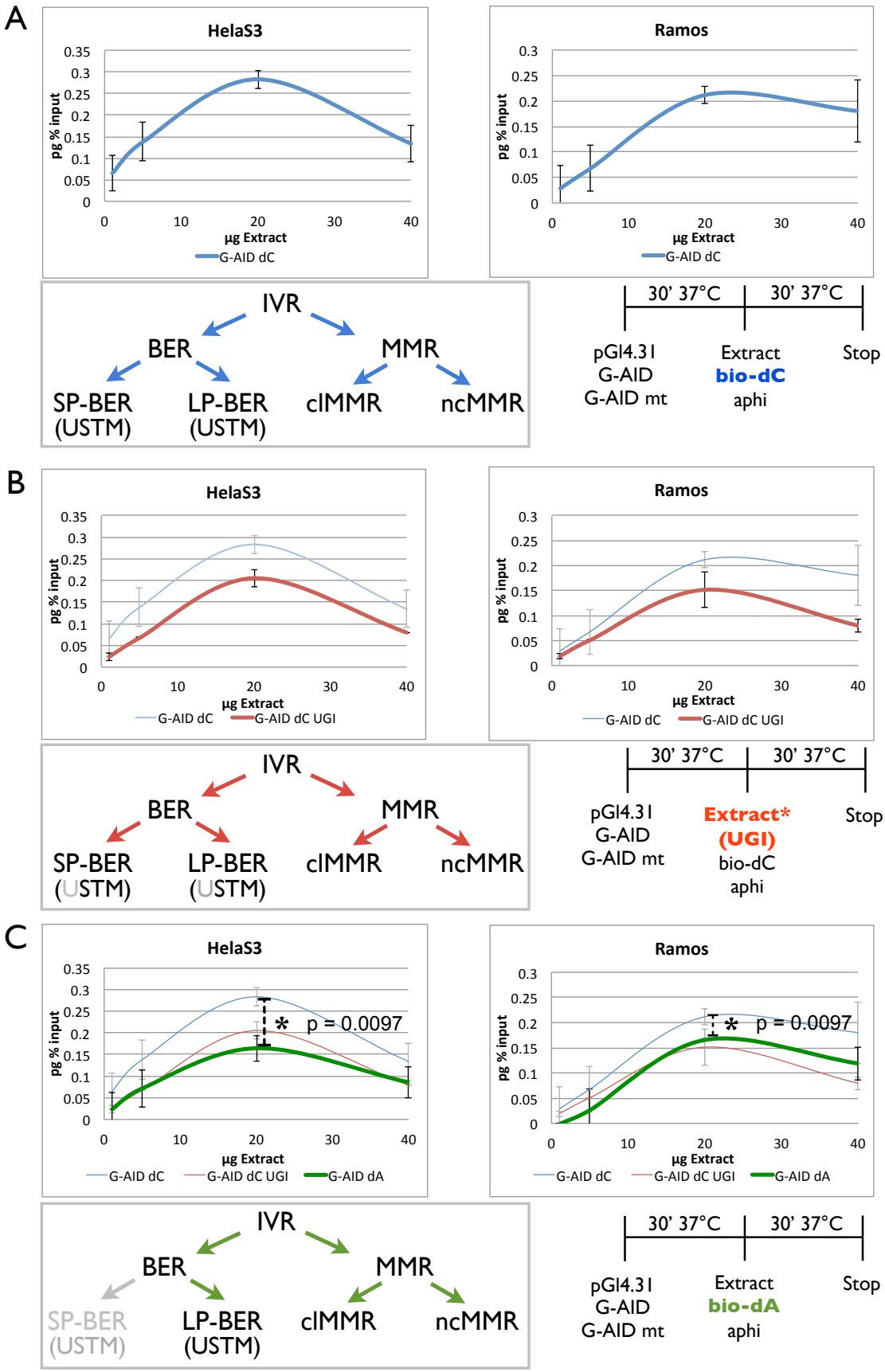


Figure 20 – IVR assay is suitable for studying AID-induced damage resolution in different cell lines

A) SC plasmids were treated with 0.66 pmols of G-AID WT or mt and then incubated with increasing amounts of HelaS3 and Ramos extracts (X-axis). Reaction was performed in the

presence of bio-dC. Deaminated and repaired plasmids were isolated and quantified as % of input (total amount of plasmids). G-AID mt results were subtracted from G-AID WT results. U indicates UNG glycosylase, while STM includes SMUG1, TDG or MBD4 glycosylases. Error bars indicate SD (n = 3). Time line of the experiment is shown below the graph. Pathways activated in this IVR assay are indicated in the panel below the chart. **B)** Increasing amounts of HelaS3 and Ramos extracts were incubated with 1 UGI unit. Reaction was performed in the presence of bio-dC. Pathways activated in this IVR assay are indicated in the panel below the chart. Error bars indicate SD (n = 3). Time line of the experiment is shown below the graph. **C)** SC plasmids were treated with 0.66 pmols of G-AID WT or mt and then incubated with increasing amounts of HelaS3 and Ramos extracts. Reaction was performed in the presence of bio-dA. Pathways able to incorporate bio-dA are indicated in the green table. Error bars indicate SD (n = 3). Time line of the experiment is shown below the graph. Pathways activated in this IVR assay are indicated in the panel below the chart. Statistical analysis (t-test) was performed on differences between A and C in Ramos and HelaS3. * p values are indicated in the figure.

HelaS3 extract was as efficient as Ramos extract in repairing AID-induced lesions (Figure 20A). In both extracts the maximum activity was reached at 20 µg. In the presence of UGI inhibitor, the repair activity was reduced (Figure 20B). In the Ramos extract, this inhibition was dependent on proteins concentration: hardly detectable at low concentrations, while more than half of the repair activity was inhibited at maximum concentration (40 µg – Figure 20B). In HelaS3 extract BER-UNG activity was independent from concentration. Furthermore, when bio-dA was incubated with the extracts, BER and MMR were more active in Ramos extract compared to HelaS3 extract (Figure 20C).

To better unravel the single repair pathways recruitment upon AID-induced lesion we decided to combine data from bio-dC and bio-dA incorporation with single or multiple specific inhibitors: UGI, T2AA and aphidicolin. UGI and aphidicolin were described before. T2AA is a small peptide able to inhibit PCNA interaction with proteins containing a PIP-box motif (161), including Pol δ and with low affinity also Pol ε and TLS polymerases. The only pathway unaffected by PCNA inhibition is SP-BER. Alternatively, some repair activity can still be observed if the ring-shaped heterotrimeric 9-1-1 complex (composed of Rad9, Rad1 and Hus1 proteins), a PCNA relative, is loaded at damaged sites, where it serves as platform for the selective recruitment of repair proteins. Although 9-1-1 clamp seems to play a major role in HR and template switching, it was shown to physically interact with

EXO1 (173) and TDG glycosylase (174); it is not clear how much and if it plays a role in BER and MMR of our extracts.

Titration experiments with UGI, aphidicolin, and T2AA identified optimal inhibition concentrations (data not shown). 1 unit of UGI, 0.2 mM T2AA and 5 µg of aphidicolin were used to inhibit their targets in 20 µg of cellular extracts (5-10 fold in excess respect to minimum inhibitory concentration).

The strategy used to identify single repair pathways involvement in AID-lesion repair can be derived from the Table in Figure 21.

A

	dNTPS	G-AID	Inhibitors	SP-BER (UNG)	SP-BER (STM) + 9-1-1 DEP	LP-BER (UNG)	LP-BER (STM)	clMMR	ncMMR	Replica-tion
1	Bio-dC	✓	Aphi	✓	✓	✓	✓		✓	
2	Bio-dC	✓	Aphi UGI		✓		✓		✓	
3	Bio-dC	✓	Aphi T2AA	✓	✓					
4	Bio-dC	✓	Aphi T2AA UGI		✓					
5	Bio-dC	✓	UGI		✓		✓	✓	✓	✓
6	Bio-dC	-	-							✓
7	Bio-dC	-	Aphi							
8	Bio-dA	✓	Aphi			✓	✓		✓	
9	Bio-dA	✓	Aphi UGI				✓		✓	

B

SP-BER (UNG)	3 - 4
SP-BER (STM) + 9-1-1 DEP	4
LP-BER (UNG)	8 - 9
ncMMR + LP-BER (STM)	9
clMMR	5 - 2 - 6

Figure 21 – Tabular representation of the strategy used to identify single repair pathways involvement in AID-induced damage repair

A) Repair pathways able to recognize and process dU lesions are indicated in the first row. Pathways unaffected by single inhibitor (or combinations of different inhibitors) are identified by the check mark. G-AID is the result of G-AID mt subtraction from G-AID WT. **B)** Calculations performed to derive single repair pathways involvement in AID-induced damage repair. Numbers refer to the first column in Figure A.

Subtracting results between the identified pathways we could calculate the single pathways involvement in repairing AID lesions (Figure 17B). For example, SP-BER (UNG-dependent) was calculated subtracting IVR results derived from incubation of the extracts with aphidicolin and T2AA from IVR results derived from the use of three inhibitors (aphidicolin, T2AA and UGI), in the presence of bio-dC. Thus, the contribution of 5 major

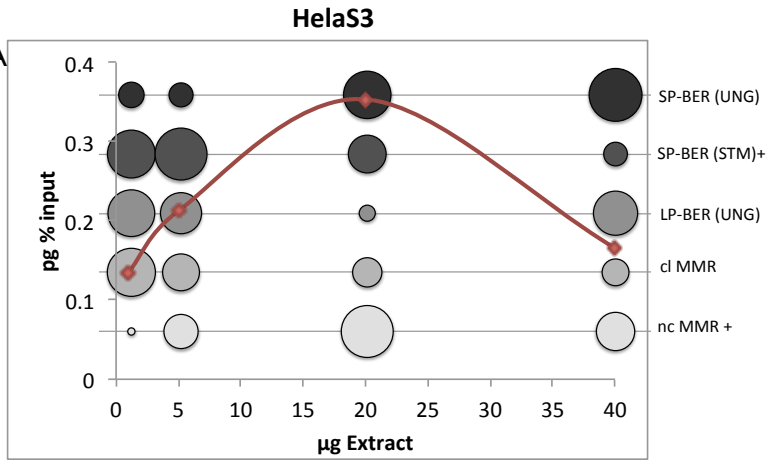
repair pathways to AID-lesion repair could be independently quantified: 1) SP-BER (UNG-dependent), 2) SP-BER (STM-dependent) and repair 9-1-1 dependent, 3) LP-BER (UNG-dependent), 4) ncMMR and LP-BER (STM-dependent) and 5) clMMR (Figure 21B).

This strategy was employed to charachterize 4 cellular extracts (HelaS3, Colo704, Ramos and BL2) (Figure 22) and the FE (Figure 23). The total repair activity (red line - Figure 22) and the contribution of single repair pathways to AID damage resolution (bubbles of different grey scale - Figure 22) were calculated using 4 different concentrations of extracts (x axis - Figure 22). All the cellular extracts analyzed displayed a maximum repair activity at 20 µg, while with higher concentrations an inhibitory effect was observed. Overall the following trends in single repair activities were observed: 1) ncMMR was predominantly activated at higher concentrations of extracts, 2) clMMR was more active in non-B cell extracts and at lower concentrations, 3) in B-cell extracts SP-BER activity was inversely proportional to the concentration of extract used, with strong activation at lower concentration (while ncMMR activity is predominant at higher concentrations), 4) LP-BER activity was variable depending on the concentrations of extracts used; in B cell extracts LP-BER was predominant when maximum repair activity was observed, while in the same conditions it was barely visible in non-B cell extracts.

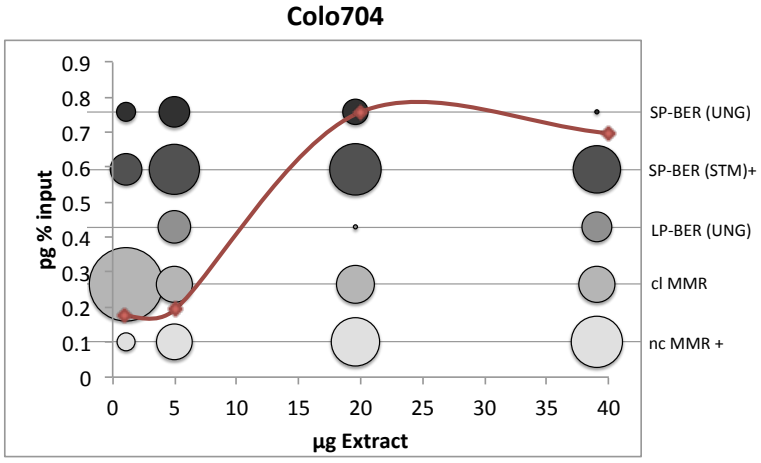
We conclude that the concentration of active repair proteins is fundamental for single repair pathway activation and could play a role in the inhibition or compensation exerted toward other repair pathways.

The same analysis was performed using FE and only at 150 µg the extract displayed the maximum repair activity, while no inhibitory effect was reported (Figure 23 - red line) even at higher concentrations (data not shown). This may be an indication of repair kinetics: substantially high number of active repair molecules may be required to trigger repair pathways activation. Comparably to B cell extracts, the FE displayed strong SP-BER activation at lower concentrations of extract, while the other repair pathways reached the same repair activity independently from the concentration used.

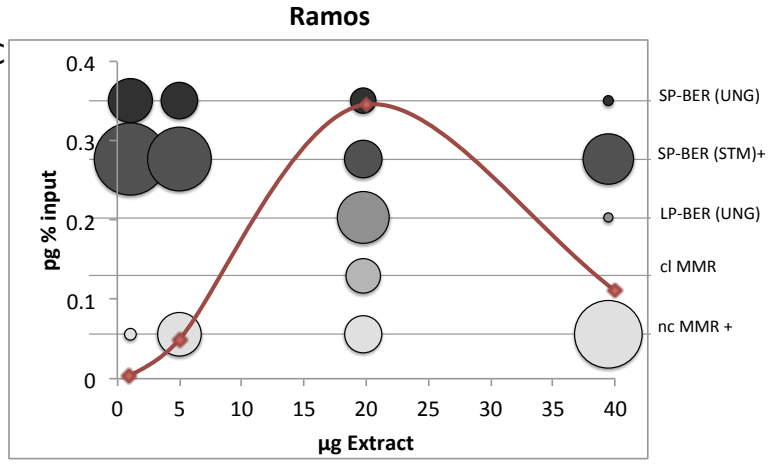
A


HelaS3
pg % input
µg Extract
SP-BER (UNG)
SP-BER (STM)+
LP-BER (UNG)
cl MMR
nc MMR +

B


Colo704
pg % input
µg Extract
SP-BER (UNG)
SP-BER (STM)+
LP-BER (UNG)
cl MMR
nc MMR +

C


Ramos
pg % input
µg Extract
SP-BER (UNG)
SP-BER (STM)+
LP-BER (UNG)
cl MMR
nc MMR +

D

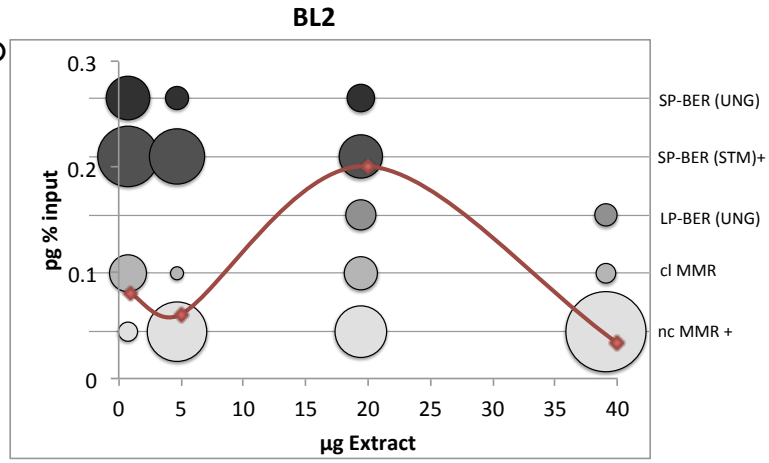

BL2
pg % input
µg Extract
SP-BER (UNG)
SP-BER (STM)+
LP-BER (UNG)
cl MMR
nc MMR +

Figure 22 – Quantification of single repair pathway activation upon AID damage in IVR assay using different cellular extracts

One, 5, 20 and 40 µg of cellular extracts (x axis) (**A** = HelaS3, **B** = Colo704, **C** = Ramos, **D** = BL2) were used. Total repair activity (red line) was calculated as absolute % of bio-dC incorporation in the absence of inhibitors, and subtracting the extract specific replication activity. For DNA repair pathway contribution each data point in the red line was set to 100 and repair pathways contribution was represented as bubbles (bubble diameter = %). Grey scale represents different repair pathways. Images are the average of 3 independent experiments.

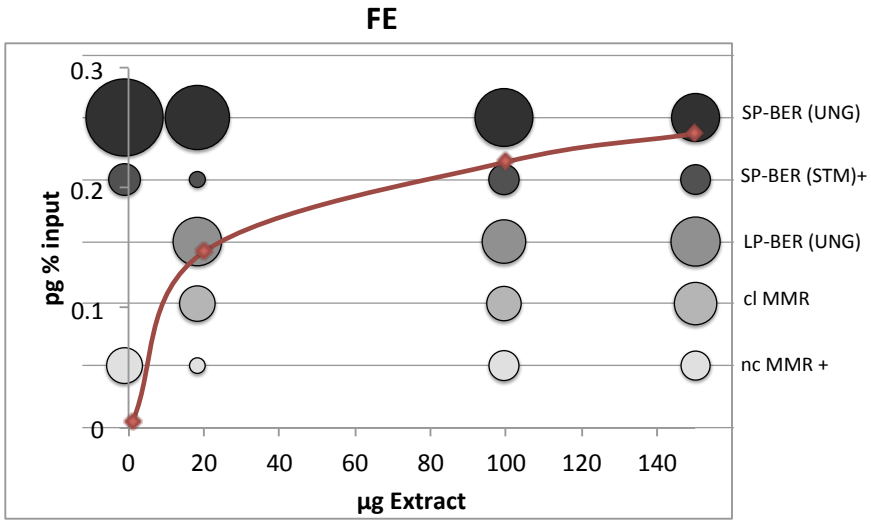


Figure 23 – Quantification of single repair pathway activation upon AID damage in IVR assay using FE

One, 20, 100 and 150 µg of FE (x axis) was used to repair *in vitro* AID-induced lesion. Total repair activity (red line) was calculated as absolute % of bio-dC incorporation in the absence of inhibitors, and subtracting the extract specific replication activity. For DNA repair pathway contribution each data point in the red line was set to 100 and repair pathways contribution was represented as bubbles (bubble diameter = %). Grey scale represents different repair pathways. Images are the average of 3 independent experiments.

To have a better overview of single repair pathways involvement in AID-induced damage repair, IVR results derived from maximum repair activity were represented in pie charts (Figure 24A). The cellular extracts displayed maximum repair activity at 20 µg while FE repair activity was complete only at 150 µg (Figure 24B). A non B-cell extract, Colo704, was 2 to 3 fold more efficient then other extracts in repairing AID lesion.

