Josip Juraj Strossmayer University of Osijek, Croatia

University of Dubrovnik, Croatia

Ruđer Bošković Institute, Zagreb, Croatia

University Postgraduate Interdisciplinary Doctoral Study

Molecular Biosciences

Christine Supina Pavić

THE ROLE OF ACRC PROTEASE AND NUCLEOTIDE EXCISION REPAIR PATHWAY IN THE REPAIR OF DNA-PROTEIN CROSSLINKS (DPCs)

DOCTORAL THESIS

Osijek, 2023.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište Josipa Jurja Strossmayera u Osijeku **Doktorska disertacija**
Sveučilište u Dubrovniku
Institut Ruđer Bošković
Poslijediplomski interdisciplinarni sveučilišni studij Molekularne bioznanosti

Znanstveno područje: Prirodne znanosti
Znanstveno polje: Biologija

Uloga proteaze ACRC i puta popravka izrezivanjem nukleotida u popravku unakrsnih veza DNA i proteina

Christine Supina Pavić

Rad je izrađen u: Institut Ruđer Bošković, Zagreb
Mentor: dr. sc. Marta Popović, znanstveni suradnik

Kratki sažetak doktorskog rada:
Unakrsna veza DNA i proteina (engl. *DNA-protein crosslink*, DPC) je vrsta DNA oštećenja koja nastaje ireverzibilnim kovalentnim vezanjem proteina za DNA. DPC-evi su česta DNA oštećenja koja predstavljaju fizičku prepreku za sve DNA transakcije. Malo je poznato o mehanizmu njihova popravka i faktorima koji sudjeluju u popravku, posebice na razini organizma. Nedavno je nekoliko grupa identificiralo nove stanične proteaze Wss1 i SPRTN, što je dovelo do spoznaje da je DPC popravak zaseban mehanizam popravka DNA. Osnovni cilj ove doktorske disertacije bio je ustanoviti sudjeluje li protein ACRC i put popravka izrezivanjem nukleotida (engl. *Nucleotide Excision Repair Pathway*, NER) u popravku DPC-eva, koristeći ribu zebricu (*Danio rerio*) kao istraživački model i CRISPR/Cas9 gensku manipulaciju. Rezultati su pokazali da NER put popravka sudjeluje u popravku manjih DPC-eva, kao što su unakrsne veze DNA i histona, te upućuju na to da NER djeluje neovisno o SPRTN proteazi. Također smo pokazali da je protein ACRC esencijalan za rani embrionalni razvoj zebrice, da je njegova katalitička funkcija neophodna za popravak DPC-eva, te da je specifično SprT proteazna domena odgovorna za stabilnost i aktivnost proteina.

Broj stranica: 136
Broj slika: 52
Broj tablica: 20
Broj literaturnih navoda: 162
Jezik izvornika: engleski
Ključne riječi: DPC, *Danio rerio*, CRISPR/Cas9, tkivna ekspresija, ACRC, XPA, NER, SPRTN proteaza
Datum obrane: 22.02.2023.

Stručno povjerenstvo za obranu:
1. Neda Slade, dr. sc., znanstvena savjetnica, Institut Ruđer Bošković, predsjednik
2. Tvrtko Smital, dr. sc., znanstveni savjetnik u trajnom zvanju, Institut Ruđer Bošković, član
3. Vera Cesar, dr. sc., redovita profesorica, Sveučilište J. J. Strossmayera u Osijeku, član
4. Andreja Ambriović Ristov, dr. sc., znanstvena savjetnica, Institut Ruđer Bošković, zamjena člana

Rad je pohranjen u: Nacionalnoj i sveučilišnoj knjižnici Zagreb, Ul. Hrvatske bratske zajednice 4, Zagreb; Gradskoj i sveučilišnoj knjižnici Osijek, Europska avenija 24, Osijek; Sveučilištu Josipa Jurja Strossmayera u Osijeku, Trg sv. Trojstva 3, Osijek.

