Mechanisms underlying mutational outcomes of DNA double-strand break repair

Kamp, J.A.

1

CHAPTER 1

DNA double-strand breaks and methods to investigate their repair in Caenorhabditis elegans

Animals are thought to be formed by nature and nurture. While environmental factors like education and nutrition can shape an individual, these factors can affect animals only to a certain extent, because their genetic information, coded in molecules called DNA, dictates their sensitivity to these external factors. However, during the last decades, it has become clear that DNA is not indifferent to external factors. DNA is wrapped around proteins termed histones, and a long DNA molecule wrapped in histones is named a chromosome. External factors can influence the packaging of DNA(Moore 2017), leading to more open or closed configurations, making it more or less accessible for transcription factors, which are molecules that influence the expression of the DNA's genes. When genes get expressed, they give orders for the production of specific proteins. These proteins can be structural proteins, or possess one or more specialized functions. Besides influencing the accessibility of DNA, it is known that DNA is also susceptible to damage by external factors, and erroneous repair of this damage can lead to alteration of the DNA molecule(Volkova, Meier et al. 2020). Thus, DNA is a molecule in which nature and nurture are encoded.

The "code" of DNA is written in four different characters: A, T, C and G. These characters represent distinct molecular building blocks of DNA (adenine, thymine, cytosine and guanine respectively), and are called nucleosides (Figure 1). These nucleosides are connected to a sugar-phosphate backbone, making DNA resemble a string of characters. A nucleoside with a phosphate group is named a nucleotide (nt). On the opposite of each nucleoside, a corresponding nucleoside is placed: an A is always paired with T, C with G, and vice versa, forming a basepair (bp). The corresponding nucleosides are also connected to a backbone, making the DNA consist of two strands (Figure 1). Therefore, we call DNA double-stranded.

When cells divide, their genome (total DNA content) has to be copied to have an identical set of DNA for each daughter cell. In order to achieve this, the DNA's two strands are separated, and enzymes named polymerases place novel nucleotides opposite to the original nucleotides, a process which is called replication(Figure 1). Each strand thus serves as a template for doubling the amount of DNA. Polymerases incorporate about 1 wrong nucleotide per every 100,000 nucleotides, but some mistakes can immediately be corrected by the polymerase itself, a process called proofreading. The structure of the separated strands where replication takes place is named the replication fork.

Nucleotides are asymmetric and have a 5' end and a 3' end. This asymmetry is called directionality. Nucleic acids can only be connected to each other by polymerases in the 5' to 3' direction, therefore replication will occur in 1 direction on each strand. The two strands in the DNA are

aligned in the opposite direction, therefore replication occurs in opposite directions in a DNA molecule, as depicted in Figure 1. Because of the directionality of the DNA, DNA of one strand can be replicated in one go, because it follows the replication fork: the leading strand. The DNA on the opposite strand (lagging strand) has to be replicated in fragments, because replication can only start after the two strands have been separated (Figure 1). The formed fragments are called Okazaki fragments.

After replication, the chromosomes consist of two connected DNA copies, named sister chromatids. These sister chromatids are held together by a protein complex called cohesion. Just before cell division, cohesion is removed to allow separation of the identical DNA copies.

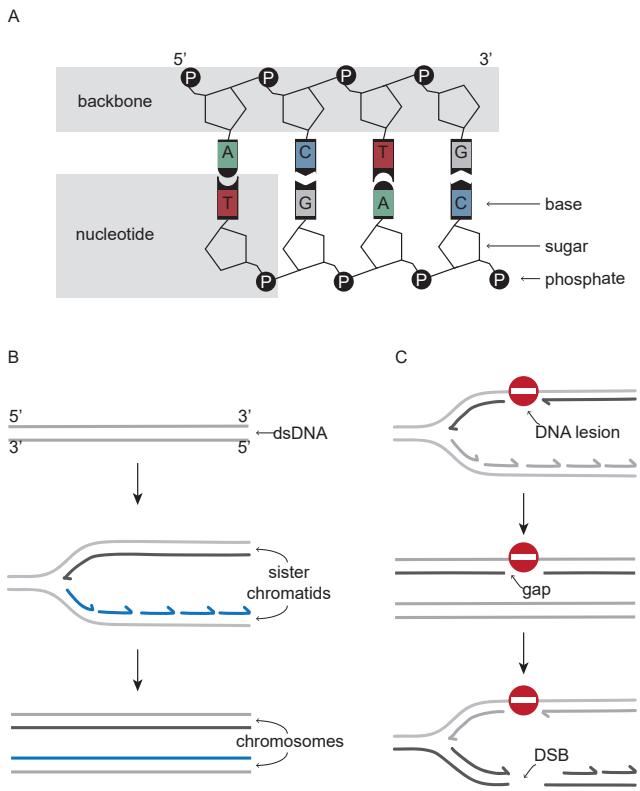

Figure 1 Illustration of DNA and DNA replication A. Graphic illustration of the building blocks of DNA. B. Model of DNA replication. Leading strand DNA is depicted in grey, lagging strand DNA is depicted in blue. C. Model of the induction of a double-strand break (DSB) resulting from a DNA lesion that interferes with replication.

Breaking of DNA

As mentioned before, DNA is prone to damage by external factors. One of the most toxic lesions is breakage of both DNA strands: a double-strand break (DSB). DSBs can be formed directly by breakage of both strands at the same time: strands can break after exposure to irradiation or chemicals that damage the sugar-phosphate backbone of the DNA. DSBs can also occur indirectly when one of the two strands is damaged: if the backbone of one of the two strands was interrupted by a break before replication, separation of the strands during replication will lead to breakage of the DNA molecule (Figure 1). Besides single-strand breaks, single-stranded gaps can also occur in the DNA: one or multiple nucleotides and the backbone connecting these nucleotides are absent in one of the two strands. These gaps arise when the replication machinery was not able to place nucleotides opposite the nucleotides of the replicated strand. This can occur when nucleotides are damaged(van Bostelen, van Schendel et al. 2020) or when the DNA is folded in a structure that cannot be resolved by the replication machinery(Lemmens, van Schendel et al. 2015). Similar to single-strand breaks, these gaps can lead to DSBs in the next replication cycle(Lemmens, van Schendel et al. 2015) (Figure 1).