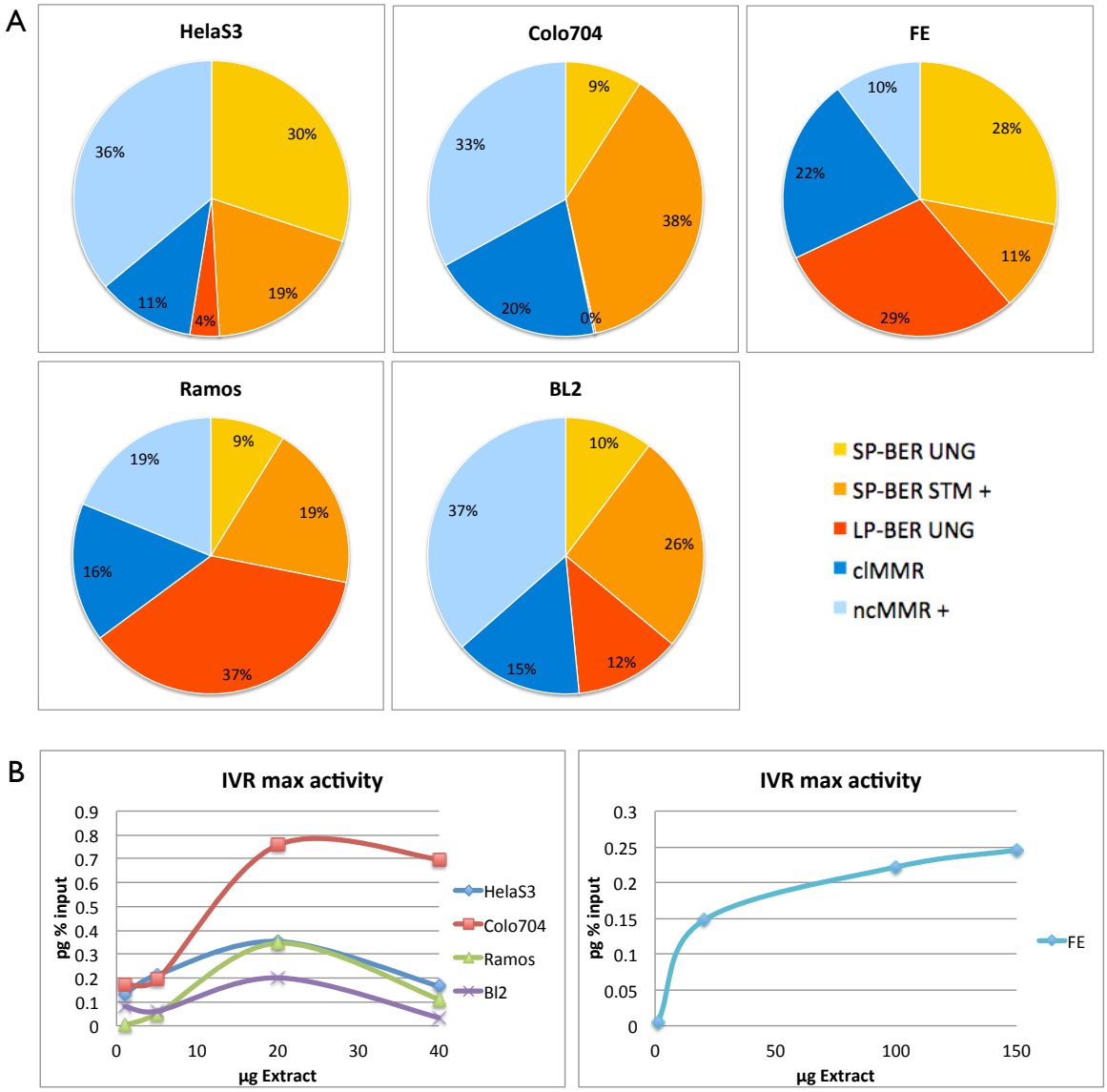


Figure 24 – Detailed analysis of single repair pathways contribution to maximum IVR activity

A) IVR results obtained with 20 µg of cellular extracts and 150 µg of FE were represented in pie charts. % of single repair pathways contribution to maximum repair activity are indicated in the pie charts. The orange portion of the pie charts correspond to SP-BER and 9-1-1 dependent repair, the turquoise portion correspond to ncMMR and LP-BER (STM-dependent). **B)** IVR total repair activity is represented as absolute recovery in relation to input, using different concentration of extracts or FE.

We observed that MMR (cool colours – Figure 24A) and BER (warm colours – Figure 24A) contributed to AID damage resolution. Comparing different B and non-B cell extracts, LP-BER activation was predominant in B cells extracts, while it was almost absent in non-B cell extracts. As observed in Figure 17A, UNG-dependent BER is the major repair pathway activated by FE to repair AID-induced lesion (yellow and red account for 57 % of

the total repair – Figure 24A). The pie chart analysis recapitulated the results observed in Figure 20C. Considering total repair activation, LP-BER and MMR activation is higher in Ramos extract (red and cool colours account for 72 % of the total repair – Ramos - Figure 20A) compared to HelaS3 extract (red and cool colours account for 51 % of the total repair – HelaS3 – Figure 20A). From Figure 20C we derived that using 20 µg of Ramos extract, 78 % of the plasmids incorporate bio-dA (LP-BER and MMR); while using 20 µg of HelaS3 extract, 57 % of the plasmids incorporate bio-dA. This is an indication of the consistency of the IVR for determining single repair pathways activity. Minor differences observed between the two experiments (Figure 20C and Figure 24A) could be technically or due to crosstalks and compensatory mechanisms of the DNA repair.

Interestingly, in Ramos, ncMMR, an error prone repair mechanism, is lower compared to other extracts. As B cells needs introduction of mutation we were expecting an opposite trend, with more error-prone repair pathway activation. This could be due to Ramos cell background, as SHM and AID are constitutively active, and thus a specific response to continuous mutation stimulation could have evolved. Alternatively, mutation incorporation can also be performed with other repair pathways, for example during LP-BER activation, if translesion polymerases are recruited.

Ramos and BL2 extracts displayed some similarities (for example higher LP-BER activation compared to non-B cell extracts) and some differences (for example in MMR activation). It would be then interesting to induce endogenous AID expression in BL2 to see if the repair profile became more similar to Ramos profile.

We concluded that different cell lines and FE quantitatively activate different repair pathways to repair AID lesion. Other processes can interfere with the dU repair activity, such as chromatin conformation, specific signal transduction, protein interacting with repair pathways, rate of mutations, cell replication state and many others.

3.6 - MMR AND BER PROTEINS ARE DIFFERENTIALLY EXPRESSED IN B CELL EXTRACTS AND NON-B CELL EXTRACTS

To understand why different cell lines displayed different activity after encountering AID-induced damage, we performed a more detailed analysis. First, we checked the amount of some of the DNA repair protein present in each extract (Figure 25A). Surprisingly, EXO1, a key protein of MMR, was more abundant in non B-cell extracts, while UNG and MBD4 glycosylases that initiate BER were more abundant in B cell extracts (Figure 25B). FE was mostly enriched in UNG, while MBD4 did not seem to contribute to BER in FE.

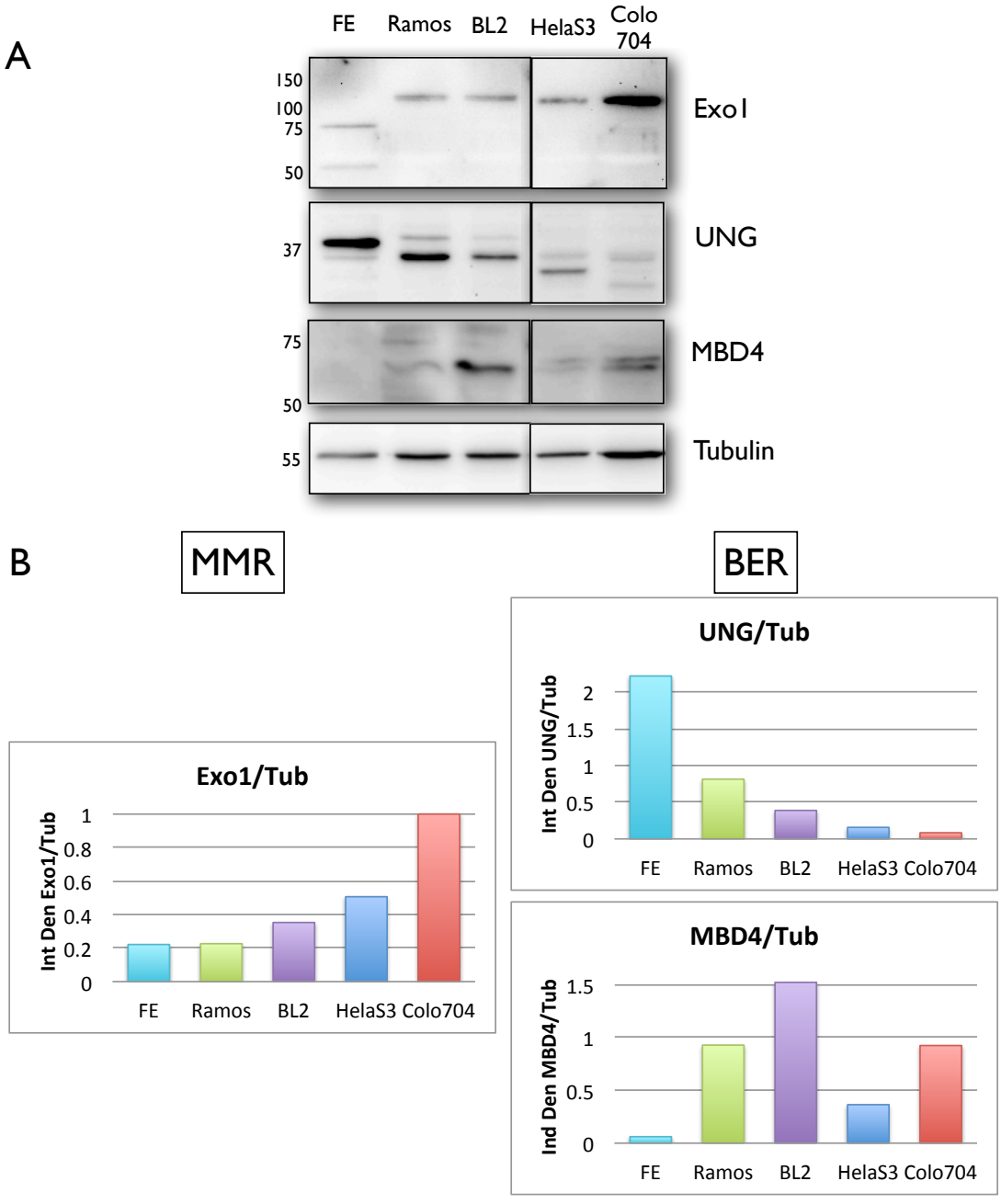


Figure 25 - BER and MMR protein expression in different cell lines

A) Western Blot analysis using different antibodies was performed using 20 µg of FE, Ramos, BL2, HelaS3 and Colo704 extracts. Differences in bands position between human and xenopus proteins are due to different molecular weight, for example Xenopus EXO1 is known to be 77 kDa compared to the human EXO1 (95 kDa). Multiple bands are possibly due to the presence of different human isoforms: 5 in UNG and 6 in MBD4. **B)** EXO1, UNG and MBD4 were normalized to tubulin in each extract. Values are expressed as integrated density. Image-J program was used to calculate bands relative area and density. Further calculations were performed to remove background specific for each single band.

Although we identified differences in repair protein expression, there was no evidence for a direct correlation with IVR activity. A more precise analysis should consider also the catalytic activity of these enzymes.

3.7 - ANALYSIS OF BER PROTEINS ACTIVITY IN DIFFERENT CELL EXTRACTS

To correlate repair protein amount with IVR resolution, we checked if repair protein amount is directly related to repair protein activity. Using a quantitative dsDNA oligonucleotide DNA glycosylases assay, we determined the activity of various DNA glycosylases enzyme able to recognize uracil (UNG, SMUG, TDG or MBD4) and initiate SP-BER (Figure 26A). ds oligonucleotides with a single U:G mismatch were incubated with cell extracts to generate UNG-dependent abasic sites, processed by APE and NaOH to cleave at the abasic site. Depending on DNA glycosylases activity, oligonucleotides will be cleaved and will run at a different position; measurement of the oligonucleotides cleavage is therefore a direct measurement of DNA glycosylases activity. Specificity of the assay was determined by incubating ss or ds oligonucleotides with commercial UNG or ENDO III enzymes (Figure 26B). UNG alone was able to cleave only a small portion of the oligonucleotides, ENDO III alone was not active in cleaving the DNA backbone if the uracil base was still present, while combining both enzymes activities we could achieve almost 100 % of cleavage, using both ss and ds oligonucleotides.

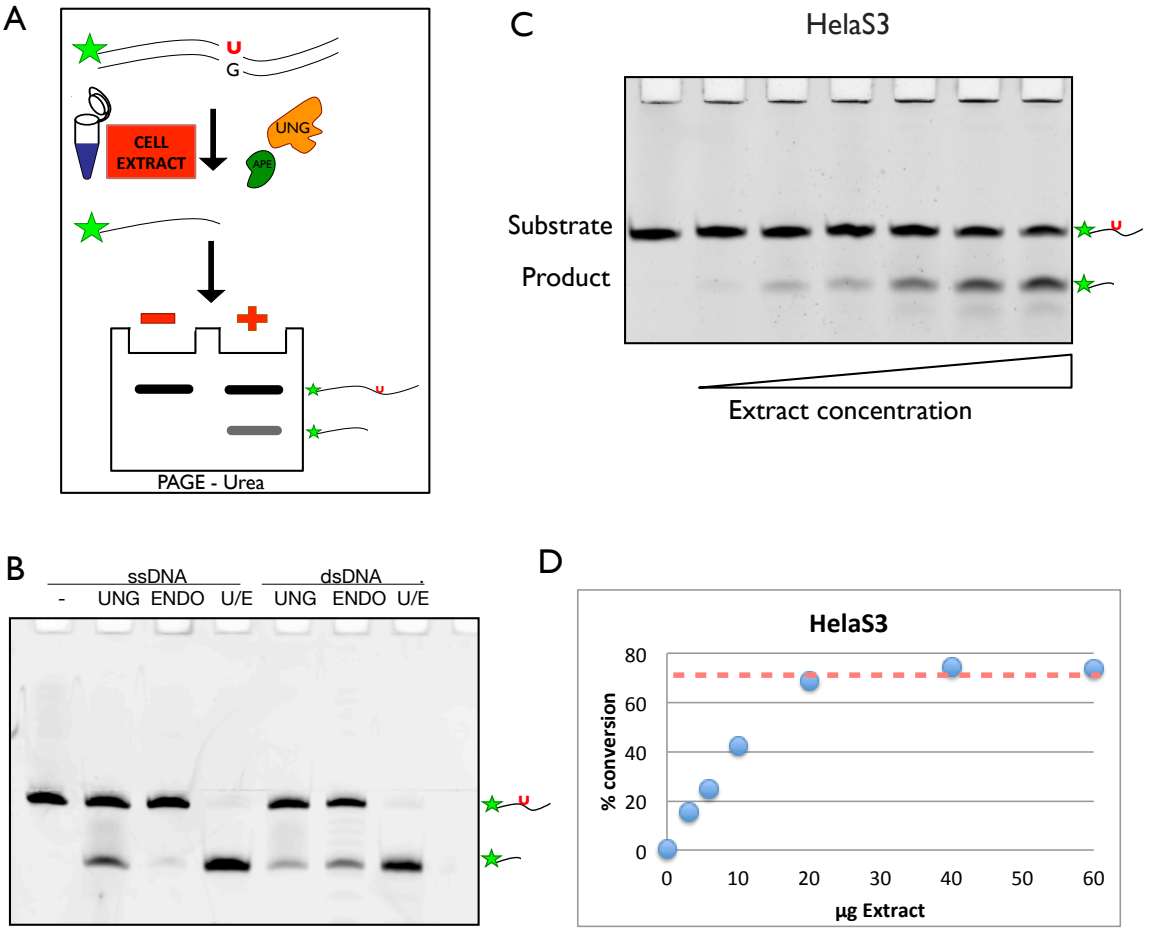


Figure 26 - ds oligomucleotide DNA glycosylases assay is suitable to evaluate DNA glycosylases activity in different cellular extracts

A) Schematic representation of ds oligonucleotide DNA glycosylases assay. 2.5 pmols of ds oligonucleotides with single U:G mismatches were incubated 15 min at 37°C with cell extracts, which contained DNA glycosylases enzymes (UNG, SMUG1, TDG and MBD4). Cleavage of the oligonucleotide was enhanced with NaOH addition. Cleaved oligonucleotides run faster on the 17.5 % PAGE gel. **B)** Commercial purified enzymes (UNG and ENDO III) were incubated 15 min at 37°C with 2.5 pmols of ss or ds oligonucleotides. Combination of the two enzyme lead to a full cleavage of the U:G oligonucleotide (lower band). **C)** different concentration of HelaS3 extract (0 to 60 µg) were incubated 15 min at 37°C with 2.5 pmols of ds oligonucleotides and run on the 17.5 % UREA PAGE gel. DNA glycosylases present in HelaS3 extract were able to cleave the U:G oligonucleotide; DNA glycosylases activity increased with increasing amount of extract **D)** Activity of DNA glycosylases in increasing concentrations of HelaS3 extract was calculated as % of oligo conversion. Red dotted line derived maximum activity from plateau. ImageJ program was used to calculate bands relative area and density. Further calculations were performed to remove background specific for each single band. % of conversion was calculated as: intensity lower band*100/[intensity upper band + intensity lower band].

We tested our cellular extracts competency by incubating increasing amount of extracts with a fixed amount of oligonucleotides (Figure 26C). There was a linear correlation between DNA glycosylases activity and global protein content, until a plateau phase was reached (Figure 26D). A threshold was set to identify maximum % conversion (Figure 26D – red line).

3.7.1 - UNG activity is dominant over the other DNA glycosylases activities

Comparing DNA glycosylases activity in HelaS3 extract and Ramos extract, we observed some differences (Figure 27A). In HelaS3 a more sensitive response was observed, with few µg of extract containing already enough DNA glycosylases activity to process 50 % of the substrate. Maximum activity is achieved using 20 µg of extract and corresponds to 65 % of conversion, before a plateau is observed. In Ramos extract a slower activation profile was registered, with maximum activity (70 % of conversion) reached at 30 µg.

Using UGI we separated UNG activity from STM activities (Figure 27B). Increasing concentrations of UGI were incubated with the extract concentration that corresponded to maximum repair activity (20 µg for HelaS3 and 30 µg for Ramos – red squares in Figure 23A). 0.05 units of UGI were sufficient to reduce UNG activity. As expected, UNG activity was dominant over the other DNA glycosylases, as the remaining activity (STM dependent) was below 10 % in both HelaS3 and Ramos extracts.

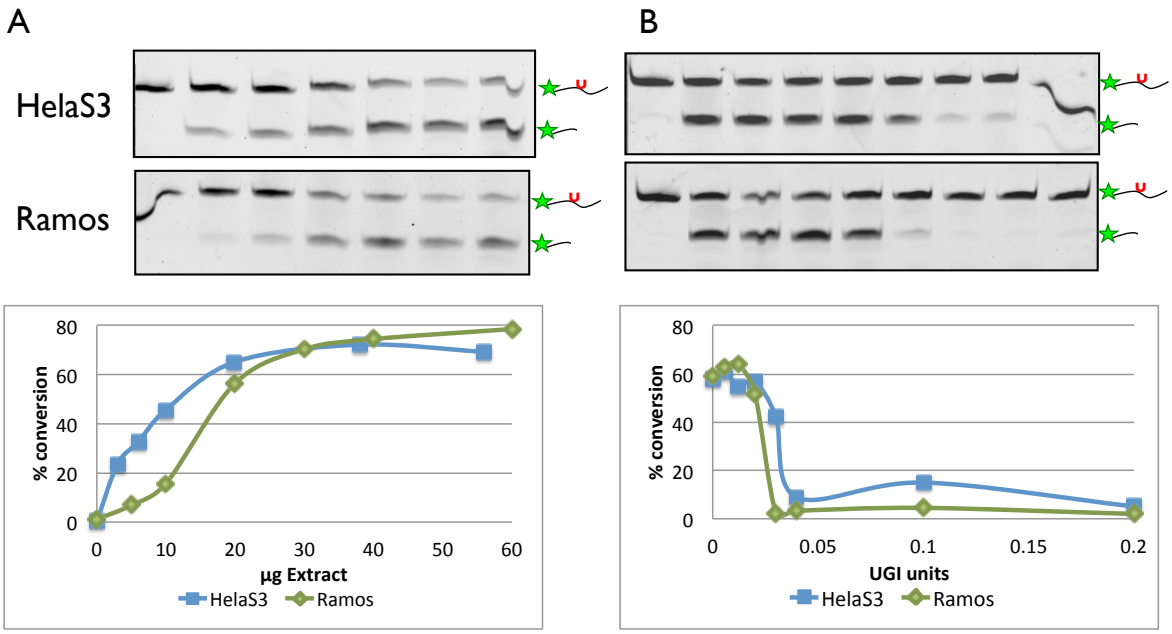


Figure 27 - UGI decreases DNA glycosylases activity in ds oligonucleotide DNA glycosylases assay

A) Increasing concentrations of HelaS3 and Ramos extract were incubated 15 min at 37°C with ds oligonucleotides and run on 17.5 % PAGE gel. DNA glycosylases activity increased with increasing amount of extract. **B)** Increasing units of UGI inhibitor were incubated on ice 10' min with 20 and 30 µg of HelaS3 and Ramos extracts, respectively. The extract was subsequently used for a ds oligonucleotide DNA glycosylases assay, as previously described. ImageJ program was used to calculate bands relative area and density. Further calculations were performed to remove background specific for each single band. % of conversion was calculated as: intensity lower band*100/[intensity upper band + intensity lower band].

Interestingly, when UGI titration was performed using a concentration of extract that was not corresponding to the maximum DNA glycosylases activity (i.e. 10 µg) we could not observe a linear decrease in DNA glycosylases activity (Figure 28) This was seen in the Colo704 extract, but similar results were also obtained with other extracts (data not shown). As UGI inhibits only active UNG molecules, this phenomenon could be explained by a concentration dependent activation of UNG molecules and possibly dimer formation.

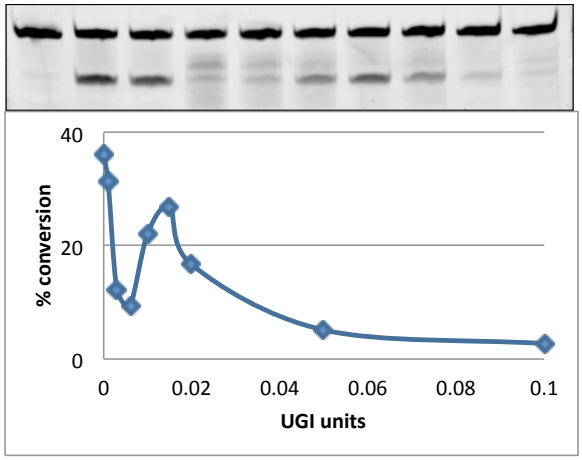


Figure 28 – UGI inhibition on not completely active UNG molecules from Colo704 extract
Increasing units of UGI inhibitor were incubated on ice 10' min with 10 µg of Colo704 extract. The extract was subsequently used for a ds oligonucleotide DNA glycosylases assay, as previously described. ImageJ program was used to calculate bands relative area and density. Further calculations were performed to remove background specific for each single band. % of conversion was calculated as: intensity lower band*100/[intensity upper band + intensity lower band].