BASIC DOCUMENTATION CARD

Josip Juraj Strossmayer University of Osijek **PhD thesis**
University of Dubrovnik
Ruđer Bošković Institute
University Postgraduate Interdisciplinary Doctoral Study of Molecular biosciences

Scientific Area: Natural sciences
Scientific Field: Biology

The role of ACRC protease and nucleotide excision repair pathway in the repair of DNA-protein crosslinks (DPCs)

Christine Supina Pavić

Thesis performed at: Ruđer Bošković Institute, Zagreb
Supervisor: Marta Popović, PhD, Senior Scientist

Short abstract:
DNA-protein crosslink (DPC) is a type of DNA damage caused by irreversible covalent binding of protein to DNA. DPCs are common DNA damage that pose a physical barrier to all DNA transactions. Not much is known about the mechanism of their repair and the factors involved, especially at the level of the organism. Recently, several groups have identified proteases Wss1 and SPRTN, leading to recognition of the DNA-protein crosslink repair as a separate DNA damage repair pathway. The main goal of this doctoral thesis was to determine whether the ACRC protein and the Nucleotide Excision Repair Pathway (NER) participate in DPC repair, using the zebrafish (*Danio rerio*) as a research model and CRISPR/Cas9 genome editing. The results showed that the NER pathway is involved in the repair of small DPCs, such as DNA- histone crosslinks, and they indicate that NER acts independently of the SPRTN protease. We have also shown that the ACRC protein is essential for the early embryonic development of the zebrafish, that its catalytic function is necessary for the repair of DPCs, and that specifically the SprT protease domain is responsible for protein stability and activity.

Number of pages: 136
Number of figures: 52
Number of tables: 20
Number of references: 162
Original in: English
Key words: DPC, *Danio rerio*, CRISPR/Cas9, tissue expression, ACRC, XPA, NER, SPRTN protease
Date of the thesis defense: 22.02.2023.

Reviewers:
1. Neda Slade, PhD, Senior Scientist, Ruđer Bošković Institute, Zagreb
2. Tvrtko Smital, PhD, Senior Scientist, Ruđer Bošković Institute, Zagreb
3. Professor Vera Cesar, PhD, Josip Juraj Strossmayer University of Osijek
4. Andreja Ambriović Ristov, PhD, Senior Scientist, Ruđer Bošković Institute, Zagreb

Thesis deposited in: National and University Library in Zagreb, Ul. Hrvatske bratske zajednice 4, Zagreb, Croatia; City and University Library of Osijek, Europska avenija 24, Osijek, Croatia; Josip Juraj Strossmayer University of Osijek, Trg sv. Trojstva 3, Osijek, Croatia.

Research for this PhD thesis was performed in the Laboratory for Molecular Ecotoxicology, Division of Marine and Environmental Research, Ruđer Bošković Institute, Zagreb, under the supervision of Marta Popović, PhD. M.P. is supported by Croatian Science Foundation installation grant (UIP-2017- 05-5258), project grant (IPS-2020-01-4225) and by the European Structural and Investment Funds STIM-REI Project (KK.01.1.1.01.0003), while this work was funded by the Croatian Science Foundation installation grant (UIP-2017- 05-5258).

Acknowledgments

First, I would like to express my sincere gratitude to my supervisor Marta Popović for giving me the opportunity to work in her group, for her guidance and transferred knowledge. I am grateful for helpful comments, suggestions and constructive criticism during these four years of my Ph.D study.

Besides my supervisor, I would like to thank our laboratory head Tvrtko Smital and my laboratory colleagues Jelena, Lana, Marin, Ivan A, Cecile, Ivan M, and Jovica for all the help, advices, stimulating discussions and pleasant and motivating working atmosphere.

Most of all, I would like to thank my family: my husband Matija for all the love, encouragement and emotional support over the past 10 years; my parents Ana-Marija and Alwin, my sister Dorotea, Ivan, Marta, Karoline, and my grandparents for their unconditional love and support throughout my education and life in general. Thank you for always standing by me and believing in me.

Last but not least, I thank my friends: Eva for all the support, care and fun times we spent together, and Sonja for being a person you can always count on and for all the help and advices.

