6 GENERAL DISCUSSION

C. ELEGANS AS A MODEL TO MONITOR GENOME EVOLUTION

The research described in this thesis started with exploring the possibility of the use of the model organism *C. elegans* to monitor mutagenesis by large scale sequencing approaches. The compact genome of *C. elegans* (about 100 Mbp) and its rapid growth properties make this model excellently suitable for analysis by whole genome sequencing. The advantage of this approach is that mutagenesis can be studied in an unbiased fashion because there is no selection for a specific mutational target.
While a similar strategy has been exploited by others to study mutation accumulation in wildtype *C. elegans* strains (Denver *et al.* 2009; Weber *et al.* 2010), we applied this method for the first time to strains that have a specific defect in replicational bypass of lesions by creating knock-out alleles for the three Y-family polymerase members that are present in the worm and monitoring their genomes over time. In chapter 3 we evaluated the effect of knocking out *polh-1* and/or *polk-1*, while chapter 5 describes the effect of a *rev-1(null)* mutation.
In order to study the full spectrum of mutations that are being induced, various bio-informatic analysis tools have been evaluated and further developed in the course of this study (chapters 3 and 5). During recent years, *C. elegans*-specific software packages have been developed for fast and straightforward analysis of whole genome sequence data that were focussed towards rapid identification of mutants (Sarin *et al.* 2008; 2010). However, for our in-depth analysis of mutation spectra, we benefited most from combining standard alignment tools with Pindel - an algorithm that searches for genomic rearrangements (Ye *et al.* 2009).
As sequencing techniques are becoming more widely available and less expensive, this newly developed sequencing and analysis pipeline will be highly valuable in evaluating other genome defence mechanisms. Currently, other mutants in key DNA repair processes are being subjected to similar analysis in our laboratory.

ERROR-PRONE POLYMERASES IN GENOME PROTECTION

It remains an intriguing question how error-prone polymerases contribute to genome stability, while they do result in mutation induction at the same time. It is proposed that error-prone polymerases prevent gross chromosomal instability by preventing prolonged replication fork stalling and eventually replication fork collapse at sites where a replication fork hits unrepaired base damage (Figure 1, Knobel and Marti 2011).
Indeed, the absence of TLS polymerases results in micronuclei induction and γ-H2Ax phosphorylation in UV-irradiated mammalian cells, both indicators of induction of double strand breaks (DSBs) that may result from replication fork collapse (Temviriyanukul *et al.* 2012). Nevertheless, it is unknown to which extent this mechanism operates to

maintain genomic stability in the absence of exogenous stressors as UV. A single study reports similar phenotypes (micronuclei induction and γ-H2Ax phosphorylation) in non-exposed Pol η knockdown cells due to defective replication of common fragile sites (Rey *et al.* 2009).

The research described in this thesis clearly demonstrates the induction of chromosomal aberrations in the absence of TLS polymerases, under normal growth conditions (chapter 3). In the absence of functional TLS, spontaneous deletions ranging from 50 - 300 basepairs are induced that were not observed in wildtype strains. In REV1 deficient strains these deletions occurred about once in every ten generations and in Pol η deficient strains about once in every three generations. Pol κ knockout on its own did not have any effect on mutagenesis, but combined knockout of Pol η and Pol κ lead to highly increased deletion induction - about two deletions per generation - that caused an easily recognizable mutator phenotype for this strain. Nevertheless, these conclusions may be difficult to derive from systems with a lower resolution analysis, stressing the strength of the fast growing and reproducing model organism *C. elegans* for these assays. From our mutation profiles, we conclude that under normal growth conditions Y-family polymerases play both redundant and non-redundant roles in bypass of endogenous lesion sites. Importantly, TLS activity on endogenous lesions appears to be mostly error-free, since wildtype substitution rates are much lower than deletion rates in specific polymerase knockouts.