3.7.2 - Number of active DNA glycosylases molecules does not directly correlate with BER activity

Contribution of different DNA glycosylases to uracil excision from DNA was quantified in different extracts (measured in pmol of oligonucleotides converted in 1 min by 1 µg of extract from the ds oligonucleotide DNA glycosylases assay) and in all the extracts UNG activity was dominant over the other UNG glycosylases. SMUG, TDG, and MBD4 had only a minor role in recognising uracil in our extracts (Figure 29A). In B-cell lines, a higher UNG activity per µg of extract was observed, compared to non-B cell lines.

As UGI is interacting 1:1 in molar stoichiometry with UNG (175),one can extrapolate the moles of active UNG enzyme for each extract (Figure 29B), while the Western blot (Figure 25) identified the total number of UNG molecules. With FE contained the highest number of UNG molecules, up to 6 times more than the other cellular extracts. Comparing the number of active UNG molecules with UNG activity, we determined UNG turnover for each extract (Figure 29C). Surprisingly, FE proteins had the lowest turnover, meaning that a high number of active molecules of UNG from FE are necessary to give small activity.

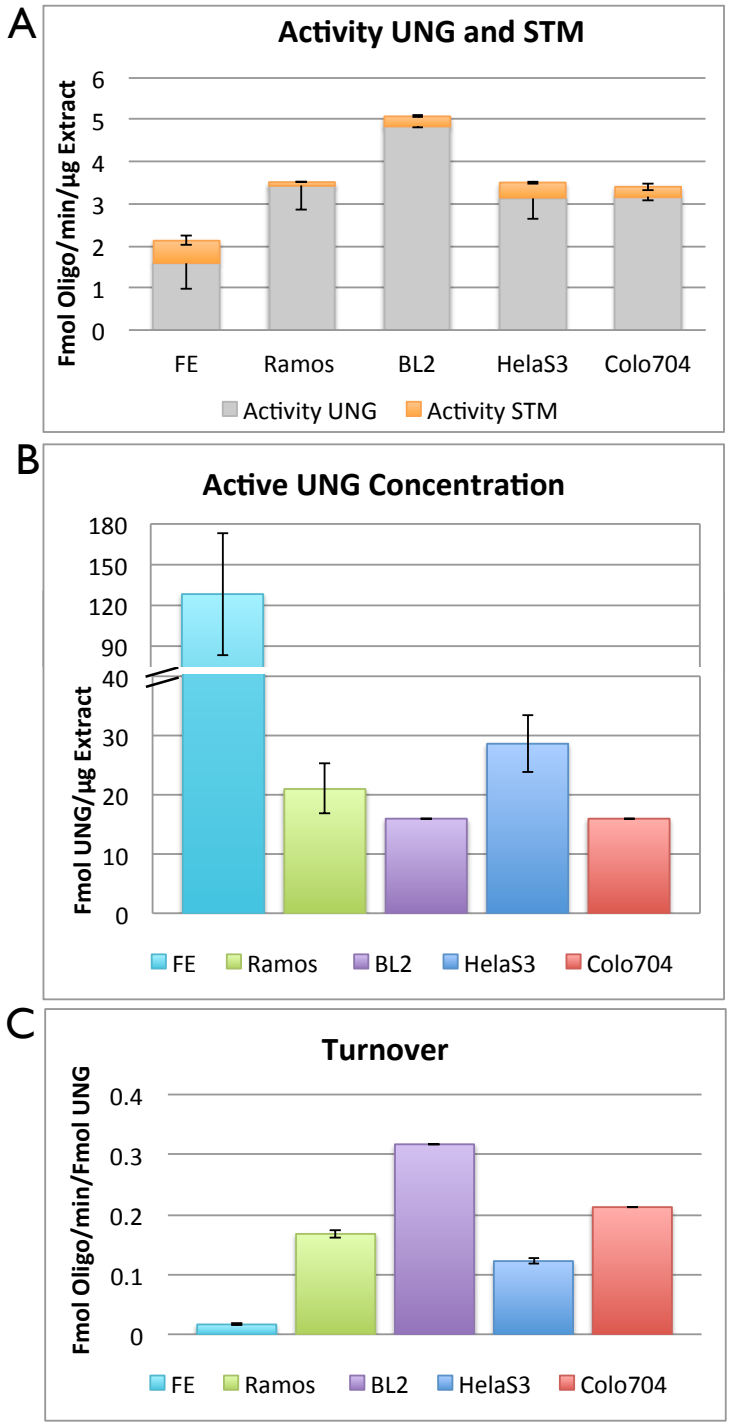


Figure 29 - Quantification of DNA glycosylases activity, concentration and turnover in different cell extracts

A) The assay was performed as previously described. Ratio of cleaved oligonucleotides is based on to 2.5 pmol of oligonucleotide input used and calculated as fmol of oligonucleotide cleaved in 1 min by 1 µg of extract. UNG is separated from other UNG enzymes using UNG2 specific inhibitor (UGI); remaining activity was due to STM (SMUG, TDG and MBD4 activity- orange colour) and was related to maximum DNA glycosylases activity. **B)** Active UNG concentration was derived from UGI (24 pmol/ 2 units) 1:1 interaction with UNG molecules. We derived the fmol of active UNG present in 1 µg of extract. **C)** Turnover was calculated from the ratio between activity of UNG and number of active UNG molecules. Error bars indicate SD (n = 3).

We conclude that in FE, UNG is slow in the recognition and removal of uracil, while few molecule of active UNG in BL2 extract are enough to give a fast and high activity in removing uracil. Using BL2 and Colo704 extracts a higher BER activation upon AID-induced damage was observed, contrary to FE.

Specific subpathways active in FE could be responsible for its kinetics of repair; SMUG, TDG and MBD4 provide 32 % of the total uracil excision (Figure 29A – orange column), and this activity may in turn be related to slow activation. Interestingly, the slow UNG turnover displayed by FE, could be correlated to IVR activity: 150 µg of FE were needed to reach the maximum repair activity upon AID-damage (Figure 23 and 24B), while in all the other extracts the repair activity was already saturated after 20 µg.

SP-BER activation analysis revealed that UNG activity is predominant over SMUG, TDG, and MBD4 (STM) activation in repairing AID lesions (Figure 29A). If we compare these observations with IVR results (Figure 24A), we can appreciate some differences in SP-BER glycosylases contribution to AID-damage resolution. We decided to integrate IVR results (Figure 24A) with specific contribution of DNA glycosylases to SP-BER (Figure 29A), to generate a more detailed analysis (Figure 30).

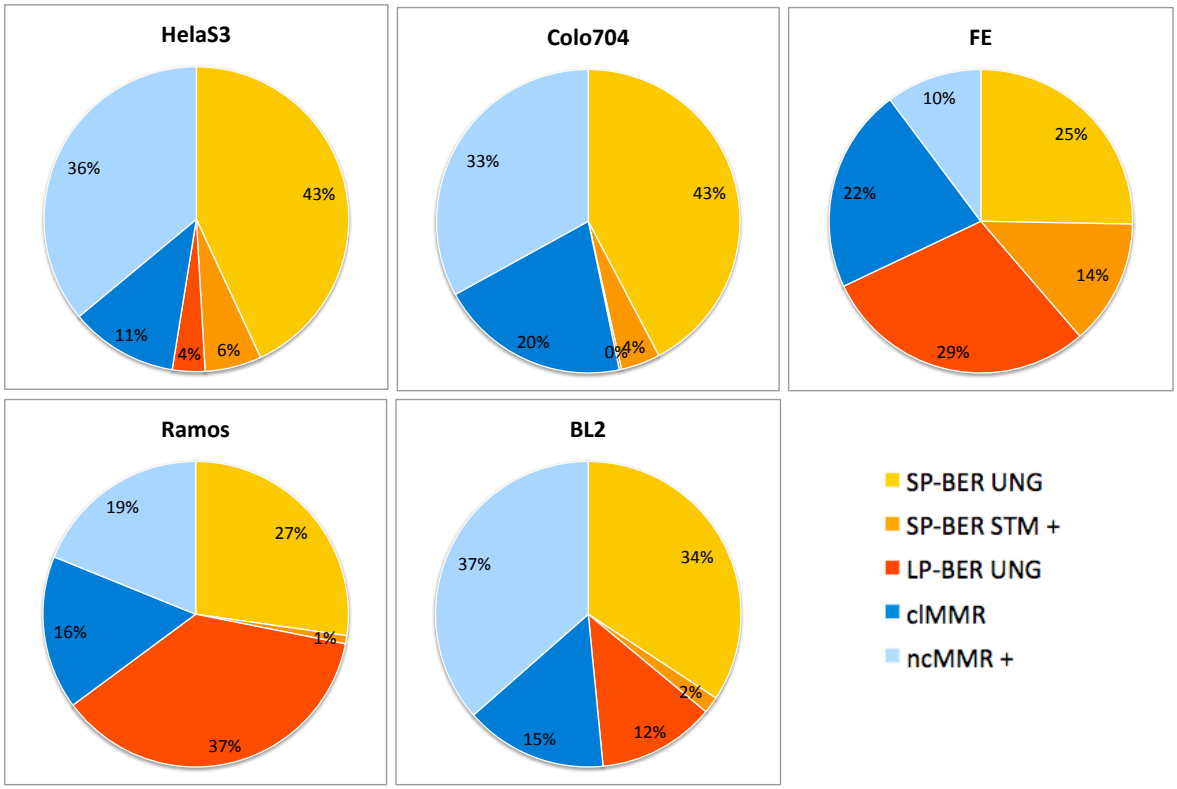


Figure 30 - Integrated analysis of single repair pathways contribution to maximum IVR activity

IVR results presented in Figure 24 were integrated with results obtained from ds oligonucleotide DNA glycosylases assay. Relative activities of SP-BER UNG and STM dependent calculated in Figure 24A were substituted to SP-BER (UNG) and SP-BER (STM) activities calculated with IVR.

Differences between Figure 24A and Figure 30 could be explained as follow: when we performed IVR analysis we were not able to separate SP-BER STM-dependent repair from 9-1-1 dependent repair. Moreover, SP-BER IVR analysis was performed using chemical inhibitors, thus compensatory mechanisms derived from chemical inhibition of specific pathways could have influenced the analysis.

Surprisingly, IVR analysis performed with FE was not visibly changed when integrated with uracil assay analysis (compare FE in Figure 24A and in Figure 30), meaning that no compensatory mechanisms occurred when UNG was chemically inhibited during IVR assay. Interestingly, FE was the only extract displaying a significant STM glycolylases activation (14 % - orange portion in FE Figure 30), indicating that UNG inhibition did not lead to an increase in STM activation to repair AID lesion.

On the contrary, in the cellular extracts IVR analysis revealed that STM-induced SP-BER activation was predominant over UNG contribution (Figure 24A). Integrating analysis with

dsDNA oligonucleotide DNA glycosylases assay we observed the opposite trend, with predominant UNG activation (Figure 30). This indicated that in cellular extracts, when UNG is inhibited a compensatory mechanism is activated, with an increase in SMUG, TDG and MBD4 activation.

Compensatory mechanisms are observed only in IVR assay and not in dsDNA oligonucleotide DNA glycosylases assay because a small oligonucleotide does not allow all the repair pathways to be active and to establish crosstalks. This hypothesis included the possibility that other proteins, activate only during IVR, are interfering and activating STM.

3.8 - TRASCRIPTIONAL ACTIVITY DOES NOT SEEM TO INFLUENCE IVR READOUT

We also checked the global transcriptional activity of different cell extracts, since both DNA repair and AID activity can be coupled to transcription (51,126,127). In the standard transcriptional assay a known DNA was incubated with the cell extracts, RNA was isolated, retrotranscribed and quantified (Figure 31). In FE, a minimal transcriptional activity was reported, while Ramos and another B cell line (DT40) produced extracts with the highest competency for transcription. The two epithelial cell lines displayed differential behaviour, with Colo704 being 5 fold more active in transcription compared to HelaS3. The FE's low transcriptional competence may be related to slow activation in recognition of the damage and activation of repair, or may be related to Xenopus oocytes peculiar maturation, via 12 rounds of DNA replication and cell division, without mRNA formation (176)

Although we identified differences in transcriptional activity, there was no evidence for a direct correlation with IVR activity.

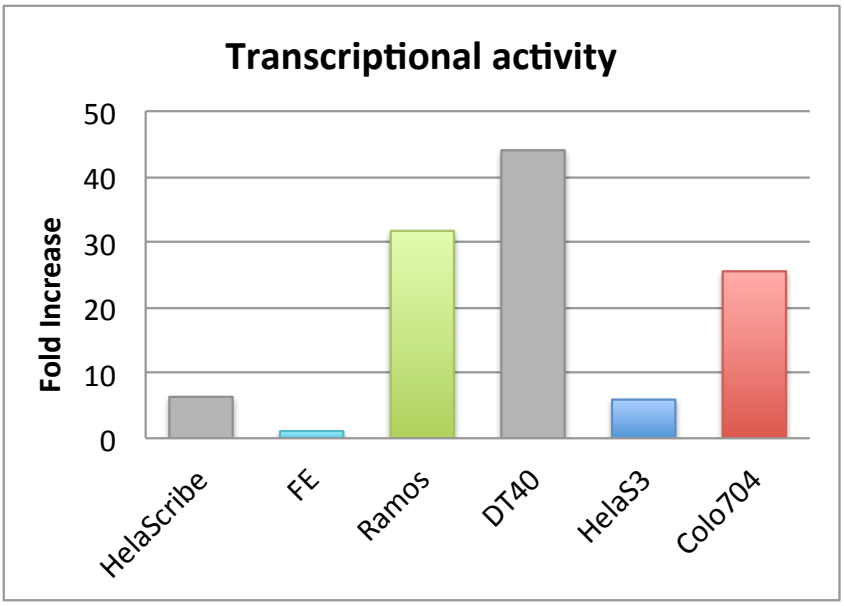


Figure 31 - Transcriptional competence of extracts

Transcriptional activity of each extract was tested with HelaScribe Nuclear Extract *in vitro* Transcription System from Promega; RNA formation was monitored by RT-qPCR. Values are normalized over negative control (extracts incubated with no HelaScribe DNA addition) and FE was set to 1. Except for FE, all extract are transcriptionally competent.

3.9 - CELL CYCLE STATE CONTRIBUTION TO DNA REPAIR PROTEIN AVAILABILITY IN THE RESOLUTION OF AID-INDUCED DAMAGE

Differential repair protein availability in the extracts could be due to proteome dynamics during cell cycle progression, and thus extract preparation could influence IVR repair activity. We synchronised cells in S/G2 phase with nocodazole to determine if the extract expression profile could alter the IVR readout. For each cell type we performed a titration analysis of nocodazole, measuring efficacy (growing curve of Ramos cells - Figure 32A) and toxicity (tryphan blue staining of dead Ramos cells - Figure 32B). Among the different concentrations of nocodazole used, 5, 10 and 20 ng per mL of culture were still not effective in blocking cell proliferation (Figure 32A – dotted lines). For synchronization we selected a concentration that provided an effective cell cycle block and with low toxicity (40 ng), and cells were incubated 14 to 16 hours before harvesting.

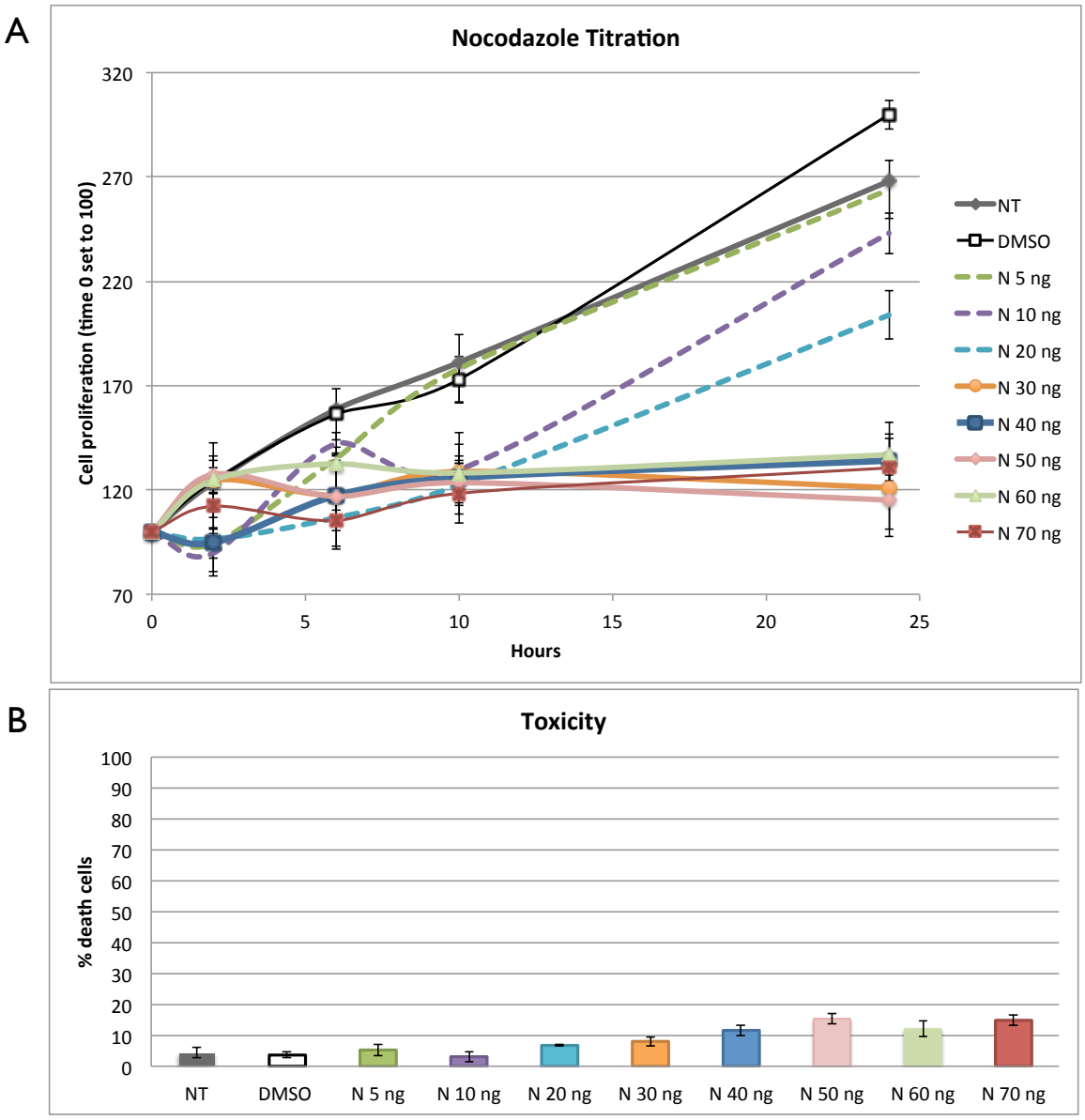


Figure 32 - Analysis of nocodazole efficacy and toxicity

A) Different concentrations of nocodazole (0 - 70 ng of nocodazole/mL of culture) were incubated 24 hours with Ramos cells (or HelaS3 cells, not shown). Cells were counted at different time points using tryphan blue. Concentrations reported by the dotted lines are not efficient in blocking cell cycle progression. **B)** Toxicity experiment was performed in parallel with A), % of death cells is reported after 24 h of nocodazole treatment, cells were counted with tryphan blue solution and hemocytometer.

Efficacy and toxicity were further tested with propidium iodide (PI) and anti-caspase3 staining during FACS analysis. Apoptotic cells accounted for 10 % of total (data not shown) and were removed prior to extract preparation. Almost 80 % of HelaS3 cells and 84 % of Ramos cells were trapped in S/G2 phase (blue and second red peak in the synchronisation model – Figure 33 A and 33C).

Extracts were tested in the IVR, and preliminary results indicate that there is no difference in IVR activity if HelaS3 cells were synchronized or not (Figure 33B), while for Ramos cells we observed some differences (Figure 33D), but further work is needed to determine statistical significance and single DNA repair pathway usage.