TABLE OF CONTENTS

1. Introduction

1.1. DNA damage ……1

1.2. DNA damage repair ……3

1.3. DNA-protein crosslinks (DPCs)……6

1.4. DNA-Protein Crosslink Repair (DPCR)……9

1.4.1. Proteolysis dependent repair……10

1.4.2. Nuclease dependent repair……11

1.4.3. Hidrolysis dependent repair……12

1.5. Proteases in DPCR ……13

1.5.1. Wss1 and SPRTN proteases……14

1.5.2. ACRC/GCNA putative protease……17

1.6. Nucleotide excision repair pathway (NER) and its role in DPCR……19

1.6.1. GG-NER i TC-NER……19

1.6.2. NER in DPCR……22

1.7. CRISPR/Cas9 genome editing……23

1.8. Zebrafish (*Danio rerio*) as a model organism……27

1.9. Summary and aims ……32

2. Materials and Methods

2.1. Materials ……33

2.2. Methods ……38

2.2.1. Phylogenetic and synteny analysis……38

2.2.2. Analysis of gene expression in adult zebrafish tissues and during embryonic development (qPCR)……38

2.2.3. Creation of zebrafish mutants using CRISPR/Cas9 system……39

2.2.4. Genotyping using High-Resolution Melting (HRM) Analysis……40

2.2.5. Genotyping using cloning and sequencing……40

2.2.6. Cell culture……41

2.2.7. Gene silencing by transfection of small interfering RNAs (siRNAs)……42

2.2.8. Morpholino-mediated gene silencing in zebrafish embyros……42

2.2.9. Isolation of DPCs using KCl/SDS precipitation method……43

2.2.10. Isolation and detection of DPCs by RADAR (Rapid Approach to DNA Adduct Recovery) assay……43

2.2.11. Western blot analysis……44
2.2.12. Slot blot analysis……..45
2.2.13. Rescue experiments……45
3. Results
3.1. Phylogenetic analysis of XPA proteins……46
3.2. Syntenic analysis of human and zebrafish *xpa* and *acrc* genes……47
3.3. Expression analysis of *acrc* in zebrafish embryos and adults……51
3.4. Creation of Xpa deficient zebrafish strain……55
3.4.1. Genotyping of Xpa F0 generation……57
3.4.2. Genotyping of Xpa F1 generation……61
3.5. Morpholino-mediated *xpa* silencing in zebrafish embryos……64
3.6. Morpholino-mediated *sprtn* silencing in zebrafish embryos……66
3.7. DPC levels in zebrafish embryos after morpholino-mediated *xpa* and *sprtn* silencing ……..68
3.8. Creation of Acrc mutant strain carrying enzymatic mutation (ΔEMCH)……73
3.8.1. Phenotyping of Acrc mutant strain carrying enzymatic mutation (ΔEMCH)……..74
3.8.2. Genotyping of Acrc mutant strain carrying enzymatic mutation (ΔEMCH) F0 generation……75
3.8.3. Genotyping of Acrc mutant strain carrying enzymatic mutation (ΔEMCH) F1 generation……77
3.8.4. Genotyping of Acrc mutant strain carrying enzymatic mutation (ΔEMCH) F2 generation……80
3.9. Creation of Acrc mutant zebrafish strain carrying C-terminal deletion (ΔC)……82
3.9.1. Phenotyping of Acrc mutant zebrafish strain carrying C-terminal deletion (ΔC)……..83
3.9.2. Genotyping of Acrc mutant zebrafish strain carrying C-terminal deletion (ΔC) F0 generation……83
3.9.3. Genotyping of Acrc mutant zebrafish strain carrying C-terminal deletion (ΔC) F1 generation……87
3.10. Acrc protein levels in zebrafish Acrc mutant lines: ΔEMCH and ΔC……89
3.11. Rescue experiments in embryos carrying C-terminal deletion (ΔC)……..90
3.12. DPC levels in Acrc deficient zebrafish embryos……92
3.13. *XPA, SPRTN* and *ACRC* silencing in cells……94
3.14. Histone 1 (H1) and Histone 3 (H3) DPC levels in XPA deficient RPE1 cells……107
4. Discussion
4.1.Characterization of zebrafish XPA protein ……109
4.2. Creation and optimization of cellular and animal models to study the role of XPA in DPC repair ……109
4.3. The role of XPA in DPC repair ……111

4.4. Characterization of zebrafish ACRC protein ……114
4.5. Creation and optimization of cellular and animal models to study the role of ACRC in DPC repair……116
4.6. The role of ACRC in DPC repair……117
5. Conclusions ……122
6. References ……123
7. Summary ……131
8. Sažetak ……133
9. Curriculum vitae ……135

1. Introduction

1.1. DNA damage

DNA is constantly exposed to a multitude of damaging agents and maintaining the integrity of the genome is important for the continuation of life. It is well known that DNA is a reactive molecule that is significantly exposed to different endogenous and exogenous agents that cause chemical modifications and mutations. Mutagenesis is an indispensible part of evolution, but it also contributes to certain human diseases, aging, and cancer. However, DNA is protected by different DNA repair and DNA damage tolerance mechanisms that allow the lesion to be removed or bypassed, while disrupted DNA repair pathways lead to genome instability (Chatterjee and Walker, 2017). There are several forms of DNA lesions, including base mismatches, apurinic/apyrimidinic (AP) sites (abasic sites), intrastrand and interstrand crosslinks (ICLs), single-strand breaks (SSBs), double-strand breaks (DSBs), and DNA-protein crosslinks (DPCs) (Martin, 2008) (Figure 1).