Induction of DNA double-strand breaks

Besides the occurrence of DSBs after replication of damaged DNA, DSBs can also be induced in other ways. Some breaks are physiological and created by the cell itself, by expressing proteins that can cut the DNA. The goal of making programmed DSBs is usually to allow recombination: shuffling of DNA. This recombination leads to more diversity in the DNA between cells. For example, in developing lymphocytes (cells of the immune system), recombination activating genes 1 and 2 (RAGs) are expressed and these RAGs introduce DSBs in antigen receptor genes, which code for the receptors that recognize foreign material. These receptors are essential for an immune response. The RAGS are guided by recombination sequences in the DNA to determine the specific location of the programmed DSB (Dudley, Chaudhuri et al. 2005). The DSBs in the antigen receptor genes are repaired in different ways to establish receptor diversity, a process called V(D)J recombination (Dudley, Chaudhuri et al. 2005). In the reproductive system, DSBs are also induced. Breaks in the germ cells are made to enable recombination between paternal and maternal DNA, leading to increased genetic diversity in egg cells or sperm(Lam and Keeney 2014). Breaks in germ cells are induced throughout the genome by the protein Spo11, the position of these Spo11-induced DSBs are influenced by multiple factors to ensure proper recombination of genetic information(Lam and Keeney 2014).

Breaks can also artificially be induced using genetic engineering. Mainly specialized proteins called nucleases that recognize a specific DNA sequence and cut at this location. One major

discovery of the last decade is the use of CRISPR/Cas9(Jinek, Chylinski et al. 2012), which allows for DNA cleavage at almost any desired location in the genome. Previously, DSBs could only be induced by restriction endonucleases that recognize a specific nucleotide sequence and cleave the DNA at a specific site. Because the recognition site for CRISPR/Cas9 can be chosen by designing a "guide RNA" differently, the CRISPR/Cas9 method provides researchers with more flexibility. Induction of DSBs at known sites allows researchers to generate genetic mutants and serves a tool to investigate the repair of DSBs.

Repair of DNA double-strand breaks

When DNA breaks, a set of signalling networks detects the break and can activate cell cycle checkpoints and signal to initiate repair. These signalling networks are called the DNA damage response (DDR). DSBs can be repaired in multiple ways, via recombination or end-joining. Which repair mechanism is used, depends on the context in which the break occurs and the nature of the break. Below, I will summarize the major repair routes.

Homologous recombination

When DSBs arise while cells are replicating their DNA, an intact copy of the broken DNA is often available, specifically the sister chromatid. This sister chromatid can be used as a template to repair the broken DNA, this process is named homologous recombination (HR, Figure 2). Because of this template usage, HR usually works without errors. Instead of the sister chromatid, the homologous chromosome can also be used as a template, or other DNA containing a nucleotide sequence identical (also called homologous) to the sequence at the break site. To access the template, break ends are resected: double-stranded break ends are made single-stranded (ssDNA). End resection is performed by multiple proteins and stimulated by BRCA1(Chen, Nievera et al. 2008, Cruz-Garcia, Lopez-Saavedra et al. 2014). The single-stranded ends of the DSB that are exposed after resection are called 3' overhangs. After end resection, the DNA ends are coated by the ssDNA binding Replication Protein A (RPA). RPA is later replaced by the recombinase RAD51, and this replacement is stimulated by BRCA2(Li and Engebrecht 2021). After coating with RAD51, the 3'overhang can invade the homologous template to find complementary nucleotides. When the invaded strand has annealed to the complementary nucleotides, polymerases extend the break end to copy the nucleotides needed for break repair. After this, resolution of the extended break end can occur via different mechanisms: when both ends of the DSB had invaded the template, the strands are cut to resolve the structure(San-Segundo and Clemente-Blanco 2020). When one end had invaded the template, a synthesis-dependent- strand annealing (SDSA) mechanism takes place(San-Segundo and Clemente-Blanco 2020). During SDSA, the extended end detaches

from the template DNA and anneals to the other break end (Figure 2). Most HR proteins are conserved across species.

Non-homologous end-joining

When the cell is not replicating, breaks are repaired via non-homologous end-joining (NHEJ). The major NHEJ factors are conserved across species: the KU dimer, which covers break ends, and ligase 4, which ligates the two break ends to seal the break (Figure 2)(Lemmens and Tijsterman 2011). The Ku dimer is known to protect break ends against resection, thereby preventing HR(Krasner, Daley et al. 2015). Moreover, in vertebrates, additional regulation by 53BP1 and the Shieldin complex stimulates NHEJ(Setiaputra and Durocher 2019).

NHEJ of DSBs can occur without changes to the original DNA sequence, but because NHEJ does not use template DNA to guide repair, errors can occur. For example, NHEJ leads to deletion of nucleotides when the break ends were prone to erosion, or fill-in-synthesis of single-stranded DSB overhangs leads to the addition of nucleotides to the DNA sequence at the break site (Schimmel, Muñoz-Subirana et al. 2021).

Alternative end-joining

When NHEJ is impaired by mutation of genes encoding core NHEJ proteins, mutagenic end-joining activity leading to deletions can still be observed(Boulton and Jackson 1996, Kabotyanski, Gomelsky et al. 1998, Ma, Kim et al. 2003). This activity is called alternative end-joining (Alt-EJ). When the deletions resulting from Alt-EJ are analyzed, an overrepresentation of potential homology usage was observed: identical nucleotide sequences on both break ends were aligned to repair the break (Figure 2). Because of this observation, the term microhomology-mediated end-joining (MMEJ) has also been used(McVey and Lee 2008). Proteins that play a role in different DNA repair pathways, like CtIP, the MRN complex, Ligase 3, XRCC1, FEN1 and PARP1, have been proposed to play a role in Alt-EJ(Sharma, Javadekar et al. 2015, Chang, Pannunzio et al. 2017). One specialized Alt-EJ protein was identified: polymerase theta, encoded by the *POLQ* gene.