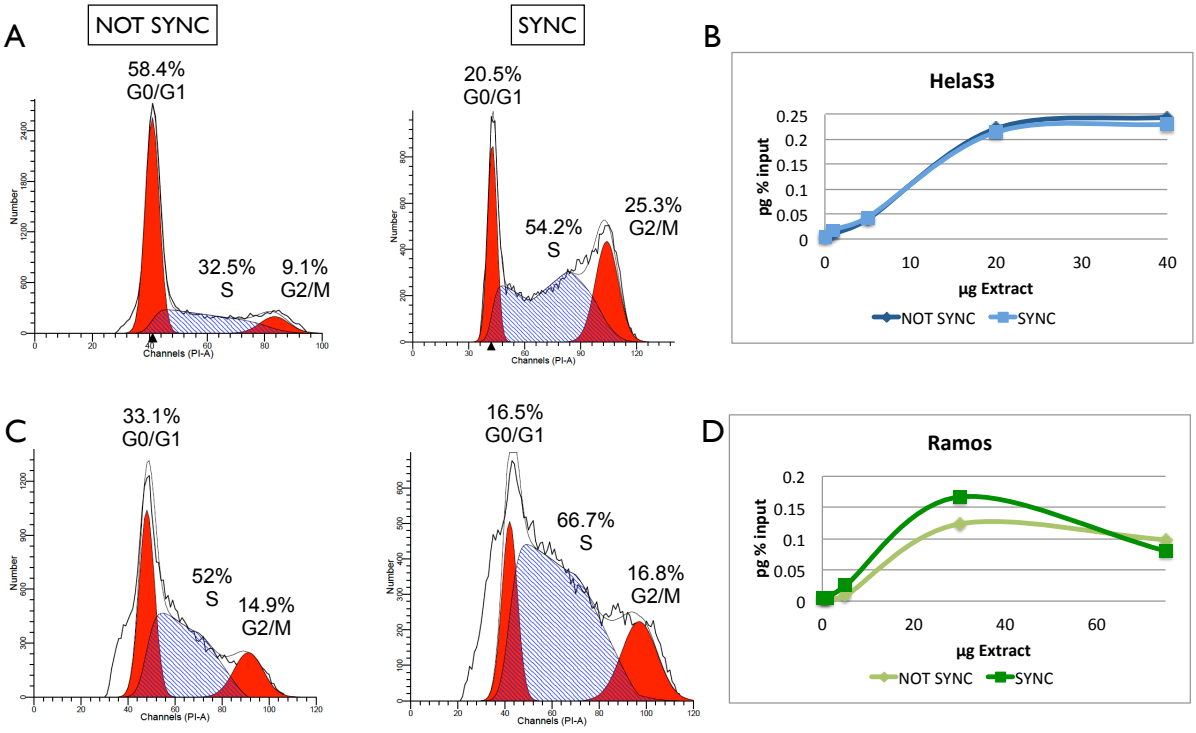


Figure 33 - Analysis of HelaS3 and Ramos synchronisation

A) HelaS3 cells were were synchronized with 40 ng/mL nocodazole for 16h, while. Cell cycle status was checked by FACS analysis after PI staining. Results were analysed with the synchronization wizard model. DNA content is indicated on the X-axis, while number of cells is reported on the Y-axis. First red peak is G1 phase, blue region is S phase and second red peak is G2 phase. % of relative phases are reported. After nocodazole treatment most of the cells were in S/G2 phase. **B)** IVR was performed using different concentrations of HelaS3 extracts synchronized or not. G-AID mt results were subtracted from G-AID WT results. **C)** Ramos cells were synchronized with 40 ng/mL nocodazole for 14h and analysis was performed as described in A. **D)** IVR was performed using different concentrations of Ramos extracts synchronized or not. G-AID mt results were subtracted from G-AID WT results.

3.10 - B CELL EXTRACTS ARE ENRICHED IN HISTONES COMPARE TO NON-B CELL EXTRACTS

Besides studying contribution of cellular microenvironment, transcription, and cell cycle to AID damage resolution, we also dissected the effect of nucleosomes and nucleosome

formation for our DNA damage response. Chromatin *status* could interfere both with AID deamination reaction and with DNA repair reaction. It is known that AID can target DNA sequences wrapped around nucleosomes only during transcription, which could have displaced nucleosomes (177). While chromatin compaction can also impede access of DNA repair protein to the damaged sites (178). We wanted to determine if the various extracts were able to form nucleosomes around the supercoiled plasmids used in the IVR assay, and if nucleosomes and other epigenetic characters influenced the IVR reaction. To this end, we analysed the histones abundance in different extracts by Western Blot analysis (Figure 34A). Normalizing H3 fraction over the total cytosolic fraction (tubulin) we observed that, B-cell extracts are more enriched in H3 while non-B cell extracts have up to 50 times less H3 protein (Figure 34B). Using H3 protein as standard we could also quantify the ng of H3 contained in 1 µg of Extract: Ramos and BL2 contained more than 12 ng of H3 per µg of extract (Figure 34B).

Low abundance of histones in Colo704 extract could be related to high IVR repair activity, as one can hypothesize that this DNA is more accessible to repair proteins. However, in HelaS3 extract we did not observed the same behaviour: even in the presence of few histones its repair activity is comparable to the one exerted by Ramos and BL2 extracts. We conclude that histones abundance is not directly related to IVR activity.

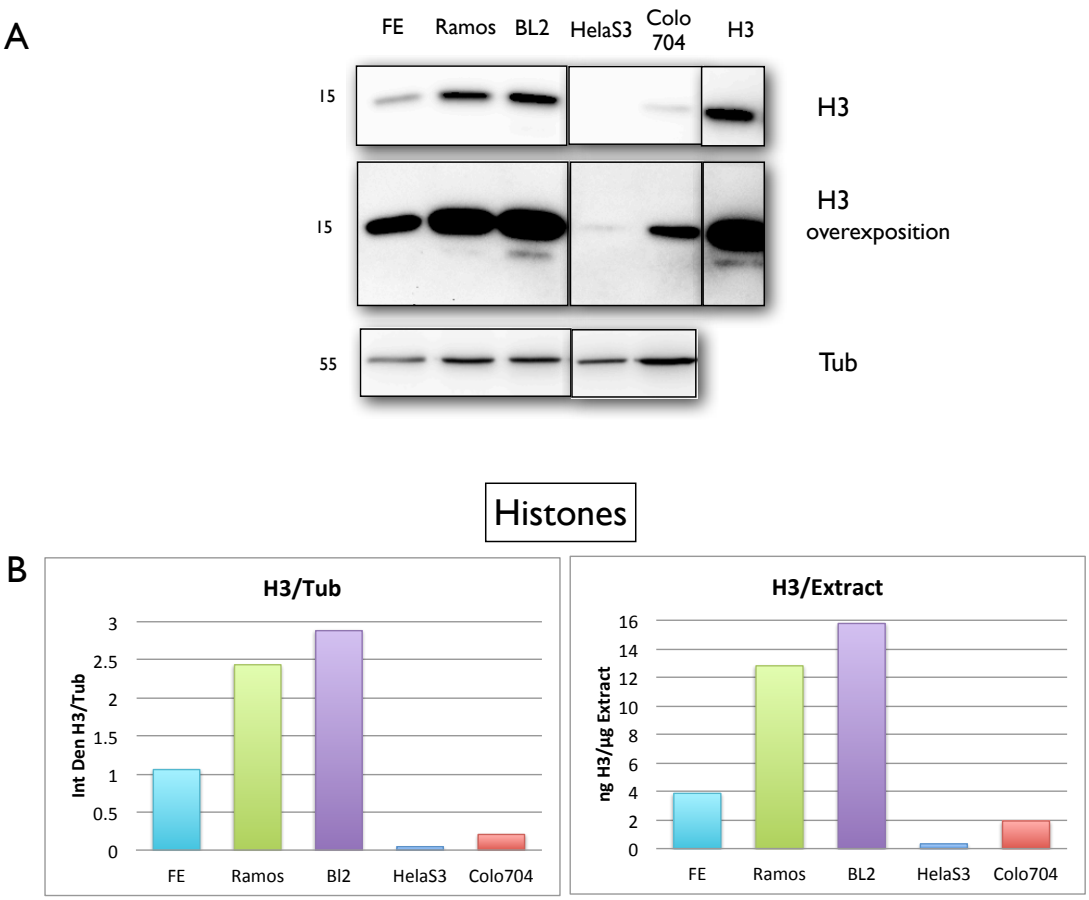


Figure 34 - Analysis of histones expression in different cell lines

A) Western Blot analysis using anti H3 antibody was performed in FE, Ramos, BL2, HelaS3 and Colo704. Overexposition showed that also HelaS3 extract contains some histones **B)** H3 was normalized on tubulin amount present in each extract. Values of the left chart are expressed as integrated density. Image-J program was used to calculate bands relative area and density. Further calculations were performed to remove background specific for each single band. Values of the right chart are expressed as ng of H3 in 1 µg of extract. ng of H3 were calculated from western blot analysis of purified H3 protein.

3.11 - ANALYSIS OF DNA SUBSTRATES ABLE TO ALTER THE CHROMATIN COMPACTION

Nucleosome formation is not exclusively related to histones abundance, but can also be influenced by specific DNA sequences. To check if the repair of supercoiled plasmid used in IVR is altered by nucleosome formation, we tested plasmid sequences with differential affinity for nucleosomes.

It is known that both base pairing and base stacking contribute to DNA double helix stability (179). AT pairing can be destabilizing and adjacent bases have a role in

determining the fluctuational opening of the DNA double helix (179). Stacked/closed and unstacked/open conformations are governed by stacking negative Gibbs free energy (-ΔG): higher ΔG correlate with dynamism. The chicken Ig lambda *locus* was analysed for the stacking ΔG of overlapping 30 consecutive bases (Figure 35) (180) Probabilities of occupancies of closed and opened DNA conformations were determined. In particular, a region between variable and constant genes (74 % AT composition) was identified as not compatible with nucleosomes formation, while a closed region (54 % GC composition) was selected for its high probability to associate with histones (Figure 35). The two regions were named “low nucleosome (LN)” and “high nucleosome (HN)” forming regions, respectively. HN region was PCR amplified and cloned into the pGL3.41 backbone. Due to high AT content in the LN template, it was not possible to get a specific product, thus the LN region was purchased from Genescript and cloned into the pGL3.41 backbone.

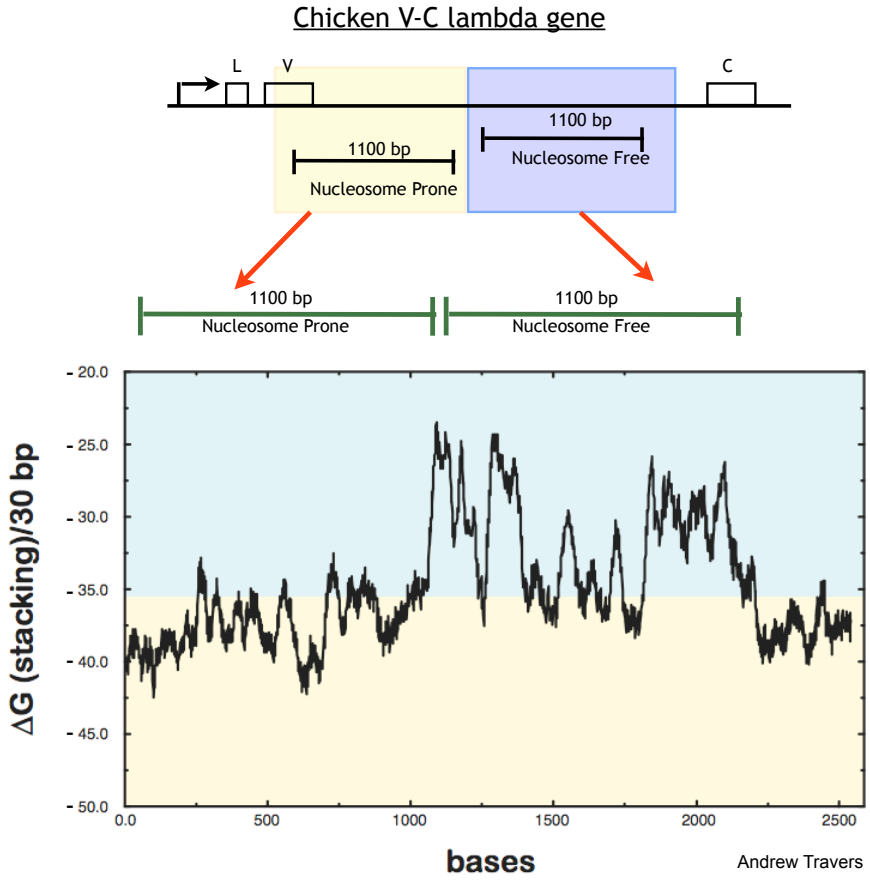


Figure 35 - Chicken variable and constant regions analysis for affinity with nucleosomes formation

Stacking -ΔG of overlapping 30 consecutive bases was analysed and regions with high -ΔGs (yellow) were separated from regions with low -ΔG (turquoise).

3.11.1 - pEI7150 has regions reluctant to form nucleosomes and pEI7151 has regions prone to form nucleosomes

Nucleosomes formation on those sequences was validated with Atomic Force Microscopy (AFM). AFM has been used to visualise plasmids in the past (160). It is a quantitative and qualitative tool, as AFM can measure real physical distances and single strandedness, as well as identify bound proteins (e.g. nucleosomes). Using multiple approaches, we visualised DNA using polyornithine-coated MICA surfaces to modulate the electrostatical repulsions between DNA phosphates and MICA. Images were obtained from a Nanowizard AFM (JPK Instruments) in AC mode in air (Figure 9), to image molecules more gently and without moving them. Indeed, in AC mode, the tip is intermittently contacting the surface with constant oscillation, without being trapped by adhesive forces coming from sample.

Both LN/HN inserts and their respective final plasmids (pEI7150 and pEI7151, respectively) were imaged, checking for plasmid integrity and topology (Figure 36A and 36B). No major differences were observed between the two plasmids.

Specific DNA sequence patterns are favoured for DNA bending and packaging in chromatin, in particular the 601 DNA was shown to bind with high affinity histones octamers to form nucleosomes *in vitro* (181). To verify nucleosome formation with AFM 601 DNA was used in a nucleosome reconstitution assay; incubating the DNA template with the core histones in the presence of high concentrations of NaCl. This allowed for the proper deposition of positively-charged histones onto the negatively-charged DNA and to prevent undesired interactions of the histones with other molecules. Various ratios of 601 DNA and recombinant octamer were combined to find the optimal conditions for the complex formation, with a nearly 2:1 molar ratio of histones to 601 DNA selected (160 fmols of DNA and 280 fmols of histones octamer) (data not shown). After dialysis, nucleosomes in low salt concentration were visualized with AFM as compacted and higher molecules, compared to the non-treated DNA (Figure 36C).

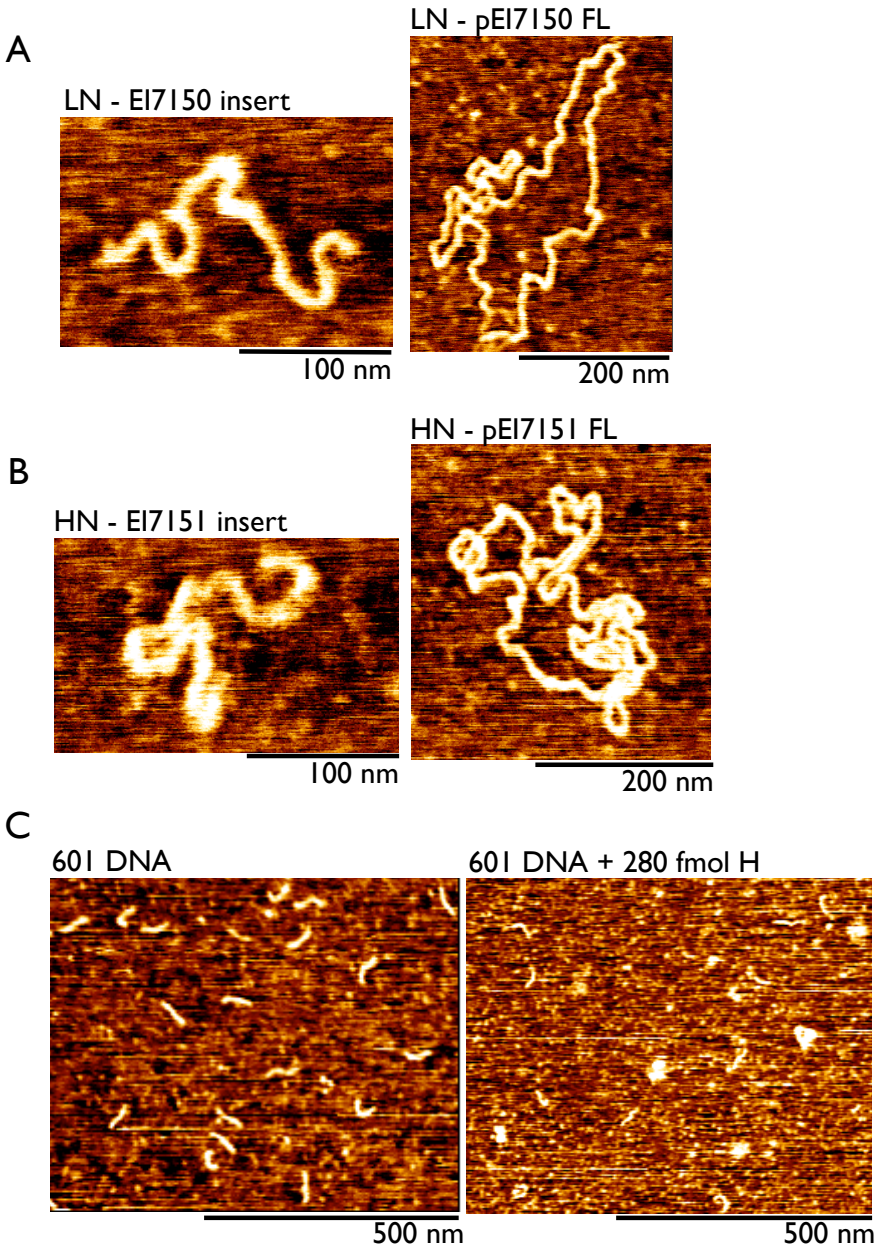


Figure 36 - Different plasmids topology and nucleosome formation visualization with AFM
A) pEI7150 and its LN insert were visualized in AC mode using JPK Nanowizard3 AFM. **B)** pEI7151 and its HN insert were visualized in AC mode using JPK Nanowizard3 AFM **C)** 160 fmols of 601 DNA and 280 fmols of recombinant human histone octamer (H) were incubated in 20 µL of high salt buffer for 30 min before dialysis. The final solution in 0.2 M NaCl, 10 mM Tris pH 8.0, 1 mM EDTA and 1 mM DTT was added onto MICA poly-L-ornithine coated surface and imaged in AC mode. Images are representative of different field of view.

In vitro nucleosome reconstitution assay using pEI7150 and pEI7151 inserts, LN and HN regions was performed and visualised. When low nucleosome forming regions were

incubated with histones and visualized with AFM, no major changes in DNA height or topology were observed (Figure 37).

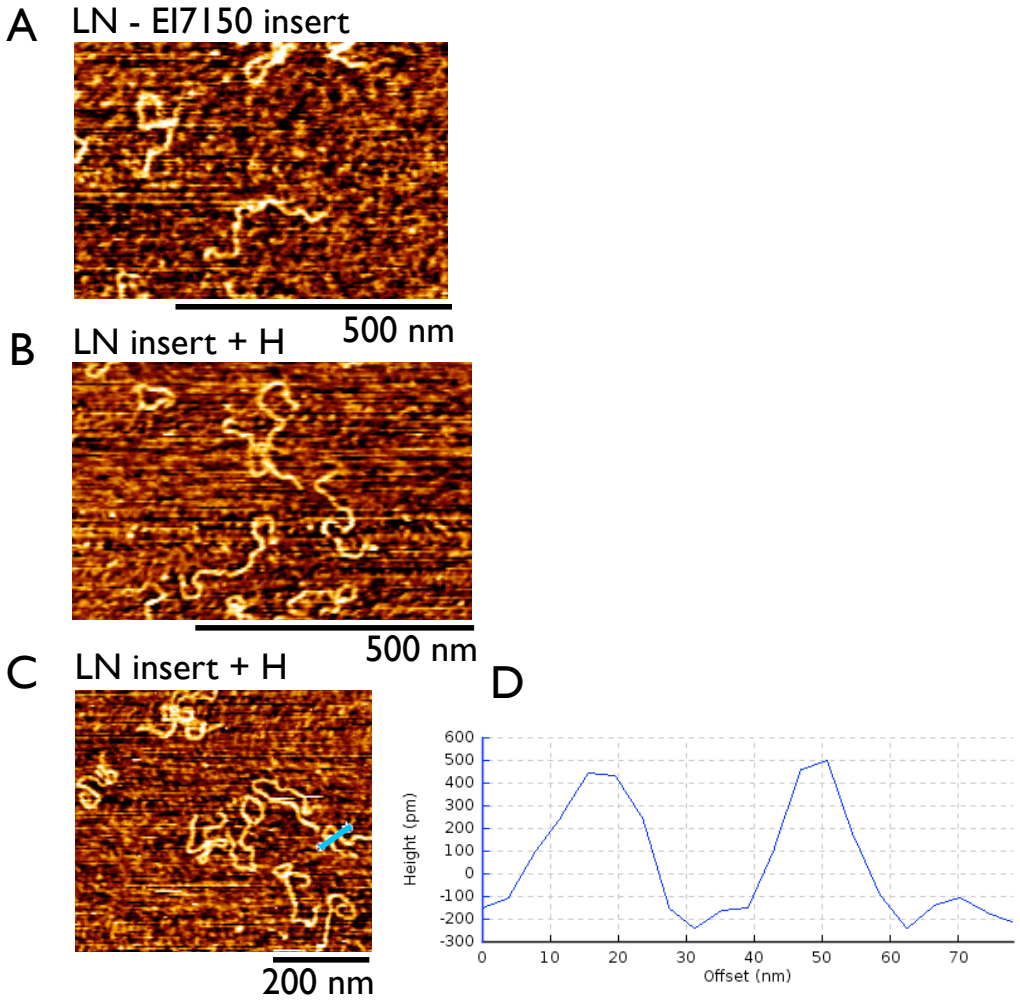


Figure 37 - LN region topology in the presence of histone octamers

A) LN insert were visualized in AC mode using JPK Nanowizard3 AFM. **B-C)** LN insert in the presence of histone octamer was visualized in AC mode using JPK Nanowizard3 AFM. 160 fmols of pEI7150 insert (1.1 Kb) and 280 fmols of recombinant human histone octamer (H) were incubated in 20 µL of high salt buffer for 30 min before dialysis. The final solution in 0.2 M NaCl was loaded onto MICA poly-L-ornithine coated surface and imaged in AC mode. JPK image processing program: turquoise lines crossed DNA and compacted DNA. **D)** The height of the regions crossed by the line (turquoise in figure C) is reported. No major changes in DNA topology or height are observed. Images are representative of different field of view.