Multiple DNA repair pathways, including base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), homologous recombination (HR), and non-homologous end joining (NHEJ), enable cells to repair DNA damage (Jeggo et al, 2016). In addition, certain types of DNA lesions are bypassed by translesion DNA synthesis, which is part of the DNA damage tolerance pathway. In higher eukaryotes, the translesion synthesis (TLS) polymerases REV1, POL ζ, POL η, POL κ and POL ι allow replication to continue on damaged DNA substrates. Cells with unrepaired DNA are subjected to apoptosis to avoid widespread genome instability (Chatterjee and Walker, 2017).

DNA damage can be categorized as endogenous or exogenous (Figure 1). Endogenous DNA damage is mainly caused by reactions of DNA with water and reactive species such as reactive oxygen species (ROS) and reactive nitrogen species (RNS) which are formed as by-products of the electron transport chain (ETC) during cellular respiration in aerobic organisms, reduction/oxidation reactions (redox reactions), normal catalytic activity of many peroxisomal enzymes, and as a result of inflammatory signals. However, an excess of reactive species causes oxidative stress leading to oxidative base lesions, modifications of 2-deoxyribose, and damaged proteins and lipids. In addition to causing oxidative DNA damage, ROS radicals can also cause single-strand breaks due to the decay of oxidized sugar or in the repair of oxidized bases by BER.

Endogenously caused mutations can also arise as a result of errors in replication. Human cells contain high fidelity replicative polymerases (δ and ε), but also other DNA polymerases (α, β, σ, γ, λ, REV1, ζ, η, ι, κ, θ, ν, μ, Tdt and PrimPol) that are more error-prone and have lower replication fidelity. Additional errors arise from replication slippage events at repetitive sequences that lead to rearrangements in the genome, such as deletions or insertions.

Another important source of mutations in human cells is base deamination. This is a type of damage in which cytosine (C), adenine (A), guanine (G), and 5-methyl-cytosine (5mC) are converted to uracil (U), hypoxanthine, xanthine and thymine (T), respectively, by the loss of their exocyclic amine. The most commonly deaminated bases are cytosine and 5-methyl-cytosine, which are repaired by uracil-DNA glycosylase and thymine DNA glycosylase (TDG), which are involved in the mismatch repair pathway.

The most common DNA damage, which occurs daily at a rate of 10,000 per cell, are abasic or AP (apurinic/apyrimidic) sites. This type of damage occurs when the glycosyl bond between the nitrogenous base and the sugar phosphate backbone is spontaneously hydrolyzed or cleaved by a DNA glycosylase during the BER pathway. Aldehydes, which can be produced endogenously, can also alter and damage DNA. Processes that generate large numbers of reactive aldehydes include lipid peroxidation, detoxification of xenobiotics in the liver, biotransformation of drug molecules, and metabolism of amino acids, carbohydrates, lipids and vitamins. Lipid peroxidation can produce highly reactive unsaturated aldehydes (acrolein, 4-hydroxy-2-nonenal, malondialdehyde), while formaldehyde is produced under both normal physiological conditions and pathophysiological conditions. Endogenous formaldehyde is mainly formed by oxidative demethylation of histones, RNA and DNA (Nakamura and Nakamura, 2020).