Theta-mediated end-joining

A role for polymerase theta in genome stability and the response to DNA damaging crosslinking agents has been described in several organisms for many years(Boyd, Sakaguchi et al. 1990, Inagaki, Suzuki et al. 2006, Muzzini, Plevani et al. 2008), and polymerase theta was often thought to prevent DSBs. *In vivo,* a role for polymerase theta in the repair of DSBs was first described in *Drosophila* in 2010(Chan, Yu et al. 2010).

The mechanism by which polymerase theta repairs DSBs is named polymerase theta-mediated end-joining (TMEJ, Figure 2). This term was introduced(Roerink, van Schendel et al. 2014) to discriminate between Alt-EJ that depends on polymerase theta versus alt-EJ that does not require polymerase theta, like MMEJ in yeast, which does not encode polymerase theta. MMEJ in yeast was shown to rely on stretches of at least 5 homologous nucleotides(McVey and Lee 2008). In species that do encode a polymerase theta ortholog, polymerase theta-independent MMEJ also exists: in mammalian cells Alt-EJ activity that involves long stretches of homology (≥ 18 nucleotides) is present, which is independent of polymerase theta activity(Kelso, Lopezcolorado et al. 2019). Polymerase theta is able to repair breaks using just one homologous nucleotide in *C. elegans*(Koole, van Schendel et al. 2014), while a study using purified human polymerase theta showed that at least two homologous nucleotides were necessary for TMEJ *in vitro*(Kent, Chandramouly et al. 2015). Usage of these homologous nucleotides can be observed by analysis of deletions resulting from breaks repaired by polymerase theta in multiple organisms(Chan, Yu et al. 2010, Roerink, van Schendel et al. 2014, van Schendel, van Heteren et al. 2016, Schimmel, Kool et al. 2017). Another hallmark of TMEJ that can be found in deletion footprints is the occurrence of templated insertions(Schimmel, van Schendel et al. 2019), which arise when two break ends detach and reanneal during TMEJ (Figure 2).

While TMEJ is often described as a backup pathway for NHEJ, TMEJ also repairs DSBs in NHEJ proficient cells and organisms(Koole, van Schendel et al. 2014, Roerink, van Schendel et al. 2014, van Schendel, Roerink et al. 2015, van Kregten, de Pater et al. 2016, Schimmel, Kool et al. 2017). Especially replication-associated breaks and breaks occurring in frequently replicating cells are described to be TMEJ substrates.

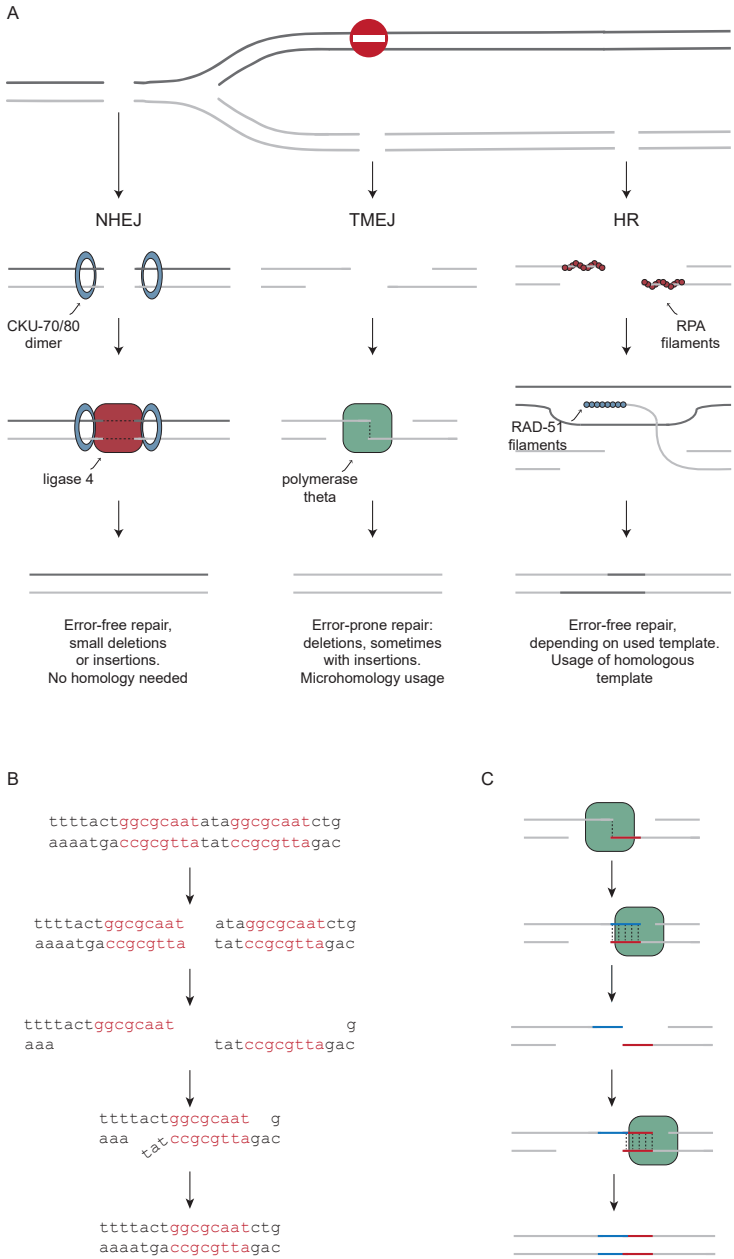

Figure 2 Overview of DNA double strand break repair pathways A. When breaks arise before replication, they are generally repaired via non-homologous end-joining (NHEJ). To enable NHEJ, the two break ends are recognized by the CKU-70/CKU-80 dimer. Ligase 4 seals the break by ligating the two break ends together. NHEJ leads to deletions when the break ends were prone to erosion, or insertions when nucleotides