On the other hand, when high nucleosome forming regions were incubated with histones under the same conditions, we could observe a significant change in DNA conformation, with an increase in DNA compaction and in the number of crosslinks intra- and inter-strand (Figure 38 B-C). Moreover, when we measured the height of the condensed

regions we could observe a 3 to 4 fold increase in those regions compared to non-condensed DNA (Figure 38 D-I). DNA is known to be 2 nm in height, while nucleosomes form a cylinder 9 nm in diameter and 5 nm in height. According to literature, low force tapping mode measurements of DNA height have reported smaller dimensions than 2 nm, both in air (less then 0.7 nm) and in buffer (0.4 nm)(182). One possible explanation is that the treated substrate (MICA surface on which DNA is attached) can be the source of morphological changes seen in the DNA. According to our measurements DNA is 0.3-0.4 nm high, while condensed regions are 0.7 to 0.8 nm high, thus we observed a 2 fold increase in high. We can conclude that pEI7150 insert is not forming nucleosomes, while pEI7151 insert has high affinity for nucleosome assembly.

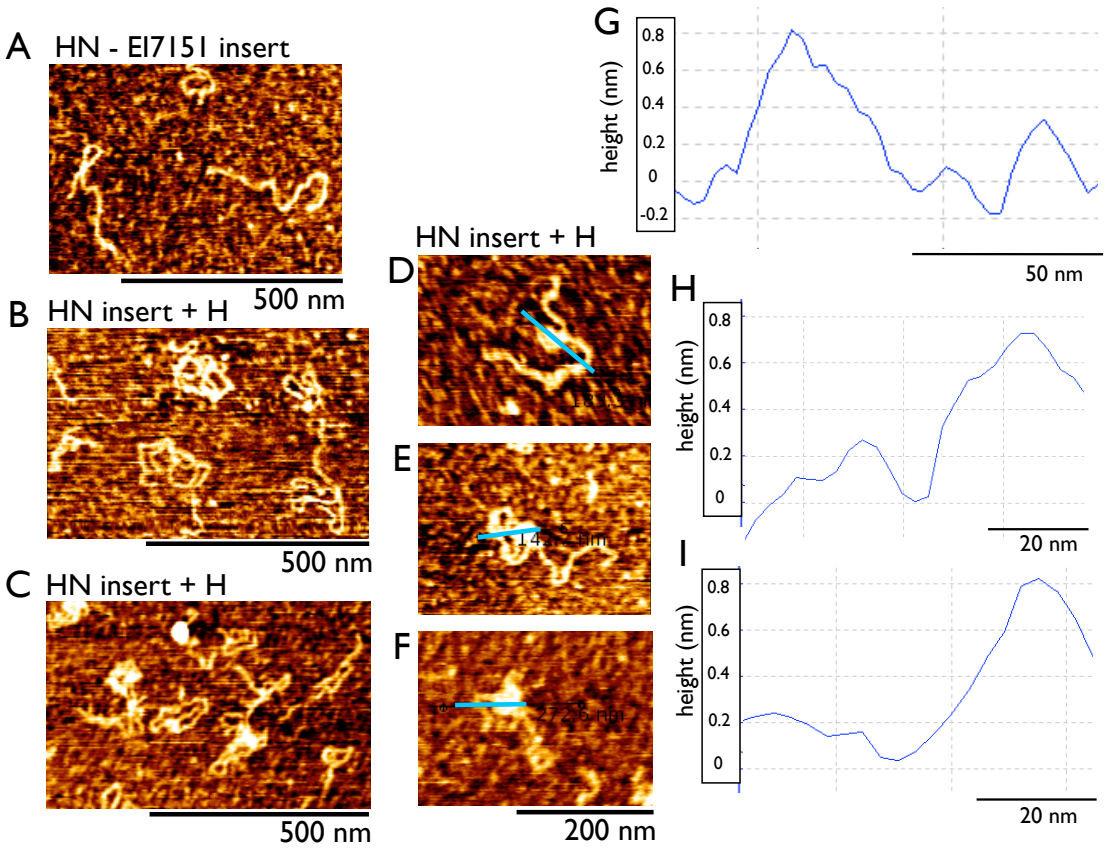


Figure 38 - HN region topology in the presence of histone octamers

A) HN insert were visualized in AC mode using JPK Nanowizard3 AFM. **B-C)** HN insert in the presence of histone octamer was visualized in AC mode using JPK Nanowizard3 AFM. 160 fmols of pEI7151 insert (1.1 Kb) and 280 fmols of recombinant human histone octamer (H) were incubated in 20 µL of high salt buffer for 30 min before dialysis. The final solution in 0.2 M NaCl was loaded onto MICA poly-L-ornithine coated surface and imaged in AC mode. Compaction of DNA is observed. **E-F)** Magnifications of HN insert in the presence of histone octamer**.** JPK image processing program: turquoise lines crossed DNA and compacted DNA. **G-I)** DNA

structures height was analysed with JPK image processing program (turquoise lines in Figure E-F). In the right schemes the height of the regions crossed by the line is reported. Images are representative of different field of view.

3.12 - NUCLEOSOME FORMATION ALTER DNA REPAIR PROTEIN RECRUITMENT UPON AID-INDUCED DAMAGE

Once confirmed that plasmids with different ΔGs have different affinity for histones proteins, we analysed those plasmids in the IVR assay. Plasmids were prepared simultaneously at 4°C and topology was monitored in a gel mobility assay (Figure 39A).

pEI7150 and pEI7151 are 5.1 kb long, while pGL4.31 is 6 kb long. Topology was conserved among different plasmids, independently from incubation with G-AID: supercoiled fraction was always over 90 % (Figure 39A).

Subsequently, IVR was performed with increasing concentrations of FE, using as control the standard plasmid pGL4.31 (Figure 39B). During qPCR amplification of the 3 plasmids an identical DNA segment was amplified, flanking the HN or LN regions. Following AID damage, differences in the total repair activity in the FE were observed. Differences could be the result of differential AID deamination activity or differential DNA repair activity or a combination of the two.

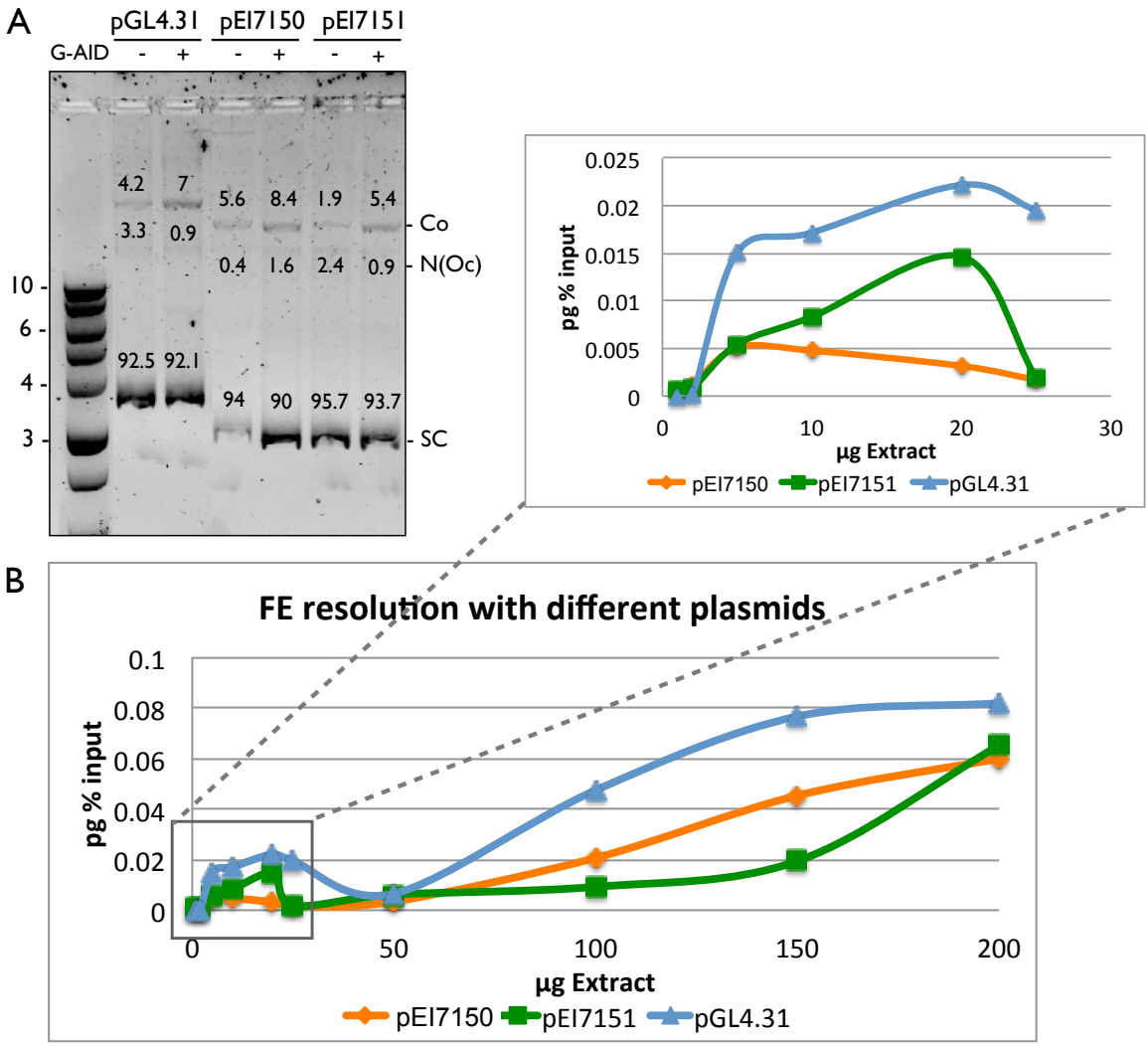


Figure 39 - Plasmids with different affinities for nucleosome formation differentially repaired AID-induced lesion

A) Gel mobility assay using different plasmids (produced simultaneously at 4°C). pGL4.31 is 6 Kb, while pEI7150 (LN) and pEI7151 (HN) are 5.1 Kb. Supercoiled DNA (SC) migrated faster than nicked/open circular (N/Oc). SC fraction is more than 90 % and no topological changes are observed if G-AID is pre-incubated 30 min with the plasmids. **B)** 300 ng of each plasmid (pGL4.31, pEI7150 or pEI7151) were incubated 30 min with G-AID WT or mt and were processed by different amounts of FE (X-axis). Reaction was performed in the presence of bio-dC. Deaminated and repaired plasmids were isolated and quantified as pg as % of input (total amount of plasmids). Primers used for the qPCR analysis amplified the same region in the different plasmids. G-AID mt results were subtracted from G-AID WT results. Magnification of lower concentration shown in insert.

Overall there was a decrease in the total repair activity when pEI7150 (LN) and pEI7151 (HN) were used as substrate, as compared to pGL4.31. This was independent from the concentration of FE used. This indicated that FE was not a limiting factor in the lesion

resolution. As observed before (Figure 17B), when 20 to 50 µg of extract were used, an inhibition of global repair activity was registered before the linear profile appeared again. This phenomenon was visible with all the 3 substrates, indicating this was not related to the sequence context or to the formation of nucleosomes, but rather to the concentration of active DNA repair molecules. One possibility is that when molecules reached a defined concentration they can interact more easily among them, exerting activating or inhibitory functions, such as competition. If we consider the two peaks as distinct repair pathways activation, we observed in the presence of pEI7151 (HN) a more pronounced activation of the first pathway and a subdued activation of the second one, while with pEI7150 we observed exactly the opposite trend. We hypothesized that these profiles were the result of nucleosome interference with repair protein recruitment.

To distinguish if the number of dU have an influence on nucleosomes or if chromatin is influencing repair activity, we performed two IVR experiments: 1) time-course of AID deamination reaction and 2) incubation of FE with UGI and aphidicolin inhibitors to distinguish single repair pathways involvement in AID lesion resolution (Figure 40).

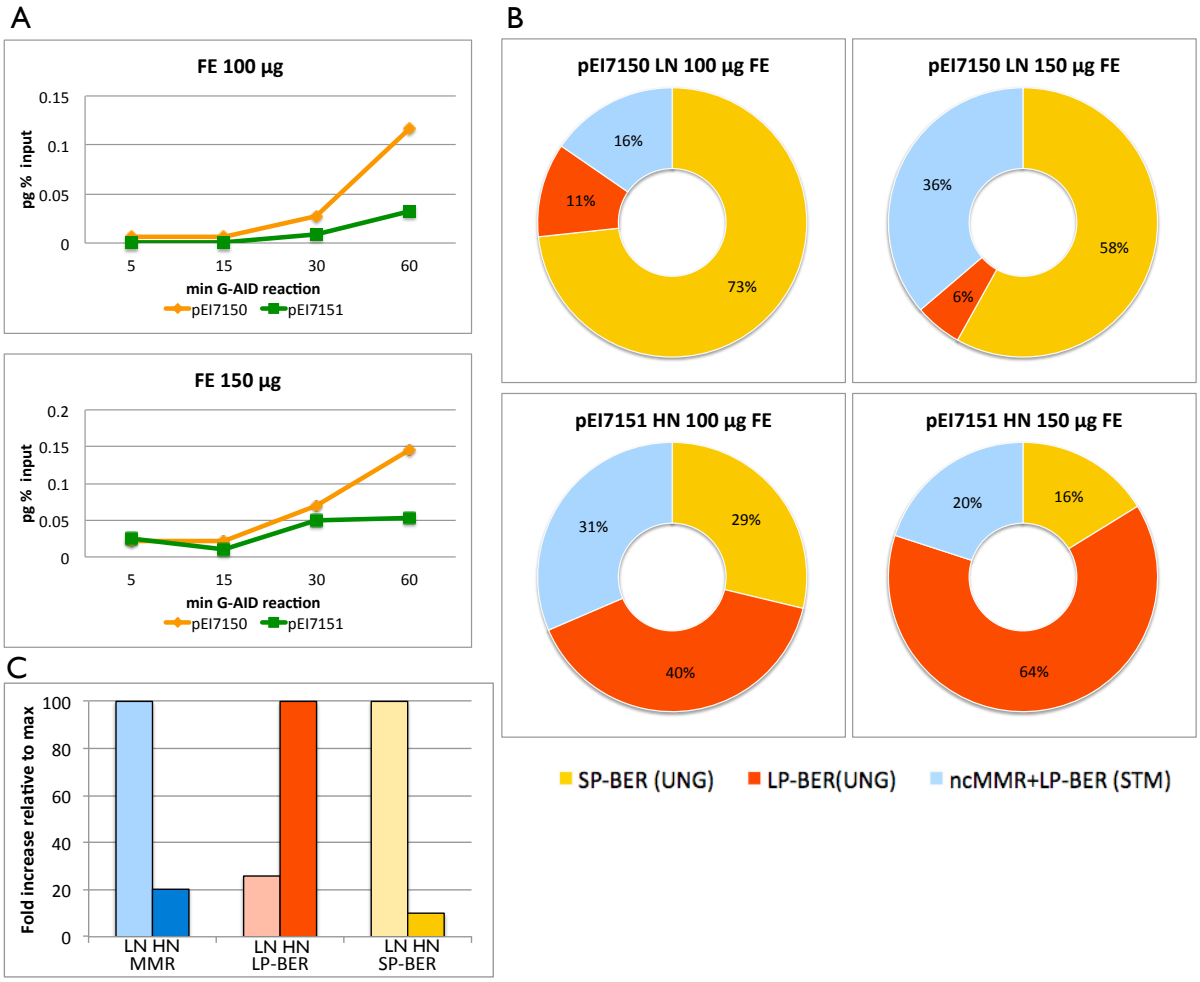


Figure 40 - DNA sequence composition influences both G-AID deamination activity and IVR repair activity

A) G-AID WT or mt were incubated with 300 ng of pEI7150 and pEI7151 and reactions were stopped at different timepoints before performing the repair phase of IVR, in the presence of 150 µg of FE and bio-dC. pEI7150 allowed an higher deamination and repair reaction in particular after 60 min. **B)** Single repair pathway analysis, AID lesions were analysed when IVR was performed in the presence of pEI7150 or pEI7151. FE was incubated with bio-dC or with UGI 10 min on ice before addition of bio-dC. White circle represents unknown pathways contribution (SP-BER STM dependent and clMMR). **C)** Single repair pathways were calculated referring to AID deamination activity (shown in A) Maximum activity between the two plasmids in each repair pathway is set to to100. Light colours refer to pEI7150 (LN); dark colours refer to pEI7151 (HN).

From the time-course analysis (Figure 40A), independently from the FE concentration used, we observed differences due to G-AID deamination reaction in the two substrates. Major differences were appreciable at 60 min, where in pEI7150 (LN) G-AID was 2-3 fold more active in IVR compared to pEI7151 (HN) (Figure 40A). Hence, it was possible that

pEI7150 was more easily targeted by G-AID protein, while pEI7151 was less deaminated and repaired.

We then analysed different repair pathways recruitment at the lesion combining results from bio-dC or bio-dA incorporation with the use of UGI and aphidicolin (Figure 40B). From the preliminary results, it is clear that nucleosome formation potentially affected the repair pathways recruitment. pEI7150 (LN) is mostly repaired by SP-BER (UNG-dependent), while pEI7151 (HN) is mostly repaired by LP-BER (UNG-dependent). When DNA is unlikely binding nucleosomes, supercoiled DNA is easily accessible for the fast component of BER (SP-BER), while when DNA is wrapped around nucleosomes it is preferentially repaired by the slow component of BER (LP-BER). Moreover, if repair pathways recruitment is calculated in relation to AID deamination activity (2.5 fold higher in pEI7150 after 60 min – Figure 40A) MMR is 4 fold reduced in pEI7151 compared to pEI7150 (Figure 40C). LP-BER and SP-BER differences are conserved. G-AID could have a role in sending signals to the repair pathways or could be directly involved in recruiting DNA repair proteins.

3.13 - PROTEIN INTERFERING WITH AID CAN MODIFY ITS ACTIVITY AND INFLUENCE REPAIR PROTEIN RECRUITMENT

Our lab recently discovered that AID has a novel function by influencing the downstream IVR activity (Willmann, Franchini, *et al*. under revision). The work showed that AID associates with the sliding clamp PCNA, which is a cofactor of numerous repair pathways, including MMR and LP-BER. This association can influence PCNA post-translational modifications at K164, a site known to influence DNA repair choice, error prone or error free (183,184). As a direct consequence of AID-PCNA interaction, PCNA K164 modification can be altered, and hence AID can influence how dU lesions are processed. To verify the association and influence of AID on PCNA in the IVR, PCNA was incubated with G-AID before IVR reaction, and the overall deamination and repair is reduced (Figure 41). This is possibly due to PCNA interfering with G-AID binding and activity. While if

PCNA is added to deaminated plasmid after AID-activity and before the extract repair reaction, overall repair could be enhanced (Figure 41). This is possibly due to AID recruiting PCNA directly to the lesion. PCNA addition to the FE post lesion generation did not change the IVR repair activity profile, possibly due to the large amount of PCNA in the FE.

Moreover if G-AID protein is removed before the repair reaction is performed, a decrease in the overall repair activity is observed (data not shown). This suggests that AID functions are important not only in the IVR deamination reaction, but also in the subsequent repair reactions.