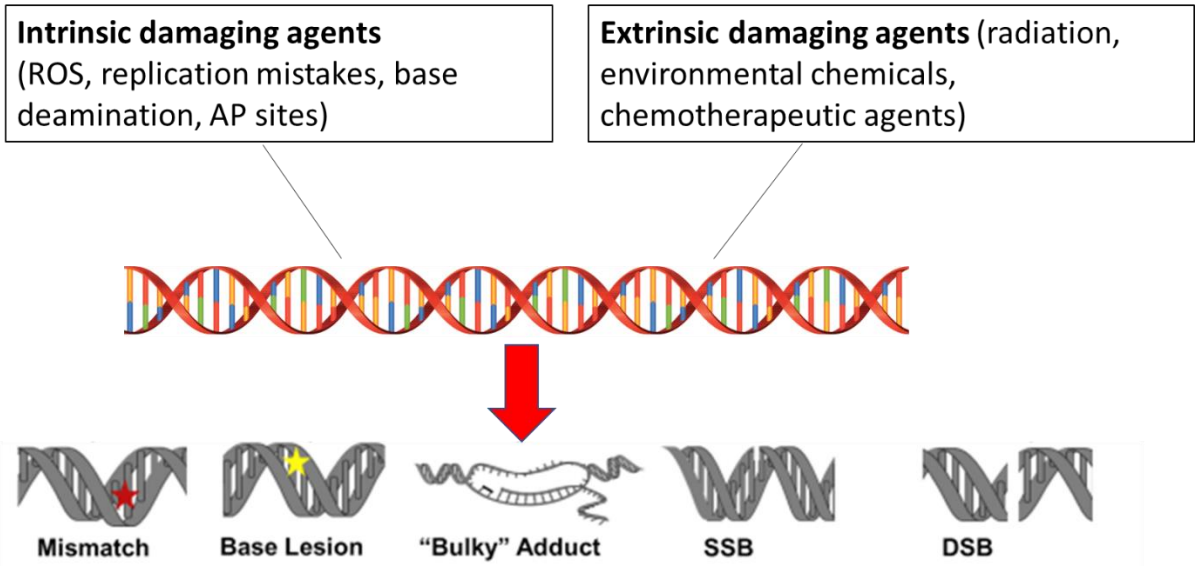


Figure 1. Schematic representation of DNA lesions which can arise from intrinsic or extrinsic agents or events (adapted from Tiwari and Wilson, 2019).

Exogenous DNA damage occurs when DNA is exposed to environmental, physical, or chemical agents (UV light, ionizing radiation (IR), alkylating agents, various drugs) (Chatterjee and Walker, 2017). IR can damage DNA directly by inducing DNA breaks or indirectly by radiolysis of water and generation of hydroxyl radicals (ROS), which in turn oxidize proteins and lipids

and also induce abasic sites and single strand breaks (SSBs). IR causes unique DSBs and SSBs in which DNA breaks bare 3′ phosphate or 3′-phosphoglycolate ends rather than the 3′-OH ends, and can also cause base damages through indirect effects. UV radiation is the main cause of skin cancer in humans and can be divided into the UV-C (190–290 nm), the UV-B (290–320 nm) and the UV-A spectra, depending on the wavelength. The UV-C spectrum is the most dangerous, damaging DNA by producing cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6 – 4) pyrimidone photoproducts ((6 – 4) PPs), two major photoproducts characterized by covalent bonds between two adjacent pyrimidines (Tiwari and Wilson, 2019). Alkylating agents, primarily derived from food ingredients, tobacco smoke, biomass combustion, industrial processing, and chemotherapeutic drugs, add the alkyl group through an SN1 or SN2 substitution reaction, generating adducts that are mutagenic and carcinogenic (Chatterjee and Walker, 2017).

1.2. DNA damage repair

To counteract DNA damage, specific repair mechanisms have evolved (Figure 2). Damaged DNA that is not accurately repaired at the cellular level leads to genomic instability, senescence, and apoptosis, while at the level of the organism it can strongly influence the ageing process and the onset of disease. Cells must respond efficiently to DNA damage to avoid the risk of immunodeficiency, neurological disorders, and cancer that arise as a result of loss of genomic integrity (Hakem, 2008).

After DNA is damaged, sensing proteins that detect the lesion initiate a DNA damage response (DDR) pathway that detects the damage, signals its presence, and initiates subsequent repair (Harper and Elledge, 2007). The DDR is mediated primarily by kinases (ataxia-telangiectasia mutated (ATM), ataxia telangiectasia and Rad3-related (ATR), DNA-dependent protein kinase (DNA-PK)), which phosphorylate various targets, and a family of poly(ADP)ribose polymerases (PARP) consisting of 16 members. The kinases ATM and DNA-PK are activated by DSBs, the ATR kinase is activated by both DSBs and single- stranded DNA regions arising at stalled replication forks, whereas the polymerases PARP1 and PARP2 are activated by both SSBs and DSBs (Cimprich and Cortez, 2008). PARP proteins attach poly(ADP-ribose) chains to the repair proteins to mobilize them to sites on chromatin where breakage or damage has occured. Activation of DDR leads to growth arrest of the damaged cells, which allows various repair mechanisms to repair the damaged DNA, after which the cells are able to continue with their cell cycle and functions (Hakem, 2008).