have been added to the break ends. When the break ends were in good condition, breaks can be repaired without errors. Breaks that occur after replication are predominantly repaired by homologous recombination (HR), except when the homologous template is damaged. When the template is damaged, breaks are repaired via polymerase theta-mediated end-joining (TMEJ). When cells are replicating, the break ends are resected to make the ends single stranded. Polymerase theta uses at least one nucleotide of a single strand overhang to anneal it with a homologous nucleotide at the other break end, thereby repairing the break. TMEJ is an error-prone repair pathway, therefore it will induce deletions, sometimes accompanied by insertions. When HR can be employed, the single-stranded overhangs are coated with replication protein A (RPA), which is subsequently exchanged by RAD51 to enable strand invasion of the homologous template, in this case the sister chromatid. The DNA sequence of the template is then copied to guide error-free repair. *C. elegans* nomenclature is used to name the proteins. B. Hypothetical example of microhomology-mediated end-joining. Homologous nucleotides are depicted in red. When the DNA breaks, the break ends are resected, allowing the homologous nucleotides to anneal. Annealing of the homologous nucleotides leads to loss of one of the original stretches of homologous nucleotides and the nucleotides between the homologous stretches compared to the original sequence: this loss is called a deletion. C. Model of templated insertion formation by polymerase theta. Polymerase theta anneals the two break ends using homologous nucleotides and extends one break end, using the other end as template. When the extended break end detaches and polymerase theta reanneals the strands at the tips of both break ends, nucleotides copied from the break end after the first annealing (depicted in blue) form an insertion.

C. elegans as a model organism for DNA repair studies

In 1974 Sidney Brenner described the worm *Caenorhabditis elegans* as a useful model organism for genetics studies(Brenner 1974) and in 1998, *C. elegans* became the first multicellular organism of which the whole genome was sequenced(Consortium 1998). *C. elegans* are usually hermaphrodites, but 1 in 1000 worms is male, enabling genetic crossing. Males, worms of different larval stages (numbered L1 to L4) and adult hermaphrodites can be distinguished under a microscope. Because of its small size, rapid reproduction cycle and easy maintenance, *C. elegans* has become a common model organism for experimental studies. Moreover, *C. elegans* consists of a fixed number of cells, and all lineages are mapped, enabling precise cell biology studies.

For DNA repair studies, often mammalian cells are used. However, cells cultured in a dish cannot completely mimic the processes in an animal, mainly because cells have to obtain cancer-like features to survive in a dish. Sometimes embryonic stem cells are used, having the benefit that they do not have to be transformed to survive in dishes. However, stem cells are rapidly replicating, therefore only representing a small part of cells in alive animals. Since replication and cell cycle stages have a significant impact on DNA damage and repair, this should be taken into consideration when interpreting data from cultured cells. Another option is the use of laboratory animals like rodents, but due to practical and ethical concerns and strict regulations, these are not the most optimal model for a lot of DNA repair studies. Because laboratory animal laws do not apply to *C. elegans*, researchers have more freedom in designing and upscaling experiments.

Multiple DNA repair assays for *C. elegans* have been developed over the last years, and I will summarize the most common techniques below.

Irradiation and exposure to DNA damaging agents

Worms can be exposed to DNA damaging agents using multiple techniques. Genotoxic substances can be dissolved in the medium on which the worms grow, worms can be soaked in genotoxic substances or the worms can be irradiated using different irradiation sources(Kim and Colaiácovo 2015). The most commonly used sources of irradiation are γ-irradiation (mainly causing DNA breakage) and UV-irradiation (causing 6-4 photoproducts and cyclobutane pyrimidine dimers). Chemical mutagenesis is often performed using ethyl methanesulfonate (alkylating agent leading to the addition of an ethyl group to guanines), camptothecin (inducing single-strand breaks), nitrogen mustard (producing interstrand crosslinks), hydroxyurea (resulting in replication fork arrest by depletion of the nucleotide pool in cells) and trimethylpsoralen (induces interstrand crosslinks in combination with UV irradiation).

The response to DNA damaging agents can be evaluated via multiple methods (Figure 3). The timing of administration and the nature of the read-out is important for the interpretation of the results(Clejan, Boerckel et al. 2006). When worms are exposed to damaging agents in the first larval stage (L1), the cells forming the vulva are not replicating, therefore the efficiency of repair mechanisms that are not dependent on replication can be assessed. For example, irradiating L1 worms with γ-irradiation (IR) and looking at the development of the vulva at the adult stage is a measure for NHEJ efficiency (Figure 3). When core NHEJ genes like *lig-4, cku-70* and *cku*-80 are knocked out by homozygous null mutations, worms irradiated at the L1-stage show protruding vulvas at the adult stage or their offspring hatches within their bodies, because of defects in vulva development. Another read-out reflecting NHEJ efficiency is scoring the development of L1-irradiated worms: while most NHEJ proficient worms will have reached the adult stage within three days after irradiation, NHEJ deficient worms develop slower after DNA damage, and often will still be in the larval stages three days after irradiation.

Alternatively, when worms are irradiated at the last larval stage, L4, their germ cells are replicating. When replication-associated mechanisms like HR are not functioning properly, this leads to a decrease in survival of their offspring. Sensitivity to IR can thus be assessed by quantifying the ratio of unhatched eggs versus alive offspring (Figure 3). Comparing the offspring survival of the irradiated worms of interest to the offspring survival of negative controls (wild-type worms) and positive controls (*brc-1* mutant worms for example), will give a measure of the capacity to repair DNA damage in replicating cells.