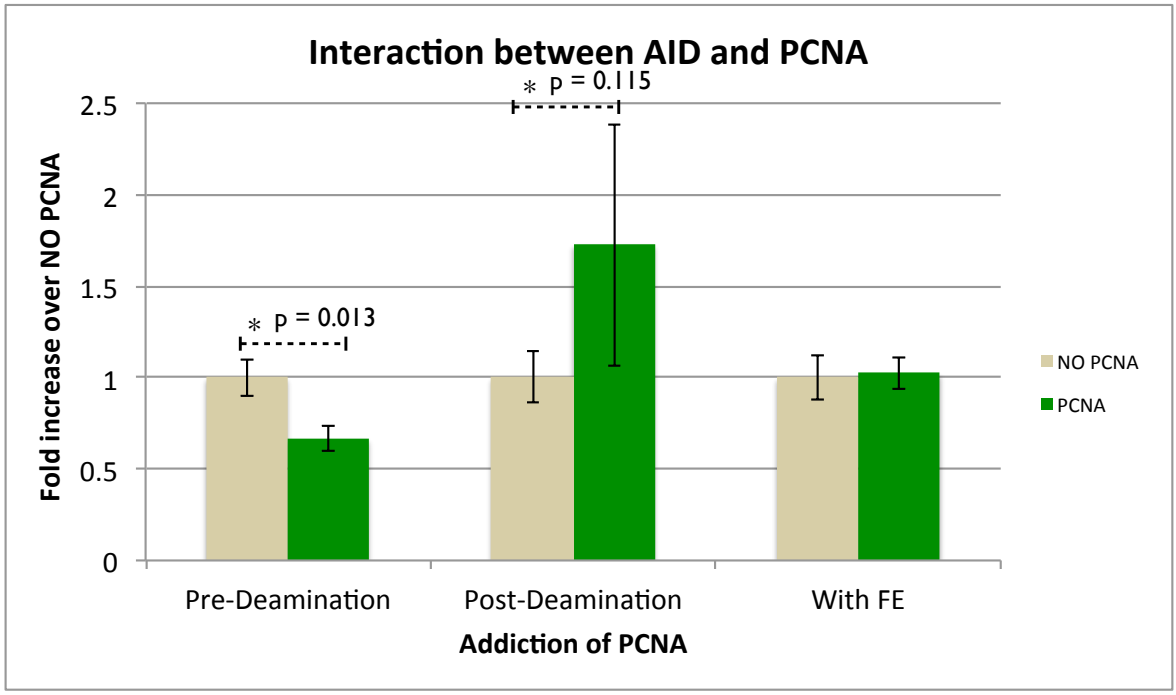


Figure 41 - PCNA interfere with G-AID induced DNA resolution

IVR assay was performed with 6 fold excess PCNA incubation with G-AID before and after deamination reaction or during repair reaction (with FE). pGL4.31 was used during the assay. Values of repaired and deaminated plasmids are expressed as pg as % of input. Error bars indicate SD (n=3). Statistical analysis (t-test) was performed on differences between A and C in Ramos and HelaS3. * p values are indicated in the figure.

CHAPTER 4 - DISCUSSION

Genomes are under constant threat from chemical, physical, and biological agents that induce DNA damage. Proficient repair of those damages by multiple DNA repair pathways ensures the survival of the organism. On the other hand, DNA lesions are necessary for some physiological mechanisms, such as the formation of a healthy adaptive immune system and proper chromosome segregation during meiosis. DNA lesion resolution plays a key role in determining the distribution of mutation *versus* repair. Understanding how physiologically induced lesions are resolved provides insight into how the balance between homeostasis and evolution regulates survival.

In the present study I investigated how a single type of lesion (uracil) is processed by different cellular systems, with a new biochemical assay named IVR. AID, the prototypical DNA deaminase and a physiological lesion generating protein, was used to introduce DNA lesions, and multiple repair pathways acting during the processing of dU lesions were characterised. In particular, with the IVR, we demonstrated that different cell extracts respond to AID-damage in a specific way and that lesion processing is influenced by different molecular pathways, such as single repair protein activity, transcription, chromatin remodelling and protein association with AID.

4.1 - ESTABLISHMENT OF AN *IN VITRO* APPROACH TO STUDY DNA LESION RESOLUTION

We established a new biochemical assay, IVR, which recapitulates the activity of AID in combination with the subsequent resolution by an extract. G-AID was targeted to SC DNA and the induced-damage was repaired by addition of FE or cellular extracts. This method allows the direct quantification of G-AID activity and extract specific repair activity.

To perform IVR assay, we ensured to quantify exclusively resolution of DNA damage generated by AID and not via other mechanisms. Each single component of the assay

was evaluated for unspecific DNA damage induction: AID, the substrate DNA, and the source of DNA repair proteins.

The principal difficulty in performing biochemical studies about AID and repair kinetics is that the protein is sticky and aggregates quickly after purification, thus possibly binding nucleases. A protocol was optimized for G-AID fusion protein production from *E. coli*. Other labs reported the advantages of AID protein purification from a bacterial system, in that the purified proteins do not contain RNA and are active (185), unlike the GST-AID purified from insect cells, which requires ribonuclease activity to deaminate a DNA substrate (116). We demonstrated that G-AID proteins produced from *E. coli* are active and endonucleases-free (Figure 12 and 13). Catalytic activity of AID is a requirement for the initiation of the IVR assay (Figure 16), while absence of nucleases ensures to quantify repair of DNA lesion exclusively due to AID.

We then controlled the DNA substrate of IVR, by maximizing SC fraction production (160) to ensure access of ssDNA to AID. Moreover, we demonstrated that the plasmid topology was not affected by AID deamination reaction.

Finally, in the IVR assay we controlled the source of DNA repair proteins. *Xenopus laevis* egg extract was used as model because it is cost-effective, easy to prepare, and more importantly it has been used for DNA repair studies since 1987 (186); indeed, in the past several experiments have been performed injecting damaged DNA into FE and monitoring DNA damage response or cell cycle progression (163,187). The major disadvantage of the system is the absence of transcription during early embryonic division (176), which could affect DNA repair.

Studying DNA repair in mammalian extracts is more complex, because lysate were shown to be less efficient in repair compared to Xenopus egg extract. Moreover, it is more difficult to obtain active DNA repair proteins which are not contaminated by unspecific nucleases. The background activity from contaminants was removed during calculations, because each experiment was performed in parallel using a G-AID catalytic mutant as control. Repair activity detected using a catalytic dead protein as damaging agent (due to

extracts unspecific repair or to G-AID co-purified contaminates), was excluded from our calculations.

Selection of different human cell lines was aimed to investigate microenvironment contribution to damage resolution. In particular, we started our analysis comparing B cell and non-B cell lines. By checking how different cell lines process DNA lesions, we can analyse different networks required for maintaining DNA stability. We used B cell lines, since AID is physiologically expressed in those cells and thus quantification of repair pathways activation can be a direct reflection of what is occurring *in vivo* during *Ig locus* diversification.

With a human cell based system the field of research can be expanded to primary cancer cells to have more clinical relevance. The study of DNA repair mechanisms activation in different cancer cells or at different stages of cancer progression can provide important insight into oncology. Cancer cells are exposed to high levels of replication stress and DNA damage. Understanding specific DNA repair pathways activation enables the use of specific DNA repair inhibitors to trigger apoptosis and cell death. Indeed, several inhibitors of DNA repair have been developed and clinical trials are ongoing (188); with the IVR assay it will be possible to calculate major repair pathways involved in cancer cells and address specific drug targeting.

The beauty of the IVR system lies in its versatility; once all the aspects of AID lesion generation have been controlled, I used the IVR assay to perform different kind of analysis:

- extracts repair protein compositions and activation, through the use of repair protein inhibitors or through the preparation of extracts from synchronized cells;
- chromatin interaction with DNA repair proteins using a modified substrate (nucleosomes and chromatin retaining plasmids *versus* not nucleosome assembling plasmids);
- analysis of AID protein activity through the association with other proteins (i.e. PCNA).

Further analysis can be performed using depleted extracts. For example, one can use extracts derived from cells genetically ablated from MMR or BER proteins (Ilaria Spadafora and Elisabetta Incorvaia, unpublished observation) or extracts depleted of

nucleosomes. Zierhut and colleagues developed a method to remove histones H3 and H4 from *Xenopus* egg extracts, such as they are incapable of forming nucleosomes (189). This extract could be used to better investigate how nucleosomes influence AID damage and subsequent DNA repair.

Also DNA substrate used during IVR can be modified, for example a methylated DNA can be used to further investigate AID function in DNA demethylation (143). Mutated DNA can be used to mimic DNA of cancer cells and repair pathways activated can be investigated.

4.2 - DIFFERENT CELL LINES ACTIVATE DIFFERENT REPAIR SYSTEMS UPON AID-INDUCED DNA LESION

To determine if the AID-induced lesions are processed in a similar manner in different biological context, we perform the IVR resolution phase using extracts from different cell lines.

All cellular extracts are capable to repair AID-induced damage, but to different extent, such that maximum repair activity is different among the extracts used. All cellular extracts display maximum repair activity at 20 µg, while FE is fully active only at >150 µg. As packing of DNA is a primary determinant of slow repair in chromatin and knowing that FE is able to support chromatin assembly *in vitro* (190), I hypothesize that chromatin assembly is involved in slowing down repair.

Besides playing a role in maximum repair activity, concentration of extracts, hence concentration of repair protein, plays an important role in the single repair pathways activation: some repair pathways are more active at low concentration of extracts, while others are predominantly active at higher concentrations (Figure 22 and 23). This trend is influenced by the cellular microenvironment: at low concentrations in B-cells and in FE, SP-BER is more active, while in non-B cells clMMR is the predominant pathway activated. Interestingly, analysing DNA repair proteins content in each extract we derived that 1) EXO1, a key protein of MMR, is more abundant in non-B cell extracts, 2) B cell are enriched in UNG and MBD4 glycosylases, and 3) FE is expressing high amounts of UNG

(Figure 21). This quantification correlate with the first specific response activated by DNA repair machinery upon AID damage (Figure 22): at lower concentrations proteins more abundant are the first to reach the lesion and actively perform repair.

Besides concentration, the time needed for each repair protein to reach the lesion site may play a role in specific pathways activation. Glycosylases can perform both lesion recognition and excision, while MutS has no cleavage activity towards DNA and needs to be associated with both ATP and a mismatch to recruit MutL and activate nucleases. Moreover, SP-BER does not require DNA re-synthesis and was reconstituted *in vitro* using only 4 purified enzymes (191), while 11 active proteins are required to carry out MMR in *E. coli* and *in vitro* (192). I hypothesise, that besides displaying easier lesion recognition and a higher turnover, SP-BER components may achieve the concentration needed to perform repair faster than MMR, which require more proteins to reach the active conformations.

This is what we observe in B cells and in FE, while in non-B cells we observe the opposite trend, with more clMMR activation al lower concentrations. Probably, due to their abundance, MMR protein can recognize and bind first the lesion, which is then less accessible for BER.

From the maximum repair activity we quantified single repair pathways active in repairing AID lesion, and we observed significant differences between different cell lines. LP-BER is hardly detectable in non-B cell lines, while in Ramos it is the predominant functional pathway. FE displays a profile of repair pathway activation very similar to Ramos. Both systems express endogenous AID. AID may be linked to a specific microenvironment and to specific repair pathways abundance of deficiency in the cells or systems. Ramos and BL2 displayed some similarities (e.g in LP-BER activation) and some differences (e.g. in MMR activation); it would then be interesting to stimulate endogenous AID expression in BL2 cell line and analyse possible variations in single repair pathways activation.

Another observation derived from this quantification is that ncMMR is activated both in B-cell lines and in non-B cell lines. Our results confirm recent findings that error-prone ncMMR can occur also outside the immune system (53). Biological relevance of error-

prone repair has been extensively discussed for healthy immune system development, however I was expecting non-B cells to activate only error-free repair, as mutation introduction can have detrimental effects. Non-B cell lines used in this project are immortalized. p53 and p21, both able to interact with PCNA, were shown to play an important role in down-regulating the error-prone TLS polymerases recruitment in mammalian cells (193). In cancer cells, p53 or p21 inactivation can lead to TLS deregulation and to mutagenic bypass of DNA lesions. It would then be interesting to look at the expression of TLS polymerases in the cells used in the IVR. Furthermore, one can silence p53 in cellular extracts and evaluate how molecules important for tumorigenesis can influence DNA repair pathways activation. At the moment, I hypothesize that ncMMR is finely regulated and that a certain balance between introduction of mutation and efficient repair is needed for cell survival.

Our assay is the first to demonstrate a quantifiable cell specific response to AID-induced damage. Cellular microenvironments play a role in determining which pathway is preferentially activated and when.

4.3 - DNA GLYCOSYLASES DISPLAYED DIFFERENT ACTIVATION KINETICS IN DIFFERENT CELL LINES

To better investigate single repair pathway role in AID lesion resolution, kinetics of DNA glycosylases involved in uracil recognition and SP-BER initiation were analysed. In all the extracts analysed, UNG activity was dominant over the other uracil glycosylases with STM playing only a minor role in SP-BER initiation (less than 10 % - Figure 29A). This is in agreement with previous observations, as UNG showed the highest turnover compared to other uracil recognising glycosylases (194) and UNG deficiency cannot be fully compensated by other uracil recognising glycosylases (130).

Contrary to cell extracts, UNG proteins in FE have a low turnover. Moreover, FE is the extract displaying higher SMUG1, TDG, and MBD4 activities (34 % - Figure 24A). Previously, SMUG1 was shown do display a very low turnover compared to UNG, and

distinct roles were proposed for the two proteins in BER: slow SMUG1 enzyme was proposed to repair deaminated cytosine in non-replicating chromatin, while highly coordinated repair by UNG was proposed for replicating DNA (134). We cannot analyse replication interference, because we inhibit replicative polymerases with aphidicolin, but we observed, at least in FE, a role for both UNG and STM glycosylases. Low UNG turnover and relevant STM activity may be indicative of coordination between DNA glycosylases to activate BER. FE's low UNG turnover may be directly related to high amount of extract needed to perform maximum repair activity in IVR assay (150 µg were needed to reach maximum activity), as 68 % of uracil repair is initiated by DNA glycosylases in FE. It is known that transcribed DNA is repaired faster than non-transcribed DNA. We confirmed that Xenopus embryonic genome is transcriptionally silent and this may be related to nucleosome interference with AID deamination and repair, resulting in slow repair pathways activation (low UNG turnover and lots of active molecules necessary to achieve maximum repair in FE) and in a major role for SMUG1.

Glycosylases activity calculated on small oligonucleotides was compared to SP-BER on longer DNA molecules, accessible by other repair proteins, such as MMR and LP-BER components. A compensatory mechanism, derived from UNG inhibition (used in IVR) is actuated by STM glycosylases. This compensation was not observed in FE, probably because basal levels of STM glycosylases are already enough to support UNG inactivation. The compensatory mechanism is likely to be stimulated by other repair pathways, as it is not occurring in small oligonucleotides.

4.4 - DNA SEQUENCE SPECIFICITY INFLUENCE AID-INDUCE LESION RESOLUTION

I also analysed if other mechanisms could influence deamination and repair activity, in particular we evaluated chromatin, since chromatin compaction can impede access of repair proteins to the damaged sites (178). We decided to change a single parameter in IVR assay, inserting in the SC plasmid sequences with altered affinity for nucleosomes. Sequences with low or high affinity for nucleosome formation (called LN and HN,

respectively) were validated via Atomic Force Microscopy (Figures 37 and 38) and were subsequently tested in the IVR assay with FE. We observed significant differences both in AID deamination activity and in single repair pathways activation upon AID damage. In particular, G-AID seems to have higher affinity for low nucleosome forming sequences, while the DNA repair pathway more affected by this substrate change is BER. Despite low nucleosome regions contains only 26 % of GC bases, G-AID induced more dC deaminations in plasmids with LN region, compared to HN sequence (54 % GC composition). It is known that AID has a preferential target site (120,129), but other DNA features seem to influence G-AID targeting. In the presence of non-nucleosome forming DNA the majority of uracil repair is performed by SP-BER and MMR pathways, while in the presence of nucleosome-forming DNA, LP-BER is the major repair pathway activated.

One study demonstrated that DNA glycosylases and AP endonuclease are also active on DNA substrates bound by nucleosomes, but their activity was decreased by 90 % compared to naked DNA (195). We didn't observe a global reduction of DNA repair when HN regions were inserted in the SC plasmid, but we reported a switch from one repair pathway to another. 1 Kb DNA sequence alteration was enough to trigger different DNA repair pathways activation. This is possibly occurring through the interplay of different mechanisms, including transcription, proteins with nucleosome disassembly activity and additional factor that allow DNA repair to proceed in the specific context.

Perturbation of nucleosome assembly and nucleosome dynamics *in vivo* will lead to aberrant gene expression and genome instability, it this then important to understand how DNA repair machinery is coordinated with a variety of DNA dependent processes.

We recently discovered that AID has a second function: it can influence PCNA post-translational modification, and thus it can select preferential repair pathway activation. Subsequent to AID lesion generation, PCNA interact with specific proteins allowing access to the damaged *locus*. This process requires the coordination and interaction between different proteins and with histones. In the IVR assay chromatin remodelling is occurring after AID damage, but if we consider AID second function in regulating PCNA, nucleosome formation upon extract addition to the deaminated plasmids may induce the

displacement of AID from the lesion site. This phenomenon can contribute to a differential regulation of PCNA and different repair pathways are activated. In this scenario DNA with nucleosome favourable base-stacking induce LP-BER activation, while SP-BER and MMR are the preferential pathways activated in the presence of non-nucleosome stacked DNA; probably because less regulation is needed in this context or repair proteins are already at the damaged site.

Based on these observations we derived multiple-models (Figure 42).

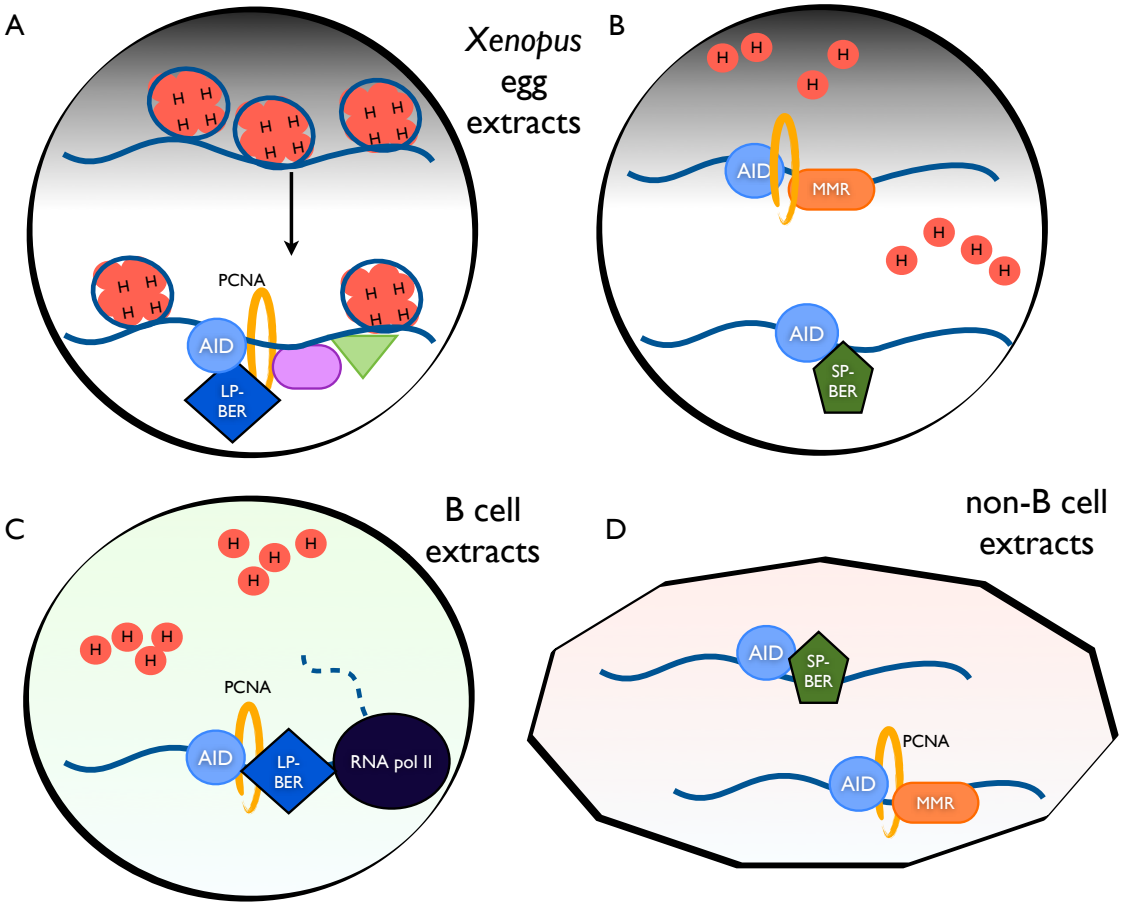


Figure 42 – Multiple-model of how cellular microenvironment influence AID-induced DNA damage repair

A) FE extract perform the repair of DNA harbouring HN forming sequences. Nucleosomes (red bubbles wrapped by blue DNA), PCNA and AID collaborate to trigger LP-BER activation **B)** FE extract perform the repair of DNA harbouring LN forming sequences. SP-BER is the major pathway recruited. **C)** In B cells extracts transcription and PCNA trigger LP-BER activation upon AID-induced damage. **D)** In non-B cells extracts specific cellular environment leads to SP-BER and MMR activation.

Cellular context influences which repair pathway is activated upon AID damage, in particular DNA dependent processes, such as chromatin and transcription, have a role in determining which repair pathway is preferentially recruited.

Depending on the DNA substrate, G-AID will insert a certain number uracil lesions. Once FE is added to the assay to induce repair, (again depending on the DNA substrate) it will either form nucleosomes or not. In the first case, other cellular proteins, chromatin-related, are necessary to disassemble chromatin and allow DNA repair machinery access to damaged sites. AID itself has a role in influencing other proteins, such as PCNA, known to interact with different repair proteins. Interaction and coordination between different processes lead to the recruitment of a specific repair pathway: LP-BER (Figure 42A). If the DNA substrate has a different sequence composition, with low nucleosome potential formation, more uracil lesions will be introduced by G-AID and FE addition will not lead to nucleosomes formation. In this scenario DNA is not protected by nucleosomes and major repair pathways recruited are SP-BER and MMR (Figure 42B).