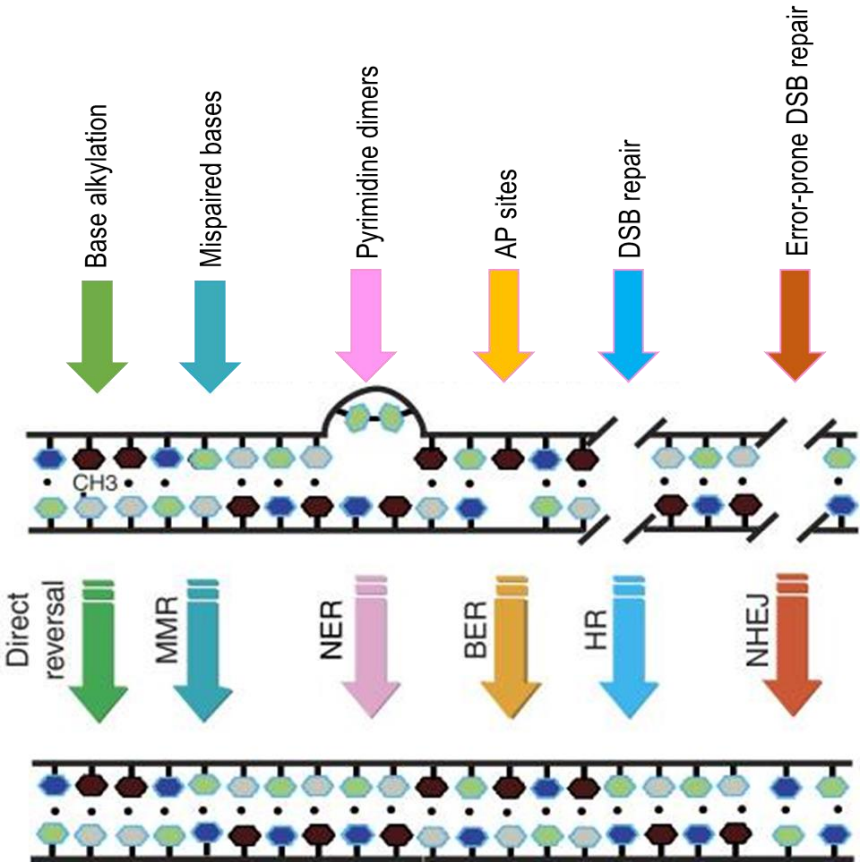


Figure 2. DNA repair pathways and types of DNA injuries they deal with. These pathways include the direct reversal pathway, MMR, NER, BER, HR, and the NHEJ pathway (adapted from Hakem, 2008).

Mispaired DNA bases are repaired by MMR pathway, whereas chemically altered DNA bases are removed by Base Excision Repair (BER) (Jiricny, 2013). Complex DNA lesions, such as UV-induced pyrimidine dimers, are repaired by removing an oligonucleotide of approximately 30 bp containing the damaged bases by Nucleotide Excision Repair (NER) pathway, while intrastrand crosslinks (ICLs) are corrected by Interstrand Crosslink Repair, also called Fanconi Anemia. SSBs are repaired by Single-Strand Break Repair (SSBR), whereas DSBs are mainly processed by error-prone non-homologous end joining (NHEJ) or homologous recombination (Waters et al., 2014). In contrast to NHEJ, which potentially leads to incorrect religation of DSBs and mutations, HR accurately restores the genomic sequence of broken DNA ends by using sister chromatid as template for repair (Hakem, 2008).

However, recent studies have identified a third pathway for DSB repair that is a subtype of NHEJ patway: microhomology-mediated end joining (MMEJ). Repair is initiated by resection of the end near the DSB, exposing short regions of complementary sequences of 2 to 20 nucleotides (microhomologies). DSB breaks are repaired by pairing DNA single strands based

on microhomology (MH), followed by nucleolytic trimming of DNA flaps, DNA gap filling synthesis and ligation, resulting in DNA deletion. MMEJ is a highly error-prone repair pathway and is always mutagenic, as one of the two MH regions and the inter-MH region are removed from the final repair product (Patterson-Fortin and D'Andrea, 2020).