Advantages of using exposure to DNA damaging agents for repair research are the relative simplicity of the methods, and the opportunity to perform epistasis analysis: to assess whether

repair genes code for proteins that work in the same pathway, double mutants can be compared to single mutants, to see if combination of mutations leads to additional IR sensitivity compared to the sensitivity of the single mutants. If not, the investigated proteins are likely to work in the same pathway leading to IR resistance. Downsides of DNA damaging agent studies is that the exact location and nature of the damage is not known. Therefore, structural variations like deletions resulting from these damaging agents are difficult to identify.

DNA repair reporters

To study DSB repair at known break sites, DSB reporter assays have been developed for *C. elegans*. Most of these assays are based on break induction by the rare-cutting endonuclease I-SceI. This enzyme recognizes an 18 bp sequence and cuts the DNA at this sequence, leaving a DSB with a 3'overhang of 4 nt. The Isce-I recognition sequence is usually not present in animal genomes, therefore it will not induce DSBs in animals where a recognition site has not been artificially introduced. I-SceI was first employed for DSB repair reporter studies in yeast(Plessis, Perrin et al. 1992), and later also used in DSB reporters in other organisms, like plants(Plessis, Perrin et al. 1992) and mouse cells(Rouet, Smih et al. 1994).

In *C. elegans* reporter systems, I-SceI is introduced into the genome by integration of plasmids (circular DNA molecules) encoding the I-sceI gene and a heat shock promoter that controls the expression of I-SceI. When worms are incubated at 34 degrees, the heat shock promoter is activated, leading to the expression of I-SceI. Plasmids containing I-SceI recognition sequences and corrupted genes encoding proteins that can be visualized by fluorescence microscopy or staining are also integrated in the *C. elegans* genome (Figure 3). When I-SceI is expressed and recognized its recognition site, it cuts the DNA. When the generated DSB gets repaired, the corrupted genes can be restored and their gene products can be visualized by researchers. Examples of DSB repair reporters developed for *C. elegans* (Figure 3) are the SSA reporter (visualizing end-joining using large stretches of homologous nucleotides)(Pontier and Tijsterman 2009) and the HDR reporter (reflecting SDSA efficiency)(Johnson, Lemmens et al. 2013). Stretches of homologous nucleotides are positioned in the reporter in a way that repair using this homology leads to restoration of the corrupted marker genes (Figure 3).

The benefit of DNA repair reporters is that the experiments are easy to scale up, making them valuable tools for screening potential DNA repair mutants. Downsides are that the break induction is difficult to measure and to titrate, and that the sequence context surrounding the break does not vary.

PCR at genomic DSB sites

Another way to analyse DNA repair is by mutational footprint analysis. When the location of a DSB is known, the DNA surrounding the break site can be sequenced to identify erroneous repair of the DSB. One way to know the location of DSBs is by inducing the break via CRISPR/Cas9. Because a guide RNA determines the position of the induced DSB, the break location is known and the sequence flanking the DSB can be analysed for repair footprints. The most conventional way to analyse CRISPR/Cas9 induced repair footprints is by amplifying the region of interest using a Polymerase Chain Reaction and read the code of this amplified DNA using Sanger sequencing. A downside of analysing CRISPR/Cas9-induced DSBs in replicating cells is that it is unknown whether the sister chromatid was intact at the time of break repair: when CRISPR/Cas9 is very efficient, it could in theory cut both sister chromatids. When the sister chromatid also contains a DSB, HR via the sister in not possible, and the CRISPR-induced deletions thus arise in a context in which HR is compromised.

While Cas9 generates blunt DSBs, the Cas9 enzymes can also be modified to only cut one of the DNA strands. These modified Cas9 enzymes are called nickases. When nickases are combined to target both strands, DSBs with single-stranded overhangs can be generated(Schimmel, Kool et al. 2017). This expands the toolset of different types of DSBs that can be studied.

Another way to study genomic DSBs is analysing DNA repair footprints generated by transposons. Transposons are DNA sequences that can 'jump' to new locations in the genome. Upon excision of a transposon, a DSB is formed. Usually, jumping of transposon is suppressed, but DNA transposition can be stimulated by genetically inactivating transposon silencing mechanisms (Sijen and Plasterk 2003). Transposon induced breaks can be repaired via HR, but also via error prone end-joining mechanisms: in somatic cells, the major mechanism leading to mutagenic footprints is NHEJ, in the germline TMEJ(Plasterk 1991, Robert and Bessereau 2007, Robert, Davis et al. 2008, van Schendel, Roerink et al. 2015). Because the original location and sequence of transposons is known, they are a good substrate for DSB repair research. Transposon-induced deletions can be identified using PCR.

As mentioned previously, an important source of DSBs is stalling of replication. The challenge of analysing replication-associated breaks is to establish the location of the replication-stalling entity in the genome. To overcome this problem, a specific type of replication-stalling secondary DNA structure has been studied: G-quadruplexes (G4s). Guanines at G4 motifs assemble into quartets when the DNA becomes single-stranded, which occurs during replication when the two strands are separated. The resulting secondary structure can be unwounded by DOG-1 helicase (FANCJ in humans). In the absence of the DOG-1 helicase, the guanines remain

folded into the G4 structure, which cannot be bypassed by the replication machinery(Cheung, Schertzer et al. 2002, Kruisselbrink, Guryev et al. 2008). The replication machinery restarts downstream of the G4(van Schendel, Romeijn et al. 2021), leaving a single strand gap. Similar to other replication blocking lesions, the G4 is present in the DNA in the next cell cycle, leading to a DSB in the second cell cycle after G4 formation(Lemmens, van Schendel et al. 2015). The *C. elegans* genome is estimated to contain 1680 G4 motifs. Performing a PCR of the DNA surrounding a G4 motif allows the identification of G4-induced deletions.