In B cell extracts histones are present, thus it will be interesting to see how a modified DNA can influence nucleosome formation and DNA repair. We know that those cells have a short cell cycle, high transcriptional activity and specific cellular environment, which influence how a lesion is processed. The specific environment and AID interacting with PCNA may collaborate to recruit LP-BER (Figure 42C). Identification and separation of SP-BER, LP-BER, and MMR pathways involvement in uracil repair is a direct reflection of AID-lesion processing *in vivo* during *Ig locus* diversification in B cells. Finally in non-B cells another specific environment, characterized by low transcription and low histones, may be related to multiple pathways recruitment at the lesion, indeed SP-BER and MMR are the preferential pathways activated in non-B cells (Figure 42D).

4.5 – FUTURE PLANS

4.5.1 - IVR future plans

I was able to demonstrate that different cell lines activate different repair mechanisms upon AID damage and that cellular context is regulating this process.

- I will complete the analysis of chromatin interference with DNA repair in B cells and non-B cells, as we did in FE.
- To assess if endogenous AID may have a role in influencing protein recruitment (independently from its catalytic activity) a comparative IVR analysis will be performed using extracts derived from activated or not activated BL2 cell lines.
- With further analysis, the physiological process of DNA demethylation can be directly investigated, using a methylated substrate in IVR.
- Other physiological or tumorigenic processes will be dissected studying repair pathways recruitment performed by primary cell lines.

4.5.1 - analysis of G-AID efficiency and G-AID function in DNA repair recruitment

I would like to better characterize G-AID deamination activity and G-AID function in recruiting DNA repair proteins:

- A more detailed quantitative analysis can be performed knowing exactly the efficiency of G-AID deamination. With AFM, I can quantify exactly the number of cytosines deaminated and repaired, and better quantify G-AID efficiency. I will directly visualise plasmids after incubation with G-AID with AFM. After IVR resolution the substrate plasmid will have biotins incorporated, which can be visualised using streptavidin proteins.
- At the same time the use of a plasmid containing a single uracil in IVR, in the absence or presence of AID, can give us more information about AID function in recruiting specific repair proteins.

- This function will be further investigated analysing if each single cell line can induce specific AID post-translational modifications, possibly involved in specific repair protein recruitment.

CHAPTER 5 - REFERENCES

1. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709-715.
2. Lindahl, T. and Andersson, A. (1972) Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. *Biochemistry (Mosc).* **11**, 3618-3623.
3. Frederico, L.A., Kunkel, T.A. and Shaw, B.R. (1990) A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry (Mosc).* **29**, 2532-2537.
4. Rydberg, B. and Lindahl, T. (1982) Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *EMBO J.*, **1**, 211-216.
5. Lindahl, T. and Barnes, D.E. (2000) Repair of endogenous DNA damage. *Cold Spring Harb. Symp. Quant. Biol.*, **65**, 127-133.
6. Kunkel, T.A. and Bebenek, K. (2000) DNA replication fidelity. *Annu. Rev. Biochem.*, **69**, 497-529.
7. Pearson, C.E., Nichol Edamura, K. and Cleary, J.D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet*, **6**, 729-742.
8. Durkin, S.G. and Glover, T.W. (2007) Chromosome fragile sites. *Annu. Rev. Genet.*, **41**, 169-192.
9. Sinha, R.P. and Hader, D.P. (2002) UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci*, **1**, 225-236.
10. Riley, P.A. (1994) Free radicals in biology: oxidative stress and the effects of ionizing radiation. *Int. J. Radiat. Biol.*, **65**, 27-33.
11. Snyder, L. and Champness, W. (1997) Molecular genetics of bacteria.
12. Toyooka, T. and Ibuki, Y. (2009) Cigarette sidestream smoke induces phosphorylated histone H2AX. *Mutat. Res.*, **676**, 34-40.
13. Msiska, Z., Pacurari, M., Mishra, A., Leonard, S.S., Castranova, V. and Vallyathan, V. (2010) DNA double-strand breaks by asbestos, silica, and titanium dioxide: possible biomarker of carcinogenic potential? *Am. J. Respir. Cell Mol. Biol.*, **43**, 210-219.
14. Calderon-Garciduenas, L., Osnaya, N., Rodriguez-Alcaraz, A. and Villarreal-Calderon, A. (1997) DNA damage in nasal respiratory epithelium from children exposed to urban pollution. *Environ. Mol. Mutagen.*, **30**, 11-20.
15. Toller, I.M., Neelsen, K.J., Steger, M., Hartung, M.L., Hottiger, M.O., Stucki, M., Kalali, B., Gerhard, M., Sartori, A.A., Lopes, M. *et al.* (2011) Carcinogenic bacterial pathogen Helicobacter pylori triggers DNA double-strand breaks and a DNA damage response in its host cells. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 14944-14949.
16. Pickering, M.T. and Kowalik, T.F. (2006) Rb inactivation leads to E2F1-mediated DNA double-strand break accumulation. *Oncogene*, **25**, 746-755.
17. Tachiwana, H., Shimura, M., Nakai-Murakami, C., Tokunaga, K., Takizawa, Y., Sata, T., Kurumizaka, H. and Ishizaka, Y. (2006) HIV-1 Vpr induces DNA double-strand breaks. *Cancer Res.*, **66**, 627-631.
18. Callegari, A.J. and Kelly, T.J. (2007) Shedding light on the DNA damage checkpoint. *Cell Cycle*, **6**, 660-666.
19. Bartek, J., Bartkova, J. and Lukas, J. (2007) DNA damage signalling guards against activated oncogenes and tumour progression. *Oncogene*, **26**, 7773-7779.
20. Berkovich, E., Monnat, R.J., Jr. and Kastan, M.B. (2007) Roles of ATM and NBS1 in chromatin structure modulation and DNA double-strand break repair. *Nat Cell Biol*, **9**, 683-690.
21. Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R., 3rd, Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lerenthal, Y. *et al.* (2007) ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science*, **316**, 1160-1166.
22. Bernstein, C., Bernstein, H., Payne, C.M. and Garewal, H. (2002) DNA repair/pro-apoptotic dual-role proteins in five major DNA repair pathways: fail-safe protection against carcinogenesis. *Mutat. Res.*, **511**, 145-178.
23. Burma, S., Chen, B.P. and Chen, D.J. (2006) Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA Repair (Amst)*, **5**, 1042-1048.
24. Burma, S. and Chen, D.J. (2004) Role of DNA-PK in the cellular response to DNA double-strand breaks. *DNA Repair (Amst)*, **3**, 909-918.
25. Sirbu, B.M. and Cortez, D. (2013) DNA damage response: three levels of DNA repair regulation. *Cold Spring Harb Perspect Biol*, **5**, a012724.
26. Dexheimer, T.S. (2013) *Chapter 2. DNA Repair Pathways and Mechanisms*.
27. Barnes, D.E. and Lindahl, T. (2004) Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu. Rev. Genet.*, **38**, 445-476.

28. Hegde, M.L., Hazra, T.K. and Mitra, S. (2008) Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Res.*, **18**, 27-47.
29. Lindahl, T. (1974) An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. *Proc. Natl. Acad. Sci. U. S. A.*, **71**, 3649-3653.
30. Caldecott, K.W., McKeown, C.K., Tucker, J.D., Ljungquist, S. and Thompson, L.H. (1994) An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. *Mol. Cell. Biol.*, **14**, 68-76.
31. Frosina, G., Fortini, P., Rossi, O., Carrozzino, F., Raspaglio, G., Cox, L.S., Lane, D.P., Abbondandolo, A. and Dogliotti, E. (1996) Two pathways for base excision repair in mammalian cells. *J. Biol. Chem.*, **271**, 9573-9578.
32. Klungland, A. and Lindahl, T. (1997) Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1). *EMBO J.*, **16**, 3341-3348.
33. Petermann, E., Ziegler, M. and Oei, S.L. (2003) ATP-dependent selection between single nucleotide and long patch base excision repair. *DNA Repair (Amst)*, **2**, 1101-1114.
34. Caldecott, K.W. (2003) XRCC1 and DNA strand break repair. *DNA Repair (Amst)*, **2**, 955-969.
35. Noren Hooten, N., Fitzpatrick, M., Kompaniez, K., Jacob, K.D., Moore, B.R., Nagle, J., Barnes, J., Lohani, A. and Evans, M.K. (2012) Coordination of DNA repair by NEIL1 and PARP-1: a possible link to aging. *Aging (Albany NY)*, **4**, 674-685.
36. Noren Hooten, N., Kompaniez, K., Barnes, J., Lohani, A. and Evans, M.K. (2011) Poly(ADP-ribose) polymerase 1 (PARP-1) binds to 8-oxoguanine-DNA glycosylase (OGG1). *J. Biol. Chem.*, **286**, 44679-44690.
37. Lynch, H.T. and de la Chapelle, A. (1999) Genetic susceptibility to non-polyposis colorectal cancer. *J. Med. Genet.*, **36**, 801-818.
38. Modrich, P. and Lahue, R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.*, **65**, 101-133.
39. Kunkel, T.A. and Erie, D.A. (2005) DNA mismatch repair. *Annu. Rev. Biochem.*, **74**, 681-710.
40. Raschle, M., Marra, G., Nystrom-Lahti, M., Schar, P. and Jiricny, J. (1999) Identification of hMutLbeta, a heterodimer of hMLH1 and hPMS1. *J. Biol. Chem.*, **274**, 32368-32375.
41. Tran, P.T., Erdeniz, N., Symington, L.S. and Liskay, R.M. (2004) EXO1-A multi-tasking eukaryotic nuclease. *DNA Repair (Amst)*, **3**, 1549-1559.
42. Johnson, R.E., Kovvali, G.K., Guzder, S.N., Amin, N.S., Holm, C., Habraken, Y., Sung, P., Prakash, L. and Prakash, S. (1996) Evidence for involvement of yeast proliferating cell nuclear antigen in DNA mismatch repair. *J. Biol. Chem.*, **271**, 27987-27990.
43. Warbrick, E. (2000) The puzzle of PCNA's many partners. *Bioessays*, **22**, 997-1006.
44. Schaetzlein, S., Chahwan, R., Avdievich, E., Roa, S., Wei, K., Eoff, R.L., Sellers, R.S., Clark, A.B., Kunkel, T.A., Scharff, M.D. *et al.* (2013) Mammalian Exo1 encodes both structural and catalytic functions that play distinct roles in essential biological processes. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, E2470-2479.
45. Rada, C., Di Noia, J.M. and Neuberger, M.S. (2004) Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell*, **16**, 163-171.
46. Bardwell, P.D., Woo, C.J., Wei, K., Li, Z., Martin, A., Sack, S.Z., Parris, T., Edelmann, W. and Scharff, M.D. (2004) Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat Immunol*, **5**, 224-229.
47. Cascalho, M., Wong, J., Steinberg, C. and Wabl, M. (1998) Mismatch repair co-opted by hypermutation. *Science*, **279**, 1207-1210.
48. Green, B., Belcheva, A., Nepal, R.M., Boulianne, B. and Martin, A. (2011) The mismatch repair pathway functions normally at a non-AID target in germinal center B cells. *Blood*, **118**, 3013-3018.
49. Chahwan, R., Edelmann, W., Scharff, M.D. and Roa, S. (2012) AIDing antibody diversity by error-prone mismatch repair. *Semin. Immunol.*, **24**, 293-300.
50. Langerak, P., Nygren, A.O., Krijger, P.H., van den Berk, P.C. and Jacobs, H. (2007) A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. *J. Exp. Med.*, **204**, 1989-1998.
51. Roa, S., Avdievich, E., Peled, J.U., Maccarthy, T., Werling, U., Kuang, F.L., Kan, R., Zhao, C., Bergman, A., Cohen, P.E. *et al.* (2008) Ubiquitylated PCNA plays a role in somatic hypermutation and class-switch recombination and is required for meiotic progression. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 16248-16253.
52. Zeng, X., Winter, D.B., Kasmer, C., Kraemer, K.H., Lehmann, A.R. and Gearhart, P.J. (2001) DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat Immunol*, **2**, 537-541.

53. Pena-Diaz, J., Bregenhorn, S., Ghodgaonkar, M., Follonier, C., Artola-Boran, M., Castor, D., Lopes, M., Sartori, A.A. and Jiricny, J. (2012) Noncanonical mismatch repair as a source of genomic instability in human cells. *Mol. Cell*, **47**, 669-680.
54. Bohr, V.A., Okumoto, D.S. and Hanawalt, P.C. (1986) Survival of UV-irradiated mammalian cells correlates with efficient DNA repair in an essential gene. *Proc. Natl. Acad. Sci. U. S. A.*, **83**, 3830-3833.
55. Pfeiffer, P., Goedecke, W. and Obe, G. (2000) Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis*, **15**, 289-302.
56. Keeney, S. and Neale, M.J. (2006) Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochem. Soc. Trans.*, **34**, 523-525.
57. Dudley, D.D., Chaudhuri, J., Bassing, C.H. and Alt, F.W. (2005) Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. *Adv. Immunol.*, **86**, 43-112.
58. Lieber, M.R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.*, **79**, 181-211.
59. Chapman, J.R., Taylor, M.R. and Boulton, S.J. (2012) Playing the end game: DNA double-strand break repair pathway choice. *Mol. Cell*, **47**, 497-510.
60. Daley, J.M., Kwon, Y., Niu, H. and Sung, P. (2013) Investigations of homologous recombination pathways and their regulation. *Yale J. Biol. Med.*, **86**, 453-461.
61. Chen, Z., Yang, H. and Pavletich, N.P. (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature*, **453**, 489-484.
62. Friedberg, E.C. (2005) Suffering in silence: the tolerance of DNA damage. *Nat Rev Mol Cell Biol*, **6**, 943-953.
63. McCulloch, S.D. and Kunkel, T.A. (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res.*, **18**, 148-161.
64. Fujii, S. and Fuchs, R.P. (2007) Interplay among replicative and specialized DNA polymerases determines failure or success of translesion synthesis pathways. *J. Mol. Biol.*, **372**, 883-893.
65. Chang, D.J. and Cimprich, K.A. (2009) DNA damage tolerance: when it's OK to make mistakes. *Nat Chem Biol*, **5**, 82-90.
66. Moynahan, M.E. and Jasin, M. (2010) Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol*, **11**, 196-207.
67. Stelter, P. and Ulrich, H.D. (2003) Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. *Nature*, **425**, 188-191.
68. Hoege, C., Pfander, B., Moldovan, G.L., Pyrowolakis, G. and Jentsch, S. (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature*, **419**, 135-141.
69. Chiu, R.K., Brun, J., Ramaekers, C., Theys, J., Weng, L., Lambin, P., Gray, D.A. and Wouters, B.G. (2006) Lysine 63-polyubiquitination guards against translesion synthesis-induced mutations. *PLoS Genet*, **2**, e116.
70. Gu, Y., Parker, A., Wilson, T.M., Bai, H., Chang, D.Y. and Lu, A.L. (2002) Human MutY homolog, a DNA glycosylase involved in base excision repair, physically and functionally interacts with mismatch repair proteins human MutS homolog 2/human MutS homolog 6. *J. Biol. Chem.*, **277**, 11135-11142.
71. Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A.T., Matsumoto, Y., Golemis, E.A., Genuardi, M. and Neri, G. (1999) MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 3969-3974.
72. Gary, R., Ludwig, D.L., Cornelius, H.L., MacInnes, M.A. and Park, M.S. (1997) The DNA repair endonuclease XPG binds to proliferating cell nuclear antigen (PCNA) and shares sequence elements with the PCNA-binding regions of FEN-1 and cyclin-dependent kinase inhibitor p21. *J. Biol. Chem.*, **272**, 24522-24529.
73. Flores-Rozas, H., Clark, D. and Kolodner, R.D. (2000) Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex. *Nat. Genet.*, **26**, 375-378.
74. Clark, A.B., Valle, F., Drotschmann, K., Gary, R.K. and Kunkel, T.A. (2000) Functional interaction of proliferating cell nuclear antigen with MSH2-MSH6 and MSH2-MSH3 complexes. *J. Biol. Chem.*, **275**, 36498-36501.
75. Matsumoto, Y. (2001) Molecular mechanism of PCNA-dependent base excision repair. *Prog. Nucleic Acid Res. Mol. Biol.*, **68**, 129-138.
76. Dianova, II, Bohr, V.A. and Dianov, G.L. (2001) Interaction of human AP endonuclease 1 with flap endonuclease 1 and proliferating cell nuclear antigen involved in long-patch base excision repair. *Biochemistry (Mosc).* **40**, 12639-12644.

77. Meyers, M., Hwang, A., Wagner, M.W., Bruening, A.J., Veigl, M.L., Sedwick, W.D. and Boothman, D.A. (2003) A role for DNA mismatch repair in sensing and responding to fluoropyrimidine damage. *Oncogene*, **22**, 7376-7388.
78. Cannon, W.B. (1929) Organization for physiological homeostasis. *Physiol. Rev.*, **9**, 399-431.
79. Hawn, M.T., Umar, A., Carethers, J.M., Marra, G., Kunkel, T.A., Boland, C.R. and Koi, M. (1995) Evidence for a connection between the mismatch repair system and the G2 cell cycle checkpoint. *Cancer Res.*, **55**, 3721-3725.
80. Zhang, H., Richards, B., Wilson, T., Lloyd, M., Cranston, A., Thorburn, A., Fishel, R. and Meuth, M. (1999) Apoptosis induced by overexpression of hMSH2 or hMLH1. *Cancer Res.*, **59**, 3021-3027.
81. Khanna, K.K., Lavin, M.F., Jackson, S.P. and Mulhern, T.D. (2001) ATM, a central controller of cellular responses to DNA damage. *Cell Death Differ.*, **8**, 1052-1065.
82. Wang, Y. and Qin, J. (2003) MSH2 and ATR form a signaling module and regulate two branches of the damage response to DNA methylation. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 15387-15392.
83. McCready, S.J., Osman, F. and Yasui, A. (2000) Repair of UV damage in the fission yeast Schizosaccharomyces pombe. *Mutat. Res.*, **451**, 197-210.
84. Norbury, C.J. and Zhivotovsky, B. (2004) DNA damage-induced apoptosis. *Oncogene*, **23**, 2797-2808.
85. Lee, E., Nakatsuma, A., Hiraoka, R., Ishikawa, E., Enomoto, R. and Yamauchi, A. (1999) Involvement of histone phosphorylation in thymocyte apoptosis by protein phosphatase inhibitors. *IUBMB Life*, **48**, 79-83.
86. Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646-674.
87. Finch, S.C. (2007) Radiation-induced leukemia: lessons from history. *Best Pract Res Clin Haematol*, **20**, 109-118.
88. Tang, M.S., Wang, H.T., Hu, Y., Chen, W.S., Akao, M., Feng, Z. and Hu, W. (2011) Acrolein induced DNA damage, mutagenicity and effect on DNA repair. *Mol Nutr Food Res*, **55**, 1291-1300.
89. Handa, O., Naito, Y. and Yoshikawa, T. (2011) Redox biology and gastric carcinogenesis: the role of Helicobacter pylori. *Redox Rep*, **16**, 1-7.
90. Wei, L., Gravitt, P.E., Song, H., Maldonado, A.M. and Ozbun, M.A. (2009) Nitric oxide induces early viral transcription coincident with increased DNA damage and mutation rates in human papillomavirus-infected cells. *Cancer Res.*, **69**, 4878-4884.
91. Kanavy, H.E. and Gerstenblith, M.R. (2011) Ultraviolet radiation and melanoma. *Semin. Cutan. Med. Surg.*, **30**, 222-228.
92. Lancaster, J.M., Powell, C.B., Kauff, N.D., Cass, I., Chen, L.M., Lu, K.H., Mutch, D.G., Berchuck, A., Karlan, B.Y. and Herzog, T.J. (2007) Society of Gynecologic Oncologists Education Committee statement on risk assessment for inherited gynecologic cancer predispositions. *Gynecol. Oncol.*, **107**, 159-162.
93. Keimling, M., Volcic, M., Csernok, A., Wieland, B., Dork, T. and Wiesmuller, L. (2011) Functional characterization connects individual patient mutations in ataxia telangiectasia mutated (ATM) with dysfunction of specific DNA double-strand break-repair signaling pathways. *FASEB J.*, **25**, 3849-3860.
94. Viktorsson, K., De Petris, L. and Lewensohn, R. (2005) The role of p53 in treatment responses of lung cancer. *Biochem. Biophys. Res. Commun.*, **331**, 868-880.
95. Bartkova, J., Tommiska, J., Oplustilova, L., Aaltonen, K., Tamminen, A., Heikkinen, T., Mistrik, M., Aittomaki, K., Blomqvist, C., Heikkila, P. *et al.* (2008) Aberrations of the MRE11-RAD50-NBS1 DNA damage sensor complex in human breast cancer: MRE11 as a candidate familial cancer-predisposing gene. *Mol Oncol*, **2**, 296-316.
96. German, J. (1997) Bloom's syndrome. XX. The first 100 cancers. *Cancer Genet. Cytogenet.*, **93**, 100-106.
97. Bohr, V.A. (2005) Deficient DNA repair in the human progeroid disorder, Werner syndrome. *Mutat. Res.*, **577**, 252-259.
98. Lehmann, A.R., McGibbon, D. and Stefanini, M. (2011) Xeroderma pigmentosum. *Orphanet J Rare Dis*, **6**, 70.
99. Manchanda, R., Menon, U., Michaelson-Cohen, R., Beller, U. and Jacobs, I. (2009) Hereditary non-polyposis colorectal cancer or Lynch syndrome: the gynaecological perspective. *Curr. Opin. Obstet. Gynecol.*, **21**, 31-38.
100. Peltomaki, P. (2001) DNA mismatch repair and cancer. *Mutat. Res.*, **488**, 77-85.
101. Iyama, T. and Wilson, D.M., 3rd. (2013) DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst)*, **12**, 620-636.
102. Fishel, R. (2001) The selection for mismatch repair defects in hereditary nonpolyposis colorectal cancer: revising the mutator hypothesis. *Cancer Res.*, **61**, 7369-7374.