It remains an intriguing question which endogenous lesions are the exact substrates for the different Y-family members. A bias towards deletion induction on dC residues in a *polh-1polk-1* defective background hints that replication fork stalling in this background

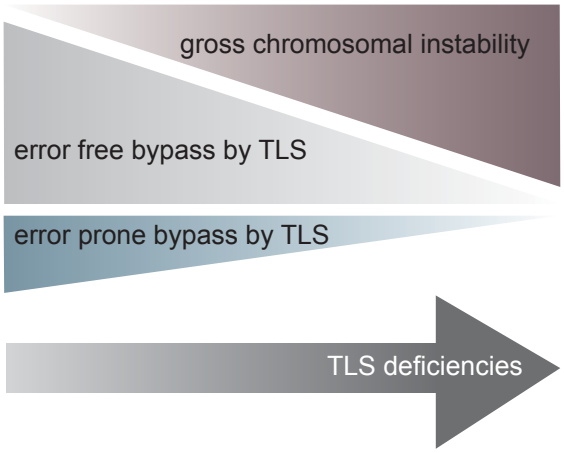

FIGURE 1. Trade off between mutation induction by TLS polymerases and chromosomal instability in the absence of TLS (adapted from (Knobel and Marti 2011).

occurs more often opposite dG residues (Figure 4, Chapter 3). The most abundant oxidative lesions are 8-oxo-dGs, responsible for the majority of transversion mutations (Barnes and Lindahl 2004). Bypass activity of yeast and mammalian Pol η on these lesions has been established in *in vitro* assays (Haracska *et al.* 2000) and may account for a large proportion of mutagenic events.

However, sensitivity assays in chapter 2 show a role of Pol η in tolerance to very different sources of DNA damage, suggesting a broad role for Pol η in damage bypass that may reflect bypass of various structurally different endogenous lesion sites. The role of Pol κ appears to be much more restricted, as *polk-1* mutants are hypersensitive to alkylating damage by methyl methane sulfonate but not to other sources of DNA damage. Nevertheless, additional knockout of Pol κ in Pol η - deficient worms increases the number of deletions under normal growth conditions about ten fold, suggesting that a large subset of lesions is bypassed by Pol κ in the absence of Pol η (Figure 1, chapter 3). Notably, this bypass is most likely error-free, as substitution rates are not increased in *polh-1* mutants. Additional knockout of Pol κ also increases cytotoxicity by MMS in Pol η - defective strains by at least a magnitude of ten, hinting at a similar redundant role for Pol κ in the absence of Pol η in bypassing alkylating damage (Chapter 2). MMS may directly methylate bases, leading to N7-methylguanines, N3-methylguanines and O^6-methylguanines but may also lead to more complex base adducts via reactions with proteins or metabolites in the cell (Fu *et al.* 2012). Endogenous byproducts of oxidation may result in a similar range of base adducts.

Chapter 2 also sheds more light on the strict regulation of error-prone polymerases during development: we show that TLS is specifically important during early embryogenesis in *C. elegans*, when embryos endure a series of fast cell divisions. Exposure to different sources of damaging agents activates a checkpoint in the first embryonic divisions in TLS mutants, resulting in embryonic lethality. Kim et al. identified the sumoylation factor GEI-17 as an important regulator in the activity of TLS protein Pol η (Kim and Michael 2008). Our genome-wide RNAi screen identified genetic interactions of both Pol η and Pol κ with GEI-17. In addition, we show involvement of three new factors: the sumo protease ULP-1, and two nuclear pore proteins (NPP-2 and NPP-22), that are essential for TLS-dependent damage bypass in the early embryo. Together these data suggest that sumoylation processes are essential for regulation of TLS polymerases after exposure to DNA damage in the early embryo. A recent paper by Mosbech and coauthors also describes a central role for the ubiquitin responsive protein p97 and its adaptor protein DVC-1 in regulation of Pol η in *C. elegans* (Mosbech *et al.* 2012).

POL THETA-MEDIATED END JOINING (TMEJ)

One of the most striking observations that we did during analysis of mutation profiles of Y-family polymerase - deficient mutants was the very distinct character of the