Analysis of germline using microscopy

One of the advantages of using *C. elegans* as a research model is the knowledge on the location and origin of all cells. One well-defined organ system in *C. elegans* is its germline: the cells leading to the worm's offspring develop here. The germline is often studied to study the repair of meiotic breaks, which are formed to enable genetic recombination in reproductive cells. Because *C. elegans* worms are usually hermaphrodites, the germline contains both oocytes and sperm. The cells leading to mature oocytes are organized in a spatiotemporal fashion: the oocyte precursors are located in the tip of the germline, and mature oocytes are located near the vulva, where the chambers containing the sperm are also located (Figure 3)(Hubbard and Greenstein). The cells thus migrate from the tip to the vulva, and all the stages between the precursor cells and mature oocytes can be distinguished using microscopy, when the DNA is visualized by staining. Cells coming from the tip enter the transition zone, where the homologous chromosomes pair(Hillers, Jantsch et al.). After pairing, DSBs are induced by SPO-11 while the cells travel through the germline.

Proteins that accumulate at break ends, like RAD-51, are often used as a marker for DSBs. Delayed repair of SPO-11 induced breaks could lead to nuclei still containing breaks further in development, visible as increased RAD-51 staining in areas further from the tip compared to repair-proficient controls. A downside of using RAD-51 as a proxy for DSB repair, is that when loading of RAD-51 on DSBs is impaired, these DSBs are not visible and the number of DSBs will be underestimated.

Mutation accumulation lines

Because of the development of next generation sequencing technology, genomes of different DNA repair defective worms can now be sequenced to identify the consequences of the DNA repair enzymes on mutational footprints(Meier and Gartner 2014). To collect sufficient mutational footprints, worms are grown over multiple generations. This is feasible in *C. elegans,* because a new generation arises approximately every three days. In order to calculate the mutation frequency for each *C. elegans* strain, worms from each generation are transferred to a new plate. We call the first worm and its progeny a mutation accumulation line(Keightley and

Caballero 1997). In our lab, after 50 generations, the DNA of the worms is sequenced to identify the mutations that have arisen. A downside of this method is its time-consuming nature: one mutation accumulation experiment typically takes 6 months, excluding the time for whole genome sequencing and bioinformatic analysis, for which a skilled bioinformatician is essential. Moreover, because mutations can have a negative result on the fitness of the worms, selection takes place during this experiment, leading to an underestimation of the mutation frequency. A major advantage of this method is the ability to investigate mutational footprints independent of sequence context. Another benefit is that DNA damage is not artificially inflicted. The exact origin of the damage leading to a mutation is however unknown: when a deletion is formed, it could be a result e.g. from a DSB induced by SPO-11, or for example via replication stress.

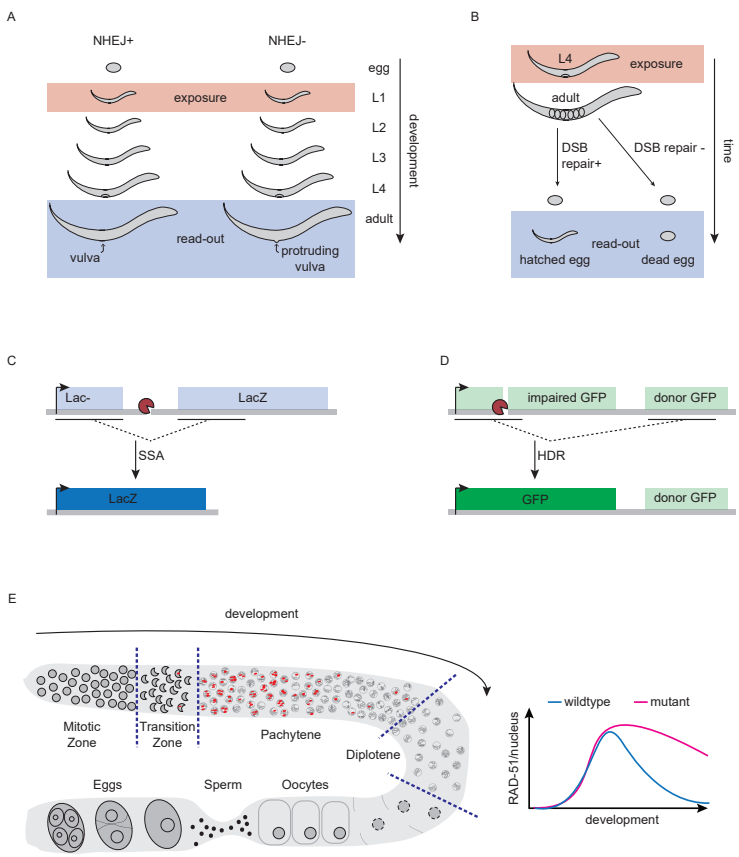

Figure 3 Methods of DNA repair research in *C. elegans* A. Graphic illustration of an L1 assay. Worms in the first larval stage are exposed to a DNA damaging agent or irradiation. At the adult stage, development of the vulva is scored as a read-out for DNA damage sensitivity. B. Graphic illustration of an L4 assay.

Worms in the fourth larval stage are exposed to a DNA damaging agent or irradiation. The eggs laid by the exposed worms are scored for their ability to hatch. The fraction of alive versus dead eggs is used as a read-out for DNA damage sensitivity. C. Schematic representation of the single-strand annealing (SSA) reporter(Pontier and Tijsterman 2009). The I-SceI endonuclease introduces a break at the I-SceI recognition site. Repair of this break by annealing of ~250 bp of identical DNA sequences up- and downstream of the I-SceI recognition site leads to a functioning LacZ open reading frame. D. Schematic representation of the homology-directed repair (HDR) reporter(Johnson, Lemmens et al. 2013). The I-SceI endonuclease introduces a break at the I-SceI recognition site, which is placed in a corrupted GFP sequence. Repair of this break by invasion and copying of the donor GFP sequence located downstream of the corrupted GFP sequence leads to a restoration of the GFP open reading frame. E. Graphic representation of one arm of the *C. elegans* germline (left). The different zones are indicated and separated by dotted lines, RAD-51 foci are represented by red dots. On the right, a hypothetical example of a RAD-51 staining experiment is depicted: the number of RAD-51 foci peak at the moment of break induction. In repair-proficient (wildtype) worms, the number of RAD-51 foci per nucleus rapidly decreases during development. In worms in which repair is impaired (mutant), the number of RAD-51 per nucleus decreases slower than wild-type worms.