103. Lin, D.P., Wang, Y., Scherer, S.J., Clark, A.B., Yang, K., Avdievich, E., Jin, B., Werling, U., Parris, T., Kurihara, N. *et al.* (2004) An Msh2 point mutation uncouples DNA mismatch repair and apoptosis. *Cancer Res.*, **64**, 517-522.
104. Weissman, L., de Souza-Pinto, N.C., Stevnsner, T. and Bohr, V.A. (2007) DNA repair, mitochondria, and neurodegeneration. *Neuroscience*, **145**, 1318-1329.
105. Lee, Y., Chong, M.J. and McKinnon, P.J. (2001) Ataxia telangiectasia mutated-dependent apoptosis after genotoxic stress in the developing nervous system is determined by cellular differentiation status. *J. Neurosci.*, **21**, 6687-6693.
106. Creighton, H.B. and McClintock, B. (1931) A Correlation of Cytological and Genetical Crossing-Over in Zea Mays. *Proc. Natl. Acad. Sci. U. S. A.*, **17**, 492-497.
107. Paigen, K. and Petkov, P. (2010) Mammalian recombination hot spots: properties, control and evolution. *Nat Rev Genet*, **11**, 221-233.
108. Meffre, E., Casellas, R. and Nussenzweig, M.C. (2000) Antibody regulation of B cell development. *Nat Immunol*, **1**, 379-385.
109. Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J. and Schatz, D.G. (2000) The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.*, **18**, 495-527.
110. Wardemann, H., Yurasov, S., Schaefer, A., Young, J.W., Meffre, E. and Nussenzweig, M.C. (2003) Predominant autoantibody production by early human B cell precursors. *Science*, **301**, 1374-1377.
111. Petersen-Mahrt, S. (2005) DNA deamination in immunity. *Immunol. Rev.*, **203**, 80-97.
112. Neuberger, M.S., Harris, R.S., Di Noia, J. and Petersen-Mahrt, S.K. (2003) Immunity through DNA deamination. *Trends Biochem. Sci.*, **28**, 305-312.
113. Manis, J.P., Tian, M. and Alt, F.W. (2002) Mechanism and control of class-switch recombination. *Trends Immunol*, **23**, 31-39.
114. Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O. and Honjo, T. (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.*, **274**, 18470-18476.
115. Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A. *et al.* (2000) Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell*, **102**, 565-575.
116. Bransteitter, R., Pham, P., Scharff, M.D. and Goodman, M.F. (2003) Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 4102-4107.
117. Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E. and Alt, F.W. (2003) Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature*, **422**, 726-730.
118. Petersen-Mahrt, S.K., Harris, R.S. and Neuberger, M.S. (2002) AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature*, **418**, 99-103.
119. Rangam, G., Schmitz, K.M., Cobb, A.J. and Petersen-Mahrt, S.K. (2012) AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS One*, **7**, e43279.
120. Morgan, H.D., Dean, W., Coker, H.A., Reik, W. and Petersen-Mahrt, S.K. (2004) Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.*, **279**, 52353-52360.
121. Coker, H.A. and Petersen-Mahrt, S.K. (2007) The nuclear DNA deaminase AID functions distributively whereas cytoplasmic APOBEC3G has a processive mode of action. *DNA Repair (Amst)*, **6**, 235-243.
122. Chelico, L., Pham, P. and Goodman, M.F. (2009) Stochastic properties of processive cytidine DNA deaminases AID and APOBEC3G. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **364**, 583-593.
123. Barreto, V., Reina-San-Martin, B., Ramiro, A.R., McBride, K.M. and Nussenzweig, M.C. (2003) C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol. Cell*, **12**, 501-508.
124. Brar, S.S., Watson, M. and Diaz, M. (2004) Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. *J. Biol. Chem.*, **279**, 26395-26401.
125. Vaidyanathan, B., Yen, W.F., Pucella, J.N. and Chaudhuri, J. (2014) AIDing Chromatin and Transcription-Coupled Orchestration of Immunoglobulin Class-Switch Recombination. *Front Immunol*, **5**, 120.
126. Willmann, K.L., Milosevic, S., Pauklin, S., Schmitz, K.M., Rangam, G., Simon, M.T., Maslen, S., Skehel, M., Robert, I., Heyer, V. *et al.* (2012) A role for the RNA pol II-associated PAF complex in AID-induced immune diversification. *J. Exp. Med.*, **209**, 2099-2111.

127. Kenter, A.L. (2012) AID targeting is dependent on RNA polymerase II pausing. *Semin. Immunol.*, **24**, 281-286.
128. Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V. *et al.* (2010) Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell*, **143**, 122-133.
129. Beale, R.C., Petersen-Mahrt, S.K., Watt, I.N., Harris, R.S., Rada, C. and Neuberger, M.S. (2004) Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J. Mol. Biol.*, **337**, 585-596.
130. Rada, C., Williams, G.T., Nilsen, H., Barnes, D.E., Lindahl, T. and Neuberger, M.S. (2002) Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol.*, **12**, 1748-1755.
131. Faili, A., Aoufouchi, S., Weller, S., Vuillier, F., Stary, A., Sarasin, A., Reynaud, C.A. and Weill, J.C. (2004) DNA polymerase eta is involved in hypermutation occurring during immunoglobulin class switch recombination. *J. Exp. Med.*, **199**, 265-270.
132. Zeng, X., Negrete, G.A., Kasmer, C., Yang, W.W. and Gearhart, P.J. (2004) Absence of DNA polymerase eta reveals targeting of C mutations on the nontranscribed strand in immunoglobulin switch regions. *J. Exp. Med.*, **199**, 917-924.
133. Di Noia, J.M. and Neuberger, M.S. (2007) Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.*, **76**, 1-22.
134. Pettersen, H.S., Sundheim, O., Gilljam, K.M., Slupphaug, G., Krokan, H.E. and Kavli, B. (2007) Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic Acids Res*, **35**, 3879-3892.
135. Di Noia, J.M., Rada, C. and Neuberger, M.S. (2006) SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *EMBO J.*, **25**, 585-595.
136. Wiesendanger, M., Kneitz, B., Edelmann, W. and Scharff, M.D. (2000) Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. *J. Exp. Med.*, **191**, 579-584.
137. Kemmerich, K., Dingler, F.A., Rada, C. and Neuberger, M.S. (2012) Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice. *Nucleic Acids Res*, **40**, 6016-6025.
138. Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H. and Schatz, D.G. (2008) Two levels of protection for the B cell genome during somatic hypermutation. *Nature*, **451**, 841-845.
139. Roa, S., Li, Z., Peled, J.U., Zhao, C., Edelmann, W. and Scharff, M.D. (2010) MSH2/MSH6 complex promotes error-free repair of AID-induced dU:G mispairs as well as error-prone hypermutation of A:T sites. *PLoS One*, **5**, e11182.
140. Pauklin, S., Sernandez, I.V., Bachmann, G., Ramiro, A.R. and Petersen-Mahrt, S.K. (2009) Estrogen directly activates AID transcription and function. *J. Exp. Med.*, **206**, 99-111.
141. Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F. *et al.* (2011) Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*, **470**, 419-423.
142. Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E. and Reik, W. (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*, **463**, 1101-1105.
143. Franchini, D.M., Chan, C.F., Morgan, H., Incorvaia, E., Rangam, G., Dean, W., Santos, F., Reik, W. and Petersen-Mahrt, S.K. (2014) Processive DNA Demethylation via DNA Deaminase-Induced Lesion Resolution. *PLoS One*, **9**, e97754.
144. Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J. *et al.* (2011) Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell*, **147**, 107-119.
145. Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A. *et al.* (2008) AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell*, **135**, 1028-1038.
146. Kuppers, R. and Dalla-Favera, R. (2001) Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene*, **20**, 5580-5594.
147. Klemm, L., Duy, C., Iacobucci, I., Kuchen, S., von Levetzow, G., Feldhahn, N., Henke, N., Li, Z., Hoffmann, T.K., Kim, Y.M. *et al.* (2009) The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. *Cancer Cell*, **16**, 232-245.
148. Patten, P.E., Chu, C.C., Albesiano, E., Damle, R.N., Yan, X.J., Kim, D., Zhang, L., Magli, A.R., Barrientos, J., Kolitz, J.E. *et al.* (2012) IGHV-unmutated and IGHV-mutated chronic lymphocytic

leukemia cells produce activation-induced deaminase protein with a full range of biologic functions. *Blood*, **120**, 4802-4811.
149. Okazaki, I.M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K. and Honjo, T. (2003) Constitutive expression of AID leads to tumorigenesis. *J. Exp. Med.*, **197**, 1173-1181.
150. Kou, T., Marusawa, H., Kinoshita, K., Endo, Y., Okazaki, I.M., Ueda, Y., Kodama, Y., Haga, H., Ikai, I. and Chiba, T. (2007) Expression of activation-induced cytidine deaminase in human hepatocytes during hepatocarcinogenesis. *Int. J. Cancer*, **120**, 469-476.
151. Matsumoto, Y., Marusawa, H., Kinoshita, K., Endo, Y., Kou, T., Morisawa, T., Azuma, T., Okazaki, I.M., Honjo, T. and Chiba, T. (2007) Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat. Med.*, **13**, 470-476.
152. Schmitz, K.M. and Petersen-Mahrt, S.K. (2012) AIDing the immune system-DIAbolic in cancer. *Semin. Immunol.*, **24**, 241-245.
153. Nik-Zainal, S., Wedge, D.C., Alexandrov, L.B., Petljak, M., Butler, A.P., Bolli, N., Davies, H.R., Knappskog, S., Martin, S., Papaemmanuil, E. *et al.* (2014) Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat. Genet.*, **46**, 487-491.
154. Incorvaia, E., Sicouri, L., Petersen-Mahrt, S.K. and Schmitz, K.M. (2013) Hormones and AID: balancing immunity and autoimmunity. *Autoimmunity*, **46**, 128-137.
155. Pauklin, S., Burkert, J.S., Martin, J., Osman, F., Weller, S., Boulton, S.J., Whitby, M.C. and Petersen-Mahrt, S.K. (2009) Alternative induction of meiotic recombination from single-base lesions of DNA deaminases. *Genetics*, **182**, 41-54.
156. Javaid, S., Manohar, M., Punja, N., Mooney, A., Ottesen, J.J., Poirier, M.G. and Fishel, R. (2009) Nucleosome remodeling by hMSH2-hMSH6. *Mol. Cell*, **36**, 1086-1094.
157. Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H.T., Difilippantonio, M.J., Wilson, P.C., Hanitsch, L., Celeste, A., Muramatsu, M., Pilch, D.R. *et al.* (2001) AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature*, **414**, 660-665.
158. Schrader, C.E., Guikema, J.E., Linehan, E.K., Selsing, E. and Stavnezer, J. (2007) Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. *J. Immunol.*, **179**, 6064-6071.
159. Franchini, D.M., Incorvaia, E., Rangam, G., Coker, H.A. and Petersen-Mahrt, S.K. (2013) Simultaneous in vitro characterisation of DNA deaminase function and associated DNA repair pathways. *PLoS One*, **8**, e82097.
160. Carbone, A., Fioretti, F.M., Fucci, L., Ausio, J. and Piscopo, M. (2012) High efficiency method to obtain supercoiled DNA with a commercial plasmid purification kit. *Acta Biochim. Pol.*, **59**, 275-278.
161. Punchihewa, C., Inoue, A., Hishiki, A., Fujikawa, Y., Connelly, M., Evison, B., Shao, Y., Heath, R., Kuraoka, I., Rodrigues, P. *et al.* (2012) Identification of small molecule proliferating cell nuclear antigen (PCNA) inhibitor that disrupts interactions with PIP-box proteins and inhibits DNA replication. *J. Biol. Chem.*, **287**, 14289-14300.
162. G. Binnig, C.F.Q., and Ch. Gerber. (1986) Atomic Force Microscope. *Phys. Rev. Lett.*, **56**.
163. Desai, A., Murray, A., Mitchison, T.J. and Walczak, C.E. (1999) The use of Xenopus egg extracts to study mitotic spindle assembly and function in vitro. *Methods Cell Biol.*, **61**, 385-412.
164. Dickerson, S.K., Market, E., Besmer, E. and Papavasiliou, F.N. (2003) AID mediates hypermutation by deaminating single stranded DNA. *J. Exp. Med.*, **197**, 1291-1296.
165. Coker, H.A., Morgan, H.D. and Petersen-Mahrt, S.K. (2006) Genetic and in vitro assays of DNA deamination. *Methods Enzymol.*, **408**, 156-170.
166. Blow, J.J. and Laskey, R.A. (1986) Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of Xenopus eggs. *Cell*, **47**, 577-587.
167. Blow, J.J. and Watson, J.V. (1987) Nuclei act as independent and integrated units of replication in a Xenopus cell-free DNA replication system. *EMBO J.*, **6**, 1997-2002.
168. Karran, P., Cone, R. and Friedberg, E.C. (1981) Specificity of the bacteriophage PBS2 induced inhibitor of uracil-DNA glycosylase. *Biochemistry (Mosc).* **20**, 6092-6096.
169. Martin, A., Bardwell, P.D., Woo, C.J., Fan, M., Shulman, M.J. and Scharff, M.D. (2002) Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas. *Nature*, **415**, 802-806.
170. Syvaoja J, S.S., Nishida C, Goldsmith JS, Chui GSJ, Jain S, Linn S. (1990) DNA polymerases alpha, delta, and epsilon: three distinct enzymes from HeLa cells. *PNAS*, **87**, 6664-6668.
171. Bryant JA, F.P., Hughes SG, Sibson DR. (1992) DNA Polymerase-α in Pea is Part of a Large Multiprotein Complex. *J. Exp. Bot.*, **43**, 31-40.

172. Asagoshi, K., Liu, Y., Masaoka, A., Lan, L., Prasad, R., Horton, J.K., Brown, A.R., Wang, X.H., Bdour, H.M., Sobol, R.W. *et al.* (2010) DNA polymerase beta-dependent long patch base excision repair in living cells. *DNA Repair (Amst)*, **9**, 109-119.
173. Karras, G.I., Fumasoni, M., Sienski, G., Vanoli, F., Branzei, D. and Jentsch, S. (2013) Noncanonical role of the 9-1-1 clamp in the error-free DNA damage tolerance pathway. *Mol. Cell*, **49**, 536-546.
174. Guan, X., Madabushi, A., Chang, D.Y., Fitzgerald, M.E., Shi, G., Drohat, A.C. and Lu, A.L. (2007) The human checkpoint sensor Rad9-Rad1-Hus1 interacts with and stimulates DNA repair enzyme TDG glycosylase. *Nucleic Acids Res*, **35**, 6207-6218.
175. Bennett, S.E., Schimerlik, M.I. and Mosbaugh, D.W. (1993) Kinetics of the uracil-DNA glycosylase/inhibitor protein association. Ung interaction with Ugi, nucleic acids, and uracil compounds. *J. Biol. Chem.*, **268**, 26879-26885.
176. Ferrell, J.E., Jr. (1999) Xenopus oocyte maturation: new lessons from a good egg. *Bioessays*, **21**, 833-842.
177. Shen, H.M., Poirier, M.G., Allen, M.J., North, J., Lal, R., Widom, J. and Storb, U. (2009) The activation-induced cytidine deaminase (AID) efficiently targets DNA in nucleosomes but only during transcription. *J. Exp. Med.*, **206**, 1057-1071.
178. Downey, M. and Durocher, D. (2006) Chromatin and DNA repair: the benefits of relaxation. *Nat Cell Biol*, **8**, 9-10.
179. Yakovchuk, P., Protozanova, E. and Frank-Kamenetskii, M.D. (2006) Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res*, **34**, 564-574.
180. Travers, A., Hiriart, E., Churcher, M., Caserta, M. and Di Mauro, E. (2010) The DNA sequence-dependence of nucleosome positioning in vivo and in vitro. *J. Biomol. Struct. Dyn.*, **27**, 713-724.
181. Lowary, P.T. and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.*, **276**, 19-42.
182. J. Vesenka, C.V., I. Kumar, T marsh, E. Henderson. (1999) The diameter of duplex and quadruplex DNA measured by Scanning Probe Microscopy." Scanning Microscopy. *Scanning Microsc.*
183. Moldovan, G.L., Pfander, B. and Jentsch, S. (2007) PCNA, the maestro of the replication fork. *Cell*, **129**, 665-679.
184. Ulrich, H.D. (2009) Regulating post-translational modifications of the eukaryotic replication clamp PCNA. *DNA Repair (Amst)*, **8**, 461-469.
185. Sohail, A., Klapacz, J., Samaranayake, M., Ullah, A. and Bhagwat, A.S. (2003) Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res*, **31**, 2990-2994.
186. Legerski, R.J., Penkala, J.E., Peterson, C.A. and Wright, D.A. (1987) Repair of UV-induced lesions in Xenopus laevis oocytes. *Mol. Cell. Biol.*, **7**, 4317-4323.
187. Prioleau, M.N., Huet, J., Sentenac, A. and Mechali, M. (1994) Competition between chromatin and transcription complex assembly regulates gene expression during early development. *Cell*, **77**, 439-449.
188. Helleday, T., Petermann, E., Lundin, C., Hodgson, B. and Sharma, R.A. (2008) DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer*, **8**, 193-204.
189. Zierhut, C., Jenness, C., Kimura, H. and Funabiki, H. (2014) Nucleosomal regulation of chromatin composition and nuclear assembly revealed by histone depletion. *Nat Struct Mol Biol*, **21**, 617-625.
190. Gaillard, P.H., Martini, E.M., Kaufman, P.D., Stillman, B., Moustacchi, E. and Almouzni, G. (1996) Chromatin assembly coupled to DNA repair: a new role for chromatin assembly factor I. *Cell*, **86**, 887-896.
191. Kubota, Y., Nash, R.A., Klungland, A., Schar, P., Barnes, D.E. and Lindahl, T. (1996) Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. *EMBO J.*, **15**, 6662-6670.
192. Ramilo, C., Gu, L., Guo, S., Zhang, X., Patrick, S.M., Turchi, J.J. and Li, G.M. (2002) Partial reconstitution of human DNA mismatch repair in vitro: characterization of the role of human replication protein A. *Mol. Cell. Biol.*, **22**, 2037-2046.
193. Avkin, S., Sevilya, Z., Toube, L., Geacintov, N., Chaney, S.G., Oren, M. and Livneh, Z. (2006) p53 and p21 regulate error-prone DNA repair to yield a lower mutation load. *Mol. Cell*, **22**, 407-413.
194. Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P.A., Hagen, L., Krokan, H.E. and 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. *J. Biol. Chem.*, **277**, 39926-39936.

195. Beard, B.C., Wilson, S.H. and Smerdon, M.J. (2003) Suppressed catalytic activity of base excision repair enzymes on rotationally positioned uracil in nucleosomes. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 7465-7470.

AKNOWLEDGEMENTS

It has been hard work, very laborious at times, but extremely rewarding and I am surprised and amazed to have made it to this point. During this doctoral journey I grew up both personally and professionally facing the whims and the needs of the outstanding science environment. I am grateful to many people for encouragement and support.

My first sincere thanks go to my project supervisor and mentor, Svend Petersen-Mahrt, for his guidance and for sparking and cultivating my interest in science. His curiosity and dedication to the prove of reasoned hypothesis was contagious. I must thank him also for introducing me to the biochemistry world and for all the precious suggestions he shared with me.

Special thanks goes to Don-Marc Franchini for introducing me to the “resolution work”; thanks for the constant support, advice, calmness, tenacity and for being my source of inspiration.

Thanks to Kerstin Maike-Schmitz for taking the time to proof read this thesis, your input was very important for me.

To Ilaria, to whom I had the pleasure to teach lablife, thanks for travelling the last year of this journey alongside with me, with your practical support and optimism you helped to keep me going and dragged me to the finish line.

Thanks to all present and past lab members, Lara, Gopinath and Marialaura for the fruitful discussions and for everything they did behind the scenes to help my thesis project to reach a defined shape.

To my friends and family, you should know that your support and encouragement was worth more than I can express on paper.

To Luca, thank you for reading this thesis with a critical eye, for your enthusiasm and pride, for your unconditional support and for encouraging me every day. I’m glad to share my map of the world with you -*my-sun-and-stars*.