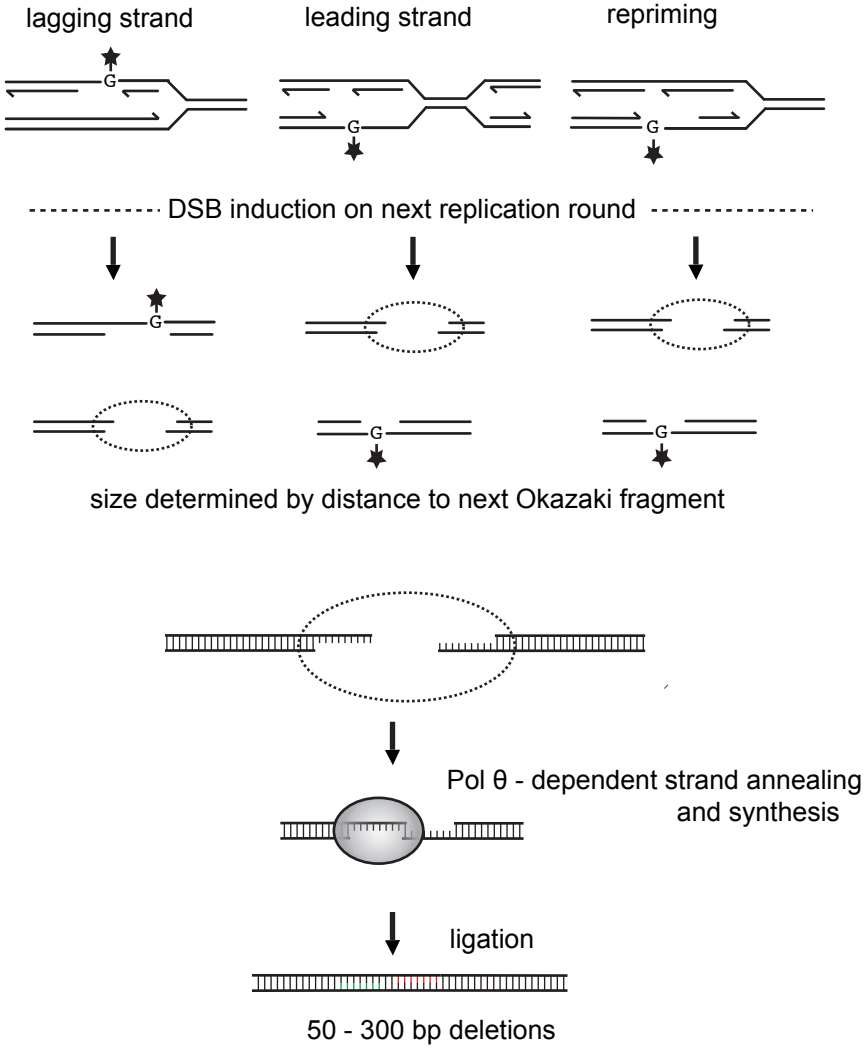

FIGURE 2. Different models explaining TMEJ - mediated deletions in absence of TLS polymerases.
After replication fork stalling at the damaged base, either a gap is formed spanning the sequence from the stall to the next Okazaki fragment, or the damaged base is looped out. Such a ssDNA gap will generate a DSB in the next round of replication, which is subsequently repaired in a Pol θ - dependent manner.

chromosomal aberrations in these strains. In *polh-1*, *polh-1polk-1* and *rev-1* mutants, we saw occurrence of a class of 50-300 bp deletions that were not seen in wildtype control strains and suggested error-prone repair of collapsed replication forks (chapters 3 and 5).

In an unrelated study - described in chapter 4 - we identified a new mechanism for germline repair of transposition-induced DSBs.

Detailed inspection of repair footprints of Tc1-induced breaks and *polh-1polk-1*-derived deletions revealed two shared characteristics in both assays: i) bias towards at least one nucleotide of microhomology and ii) insertion of fragments up to ~20 nucleotides duplicated from the flanks. A third assay, described by W. Koole et al. (submitted for publication) displays similar outcomes in FANCJ/*dog-1* mutants at genomic sequences that can fold in G-quadruplex structures. We show that all of these repair outcomes are the result of the action of the A-family Polymerase θ, hereby describing a novel repair pathway operating in the *C. elegans* germline, termed Pol θ - mediated end joining (TMEJ). This mechanism is based on one or more cycles of annealing of both flanking sites based on at least a single basepair homology and subsequent primer extension, followed by ligation.

At present, we do not know the determinants for the choice for TMEJ repair of DSBs in the *C. elegans* germline as opposed to the two canonical routes non-homologous end joining (NHEJ) and homologous repair (HR). Studies from our lab and others showed HR as the preferred pathway in the germline, while NHEJ is operating in somatic cells (Clejan *et al.* 2006; Pontier and Tijsterman 2009). However, HR depends on the availability of a sister chromatid as a repair template and may operate only in S-phase, when DNA has already replicated.