This thesis

In this thesis, I will describe my approach to identify factors influencing DSB repair outcomes using *C. elegans*. We investigate repair of DSBs that occur spontaneously or that we artificially introduce. I aimed to establish how DSB context affects repair outcomes and to identify novel proteins affecting DSB repair. I will show how I made use of the unique features and methods of *C. elegans* to study DSB repair in a way that is not possible in other common model organisms and cells, leading to insight into the interplay of different proteins that affect DSB repair.

In Chapter 2, we describe how HR impairment results in three different classes of mutations that are also observed in HR-deficient tumor cells: single nucleotide variants, small deletions with overrepresentation of micro-homology at the junction and tandem duplications. We subsequently demonstrate that the structural variations are the result of TMEJ, and not NHEJ. Moreover, I propose a model explaining the emergence of structural variations in BRCA1-deficient contexts.

In Chapter 3, by using validated transgenic DSB-reporter animals and mutational footprint analysis at G-quadruplex and CRISPR-induced DSBs, we show that HELQ-1 is necessary for multiple DSB-repair mechanisms that are guided by annealing of extensive stretches of complementary bases at break ends. Additionally, we show that helicase Q prevents tandem duplications.

In Chapter 4, we describe how we identified RNA-processing proteins that influence NHEJ efficiency by performing unbiased forward genetics screens using worms that carry a novel dual NHEJ reporter system in their genomes.

In Chapter 5, I discuss the findings described in my thesis and put them in the context of the current DNA DSB repair field.

REFERENCES

Boulton, S. J. and S. P. Jackson (1996). "Identification of a Saccharomyces cerevisiae Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance." Nucleic Acids Res **24**(23): 4639-4648.

Boyd, J. B., K. Sakaguchi and P. V. Harris (1990). "mus308 mutants of Drosophila exhibit hypersensitivity to DNA cross-linking agents and are defective in a deoxyribonuclease." Genetics **125**(4): 813-819.

Brenner, S. (1974). "The genetics of Caenorhabditis elegans." Genetics **77**(1): 71-94.

Chan, S. H., A. M. Yu and M. McVey (2010). "Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila." PLoS Genet **6**(7): e1001005.

Chang, H. H. Y., N. R. Pannunzio, N. Adachi and M. R. Lieber (2017). "Non-homologous DNA end joining and alternative pathways to double-strand break repair." Nat Rev Mol Cell Biol **18**(8): 495-506.

Chen, L., C. J. Nievera, A. Y. Lee and X. Wu (2008). "Cell cycle-dependent complex formation of BRCA1. CtIP.MRN is important for DNA double-strand break repair." J Biol Chem **283**(12): 7713-7720.

Cheung, I., M. Schertzer, A. Rose and P. M. Lansdorp (2002). "Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA." Nat Genet **31**(4): 405-409.

Clejan, I., J. Boerckel and S. Ahmed (2006). "Developmental modulation of nonhomologous end joining in Caenorhabditis elegans." Genetics **173**(3): 1301-1317.

Consortium, C. e. S. (1998). "Genome sequence of the nematode C. elegans: a platform for investigating biology." Science **282**(5396): 2012-2018.

Cruz-Garcia, A., A. Lopez-Saavedra and P. Huertas (2014). "BRCA1 accelerates CtIP-mediated DNA-end resection." Cell Rep **9**(2): 451-459.

Dudley, D. D., J. Chaudhuri, C. H. Bassing and F. W. Alt (2005). "Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences." Adv Immunol **86**: 43-112.

Hillers, K. J., V. Jantsch, E. Martinez-Perez and J. L. Yanowitz Meiosis. WormBook. T. C. e. R. Community, WormBook.

Hubbard, E. J. A. and D. Greenstein Introduction to the germ line. WormBook. T. C. e. R. Community, WormBook.

Inagaki, S., T. Suzuki, M. A. Ohto, H. Urawa, T. Horiuchi, K. Nakamura and A. Morikami (2006). "Arabidopsis TEBICHI, with helicase and DNA polymerase domains, is required for regulated cell division and differentiation in meristems." Plant Cell **18**(4): 879-892.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science **337**(6096): 816-821.

Johnson, N. M., B. B. Lemmens and M. Tijsterman (2013). "A role for the malignant brain tumour (MBT) domain protein LIN-61 in DNA double-strand break repair by homologous recombination." PLoS Genet **9**(3): e1003339.

Kabotyanski, E. B., L. Gomelsky, J. O. Han, T. D. Stamato and D. B. Roth (1998). "Double-strand break repair in Ku86- and XRCC4-deficient cells." Nucleic Acids Res **26**(23): 5333-5342.

Keightley, P. D. and A. Caballero (1997). "Genomic mutation rates for lifetime reproductive output and lifespan in Caenorhabditis elegans." Proc Natl Acad Sci U S A **94**(8): 3823-3827.

Kelso, A. A., F. W. Lopezcolorado, R. Bhargava and J. M. Stark (2019). "Distinct roles of RAD52 and POLQ in chromosomal break repair and replication stress response." PLoS Genet **15**(8): e1008319.

Kent, T., G. Chandramouly, S. M. McDevitt, A. Y. Ozdemir and R. T. Pomerantz (2015). "Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ." Nat Struct Mol Biol **22**(3): 230-237.

Kim, H. M. and M. P. Colaiácovo (2015). "DNA Damage Sensitivity Assays in Caenorhabditis elegans." Bio Protoc **5**(11).

Koole, W., R. van Schendel, A. E. Karambelas, J. T. van Heteren, K. L. Okihara and M. Tijsterman (2014). "A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites." Nat Commun **5**: 3216.