Unlike other DNA damage repair pathways that involve multiple steps and require the activity of multiple proteins, direct reversal of DNA damage is a simple process that does not require excision of the damaged bases (Sedgwick et al., 2007) O6-alkylguanine is an example of a DNA lesion that is repaired by direct reversal. The enzyme O6-methylguanine-DNA methyltransferase (MGMT) transfers the alkyl group from the oxygen in the DNA to its active site, resulting in the reversal of the base damage (Hakem, 2008).

Rare small insertions and deletions occur during replication due to lower replication fidelity and errors that have escaped proofreading by replicative polymerases. These small mutations are repaired by the MMR pathway, which reduces the rate of spontaneous mutations (Jiricny, 2013). Mutations in MMR proteins are associated with hereditary and sporadic cancers in human cells, and the MMR system is also required for cell cycle arrest and apoptosis in response to certain types of DNA damage (Li, 2008).

The NER pathway is a multistep process that can be divided into two subpathways: global genome NER (GG-NER) which can occur anywhere in the genome, and transcription-coupled NER (TC-NER) which repairs lesions in actively transcribed genes. The NER pathway repairs bulky DNA lesions caused by mutagenic chemicals, UV radiation, or chemotherapeutic agents. After the DNA lesion is recognised by specific proteins, endonucleases make the incisions at the sites surrounding the lesion, and oligonucleotide containing the lesion is removed. Repair is completed by gap-filling synthesis of the oligonucleotide complementary to the pre-existing strand and ligation. Inactivation of the NER pathway has been associated with extreme UV sensitivity, growth defects, and increased risk of skin cancer (Hakem, 2008).

The BER pathway deals with the most common type of DNA lesion- base damage (Tiwari and Wilson, 2019). The BER pathway can be divided into two subpathways: the short-patch BER, in which a single nucleotide is replaced, and the long-patch BER, which leads to the incorporation of 2–13 nucleotides. The first step of the two subpathways involves removal of the damaged base by glycosylases, such as 8-oxoguanine DNA glycosylase (Ogg1) and mutY DNA glycosylase (Myh). After the glycosylase creates an apurinic or apurimidinic (AP) site, the endonuclease APE1 performs strand nicking. Several DNA polymerases are involved in gap-filling synthesis and nucleotide incorporation, followed by strand ligation.

Two important patways for repairing DNA double-strand breaks are HR and NHEJ (Li, 2008). HR is a complex process that occurs in the S or G2 phase of the cell cycle and requires several proteins. It provides template-dependent and error-free repair of DNA gaps, DSBs, and DNA interstrand crosslinks (ICLs), the repair of which also involves the Fanconi anemia pathway.

HR critically supports DNA replication and telomere maintenance and plays an important role in faithful duplication of the genome and maintenance of genome integrity (Li and Heyer, 2008). HR is initiated by generation of ssDNA, which is promoted by the MRE11–RAD50–NBS1 (MRN) complex. ssDNA invades the undamaged template in reaction catalysed by the RAD51 protein and breast-cancer susceptibility proteins 1 (BRCA1) and 2 (BRCA2), followed by the actions of polymerases, nucleases, and helicases, after which DNA ligation and substrate resolution occur (Jackson and Bartek, 2009).

NHEJ is active throughout the cell cycle and repairs DSBs in both dividing and non-dividing somatic cells, especially outside S and G2 phases, in the absence of HR (Li and Heyer, 2008). NHEJ is the dominant pathway for repair of DSBs at all phases of the cell cycle and is essential for T-cell receptor-α/β and Ig V(D)J recombination, demonstrating its key role in the development of the T and B-cell repertoires (Hakem, 2008). NHEJ is error-prone and often associated with the presence of small insertions and deletions (indels). DSBs are recognized by Ku proteins, which activate the protein kinase DNA-PK followed by activation of end-processing enzymes (polymerases and DNA ligase IV) that can excise, modify, or add nucleotides, resulting in mutations at the break site. Microhomology-mediated end-joining (MMEJ) is an NHEJ subpathway independent of Ku proteins, and repair products generated by this pathway always have sequence deletions. Although both NHEJ and MMEJ are imprecise repair pathways, they are active at all stages of the cell cycle (Jackson and Bartek, 2009).