TMEJ at replication-blocking lesions in *polh-1polk-1*-defective backgrounds and at G4 stretches in *dog-1* mutants suggests that the mechanism is linked to replication. Besides, the distinct size of the deletions may give clues about the molecular mechanism: different scenarios can be envisioned that result in DSBs where the ends are approximately 50-300 bps apart and would thus result in deletions with such a distinct size (Figure 2). Notably, the average distance between Okazaki fragments in the lagging strand in eukaryotes is in the range of 100 - 200 bps (Abdurashidova *et al.* 2000; Smith and Whitehouse 2012), offering an attractive explanation for the deletion length as the gap length determined by the distance of the stalled fork to the next Okazaki fragment. Alternatively, repriming may occur downstream of the lesion site. In both scenarios a single stranded gap may be converted into a DSB intermediate in the next round of replication. The analogy to repair of Tc1-induced breaks by Pol θ strongly suggests that the repair of stalled forks involves a DSB intermediate. However, we cannot exclude an alternative scenario for replication-blocking lesions that does not require a DSB intermediate but is explained by direct priming on the same strand after loop-out of the damaged base. In this scenario, deletion size may be governed by the size of the loop.

FUTURE PERSPECTIVES

One of the most prominent questions raised by the research described in this thesis is to what extent the described error-prone TMEJ mechanism is active in higher eukaryotes under physiological or cancerous conditions. In this thesis, we show that for different sources of replicational stress or DSBs, TMEJ dominates repair in the *C. elegans* germline. Possibly, this repair mode is developmentally strictly regulated and therefore missed in many previous studies on break repair in mammals. Expression analysis showed only low levels of human and mice Pol θ in somatic tissues other than testes (Seki *et al.* 2003).

Nevertheless, Pol θ-dependent footprints bear similarity to certain deletion or translocation events identified in cancer genomes ((Roth *et al.* 1985; Welzel *et al.* 2001; Murga Penas *et al.* 2010), chapter 3), which could argue that cancer cells have acquired the capacity to deal with replication stress by employing Pol θ. Indeed, Pol θ expression was found upregulated in several cancers including breast cancer and colorectal cancer, and higher Pol θ levels are correlating with a reduced prognosis (Kawamura *et al.* 2004; Pillaire *et al.* 2009; Lemée *et al.* 2010).

The work described in this thesis, revealing an unanticipated role for Pol θ in the maintenance of genome stability in *C. elegans*, strongly encourages an in-depth analysis of the function of this protein under physiological and cancerous conditions in humans, as this may provide a new target for therapy.

REFERENCES

Abdurashidova G., Deganuto M., Klima R., Riva S., Biamonti G., Giacca M., Falaschi A., 2000 Start sites of bidirectional DNA synthesis at the human lamin B2 origin. Science **287**: 2023–2026.

Barnes D. E., Lindahl T., 2004 Repair and genetic consequences of endogenous DNA base damage in mammalian cells. Annu Rev Genet **38**: 445–476.

Clejan I., Boerckel J., Ahmed S., 2006 Developmental modulation of nonhomologous end joining in Caenorhabditis elegans. Genetics **173**: 1301–1317.

Denver D. R., Dolan P. C., Wilhelm L. J., Sung W., Lucas-Lledó J. I., Howe D. K., Lewis S. C., Okamoto K., Thomas W. K., Lynch M., Baer C. F., 2009 A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. Proc Natl Acad Sci USA **106**: 16310–16314.

Fu D., Calvo J. A., Samson L. D., 2012 Balancing repair and tolerance of DNA damage caused by alkylating agents. Nat Rev Cancer **12**: 104–120.

Haracska L., Yu S. L., Johnson R. E., Prakash L., Prakash S., 2000 Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase eta. Nat Genet **25**: 458–461.

Kawamura K., Bahar R., Seimiya M., Chiyo M., Wada A., Okada S., Hatano M., Tokuhisa T., Kimura H., Watanabe S., Honda I., Sakiyama S., Tagawa M., O-Wang J., 2004 DNA polymerase theta is preferentially expressed in lymphoid

tissues and upregulated in human cancers. Int. J. Cancer **109**: 9–16.

Kim S.-H., Michael W. M., 2008 Regulated proteolysis of DNA polymerase eta during the DNA-damage response in C. elegans. Mol Cell **32**: 757–766.

Knobel P. A., Marti T. M., 2011 Translesion DNA synthesis in the context of cancer research. Cancer Cell Int. **11**: 39.