Krasner, D. S., J. M. Daley, P. Sung and H. Niu (2015). "Interplay between Ku and Replication Protein A in the Restriction of Exo1-mediated DNA Break End Resection." J Biol Chem **290**(30): 18806-18816.

Kruisselbrink, E., V. Guryev, K. Brouwer, D. B. Pontier, E. Cuppen and M. Tijsterman (2008). "Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans." Curr Biol **18**(12): 900-905.

Lam, I. and S. Keeney (2014). "Mechanism and regulation of meiotic recombination initiation." Cold Spring Harb Perspect Biol **7**(1): a016634.

Lemmens, B., R. van Schendel and M. Tijsterman (2015). "Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers." Nat Commun **6**: 8909.

Lemmens, B. B. and M. Tijsterman (2011). "DNA double-strand break repair in Caenorhabditis elegans." Chromosoma **120**(1): 1-21.

Li, Q. and J. Engebrecht (2021). "BRCA1 and BRCA2 Tumor Suppressor Function in Meiosis." Front Cell Dev Biol **9**: 668309.

Ma, J. L., E. M. Kim, J. E. Haber and S. E. Lee (2003). "Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences." Mol Cell Biol **23**(23): 8820-8828.

McVey, M. and S. E. Lee (2008). "MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings." Trends Genet **24**(11): 529-538.

Meier, B. and A. Gartner (2014). "Having a direct look: analysis of DNA damage and repair mechanisms by next generation sequencing." Exp Cell Res **329**(1): 35-41.

Moore, D. S. (2017). "Behavioral epigenetics." Wiley Interdiscip Rev Syst Biol Med **9**(1).

Muzzini, D. M., P. Plevani, S. J. Boulton, G. Cassata and F. Marini (2008). "Caenorhabditis elegans POLQ-1 and HEL-308 function in two distinct DNA interstrand cross-link repair pathways." DNA Repair (Amst) **7**(6): 941-950.

Plasterk, R. H. (1991). "The origin of footprints of the Tc1 transposon of Caenorhabditis elegans." Embo j **10**(7): 1919-1925.

Plessis, A., A. Perrin, J. E. Haber and B. Dujon (1992). "Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus." Genetics **130**(3): 451-460.

Pontier, D. B. and M. Tijsterman (2009). "A robust network of double-strand break repair pathways governs genome integrity during C. elegans development." Curr Biol **19**(16): 1384-1388.

Robert, V. and J. L. Bessereau (2007). "Targeted engineering of the Caenorhabditis elegans genome following Mos1-triggered chromosomal breaks." Embo j **26**(1): 170-183.

Robert, V. J., M. W. Davis, E. M. Jorgensen and J. L. Bessereau (2008). "Gene conversion and end-joining-repair double-strand breaks in the Caenorhabditis elegans germline." Genetics **180**(1): 673-679.

Roerink, S. F., R. van Schendel and M. Tijsterman (2014). "Polymerase theta-mediated end joining of replication-associated DNA breaks in C. elegans." Genome Res **24**(6): 954-962.

Rouet, P., F. Smih and M. Jasin (1994). "Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease." Mol Cell Biol **14**(12): 8096-8106.

San-Segundo, P. A. and A. Clemente-Blanco (2020). "Resolvases, Dissolvases, and Helicases in Homologous Recombination: Clearing the Road for Chromosome Segregation." Genes (Basel) **11**(1).

Schimmel, J., H. Kool, R. van Schendel and M. Tijsterman (2017). "Mutational signatures of non-homologous and polymerase theta-mediated end-joining in embryonic stem cells." Embo j **36**(24): 3634-3649.

Schimmel, J., N. Muñoz-Subirana, H. Kool, R. van Schendel and M. Tijsterman (2021). "Small tandem DNA duplications result from CST-guided Pol α-primase action at DNA break termini." Nat Commun **12**(1): 4843.

Schimmel, J., R. van Schendel, J. T. den Dunnen and M. Tijsterman (2019). "Templated Insertions: A Smoking Gun for Polymerase Theta-Mediated End Joining." Trends Genet **35**(9): 632-644.

Setiaputra, D. and D. Durocher (2019). "Shieldin - the protector of DNA ends." EMBO Rep **20**(5).

Sharma, S., S. M. Javadekar, M. Pandey, M. Srivastava, R. Kumari and S. C. Raghavan (2015). "Homology and enzymatic requirements of microhomology-dependent alternative end joining." Cell Death Dis **6**(3): e1697.

Sijen, T. and R. H. Plasterk (2003). "Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi." Nature **426**(6964): 310-314.

van Bostelen, I., R. van Schendel, R. Romeijn and M. Tijsterman (2020). "Translesion synthesis polymerases are dispensable for C. elegans reproduction but suppress genome scarring by polymerase theta-mediated end joining." PLoS Genet **16**(4): e1008759.

van Kregten, M., S. de Pater, R. Romeijn, R. van Schendel, P. J. Hooykaas and M. Tijsterman (2016). "T-DNA integration in plants results from polymerase-θ-mediated DNA repair." Nat Plants **2**(11): 16164.

van Schendel, R., S. F. Roerink, V. Portegijs, S. van den Heuvel and M. Tijsterman (2015). "Polymerase Theta is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis." Nat Commun **6**: 7394.

van Schendel, R., R. Romeijn, H. Buijs and M. Tijsterman (2021). "Preservation of lagging strand integrity at sites of stalled replication by Pol α-primase and 9-1-1 complex." Sci Adv **7**(21).

van Schendel, R., J. van Heteren, R. Welten and M. Tijsterman (2016). "Genomic Scars Generated by Polymerase Theta Reveal the Versatile Mechanism of Alternative End-Joining." PLoS Genet **12**(10): e1006368.

Volkova, N. V., B. Meier, V. González-Huici, S. Bertolini, S. Gonzalez, H. Vöhringer, F. Abascal, I. Martincorena, P. J. Campbell, A. Gartner and M. Gerstung (2020). "Mutational signatures are jointly shaped by DNA damage and repair." Nat Commun **11**(1): 2169.