Lemée F., Bergoglio V., Fernandez-Vidal A., Machado-Silva A., Pillaire M.-J., Bieth A., Gentil C., Baker L., Martin A.-L., Leduc C., Lam E., Magdeleine E., Filleron T., Oumouhou N., Kaina B., Seki M., Grimal F., Lacroix-Triki M., Thompson A., Roché H., Bourdon J.-C., Wood R. D., Hoffmann J.-S., Cazaux C., 2010 DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. Proceedings of the National Academy of Sciences **107**: 13390–13395.

Mosbech A., Gibbs-Seymour I., Kagias K., Thorslund T., Beli P., Povlsen L., Nielsen S. V., Smedegaard S., Sedgwick G., Lukas C., Hartmann-Petersen R., Lukas J., Choudhary C., Pocock R., Bekker-Jensen S., Mailand N., 2012 DVC1 (C1orf124) is a DNA damage-targeting p97 adaptor that promotes ubiquitin-dependent responses to replication blocks. Nat Struct Mol Biol **19**: 1084–1092.

Murga Penas E. M., Callet-Bauchu E., Ye H., Gazzo S., Berger F., Schilling G., Albert-Konetzny N., Vettorazzi E., Salles G., Wlodarska I., Du M.-Q., Bokemeyer C., Dierlamm J., 2010 The t(14;18)(q32;q21)/IGH-MALT1 translocation in MALT lymphomas contains templated nucleotide insertions and a major breakpoint region similar to follicular and mantle cell lymphoma. Blood **115**: 2214–2219.

Pillaire M. J., Selves J., Gordien K., Gouraud P.-A., Gentil C., Danjoux M., Do C., Negre V., Bieth A., Guimbaud R., Trouche D., Pasero P., Méchali M., Hoffmann J.-S., Cazaux C., 2009 A "DNA replication" signature of progression and negative outcome in colorectal cancer. Oncogene **29**: 876–887.

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

Rey L., Sidorova J. M., Puget N., Boudsocq F., Biard D. S. F., Monnat R. J., Cazaux C., Hoffmann J.-S., 2009 Human DNA Polymerase Is Required for Common Fragile Site Stability during Unperturbed DNA Replication. Mol Cell Biol **29**: 3344–3354.

Roth D. B., Porter T. N., Wilson J. H., 1985 Mechanisms of nonhomologous recombination in mammalian cells. Mol Cell Biol **5**: 2599–2607.

Sarin S., Bertrand V., Bigelow H., Boyanov A., Doitsidou M., Poole R. J., Narula S., Hobert O., 2010 Analysis of multiple ethyl methanesulfonate-mutagenized Caenorhabditis elegans strains by whole-genome sequencing. Genetics **185**: 417–430.

Sarin S., Prabhu S., O'Meara M. M., Pe'er I., Hobert O., 2008 Caenorhabditis elegans mutant allele identification by whole-genome sequencing. Nat Methods **5**: 865–867.

Seki M., Marini F., Wood R. D., 2003 POLQ (Pol theta), a DNA polymerase and DNA-dependent ATPase in human cells. Nucleic Acids Research **31**: 6117–6126.

Smith D. J., Whitehouse I., 2012 Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature **483**: 434–438.

Temviriyanukul P., van Hees-Stuivenberg S., Delbos F., Jacobs H., de Wind N., Jansen J. G., 2012 Temporally distinct translesion synthesis pathways for ultraviolet light-induced photoproducts in the mammalian genome. DNA Repair **11**: 550–558.

Weber K. P., De S., Kozarewa I., Turner D. J., Babu M. M., de Bono M., 2010 Whole genome sequencing highlights genetic changes associated with laboratory domestication of C. elegans. PLoS ONE **5**: e13922.

Welzel N., Le T., Marculescu R., Mitterbauer G., Chott A., Pott C., Kneba M., Du M. Q., Kusec R., Drach J., Raderer M., Mannhalter C., Lechner K., Nadel B., Jaeger U., 2001 Templated nucleotide addition and immunoglobulin JH-gene utilization in t(11;14) junctions: implications for the mechanism of translocation and the origin of mantle cell lymphoma. Cancer Res **61**: 1629–1636.

Ye K., Schulz M. H., Long Q., Apweiler R., Ning Z., 2009 Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics **25**: 2865–2871.