1.3. DNA-protein crosslinks (DPCs)

DNA-protein crosslinks (DPCs) are DNA lesions which ocurr when protein ireversibly covalently binds to DNA (Vaz, Popovic and Ramadan, 2017). They are second most common lesions occuring daily in human genome, while most frequent are AP sites (Ruggiano and Ramadan, 2021). It is estimated that they are occuring at the rate of approximately 6000 per day in mammalian cells (Oleinick et al., 1987). DPCs are created under physiological conditions caused by reactive oxygen and nitrogen species (ROS and NOS), aldehydes and DNA helical alterations (Vaz, Popovic and Ramadan, 2017). Endogenous formaldehyde forms as a byproduct of lipid peroxidation and histone, DNA and RNA demethylation (Tretyakova, Groehler and Ji, 2015) and can reach 66–100 micromolar concentration in human blood (Klages-Mundt and Li, 2017). DPCs are aslo caused by exogenous sources including ultraviolet light (UV), ionizing radiation (IR), aldehydes, metal ions, and certain types of anticancer agents (Ide et al., 2018). DPCs pose major risk to genomic stability and are highly harmful to living cells, as they constitute steric blockage to all DNA transactions: transcription, replication, repair and chromatin remodeling. If not properly repaired, DPC lesions can lead to mutations, genomic instability, and/or cell death (Ruggiano et al., 2021). On the organismal level

impaired DPC repair was so far shown to cause premature aging phenotypes and cancer (Ruggiano and Ramadan, 2021).

Proteins of different size, function, cellular localization, and physiochemical properties can become crosslinked to DNA, contributing to the complexity and heterogeneity of DPCs. In addition, covalent bond can form between different DNA sites (N7 of guanine, the C-5 methyl group of thymine, and the exocyclic amino groups of guanine, cytosine, and adenine) and numerous amino acid side chains on proteins. Analysis of crosslinked proteins identified several DNA-binding proteins (transcription factors, histones, proteins involved in DNA repair and replication) and non-DNA-binding proteins (Kiianitsa and Maizels, 2013; Vaz, Popovic and Ramadan, 2017). Studies using a proteomic approach to identify proteins crosslinked to DNA have shown that the repertoire of adducted proteins is enriched in histones, high mobility group (HMG) proteins, and proteins that are part of the spliceosome. A significant proportion of DPCs consisted of proteins that form reversible covalent intermediates with DNA, such as topoisomerase 1 and 2, DNA-(cytosine-5)-methyltransferase 1 (DNMT1), and Ku (XRCC5/6) proteins. Histones and HMG proteins were also highly enriched, especially after treatment with the DPC-inducing agent formaldehyde (Kiianitsa and Maizels, 2013).

DPCs differ according to properties of the crosslinked proteins, the size, the chemistry of crosslinking bonds, and whether they are associated with flanking DNA nicks. Since both enzymes and general proteins can be covalently crosslinked to DNA, DPCs can be classified as non-enzymatic (general) or enzymatic (Figure 3). Non-enzymatic (general) DPCs are formed by covalent binding of proteins to unbroken DNA strand, whereas enzymatic DPCs are formed by covalent attachment of proteins near a single-strand or double-strand break (Tretyakova, Groehler and Ji, 2015). Non-enzymatic DPCs are most commonly caused by endogenously present reactive metabolites that cause nonspecific trapping of proteins near DNA, but they can also be caused by exposure to exogenous DPC inducers. Enzymatic DPCs are formed when enzymes that form short-lived covalent reaction intermediates with DNA (topoisomerases, DNA polymerases, and DNA methyltransferases) are trapped on DNA in the presence of crosslinking agents (Vaz, Popovic and Ramadan, 2017).

Non-enzymatic (general) DPCs are formed when proteins located near DNA become crosslinked to undisrupted DNA strands in the presence of endogenous or exogenous DPC-inducing agents (Zhang, Xiong and Chen, 2020). This type of DPCs is most prevalent under physiological conditions. Processes such as histone demethylation, amino acid metabolism, AlkB-type repair, and lipid peroxidation produce various reactive aldehydes in chromatin proximity (Vaz, Popovic and Ramadan, 2017). Aldehydes can react with lysine, cysteine, and histidine protein residues to form an adduct, which can then further react with a primary amine of a nearby DNA base, resulting in the formation of a stable amide bond and DPC (Klages-Mundt and Li, 2017). In addition to aldehydes, reactive oxygen and nitrogen species (ROS and RNS) can also form free radicals and electrophilic lesions that can react with both

DNA and proteins, leading to a crosslinking reaction and DPCs (Vaz, Popovic and Ramadan, 2017).

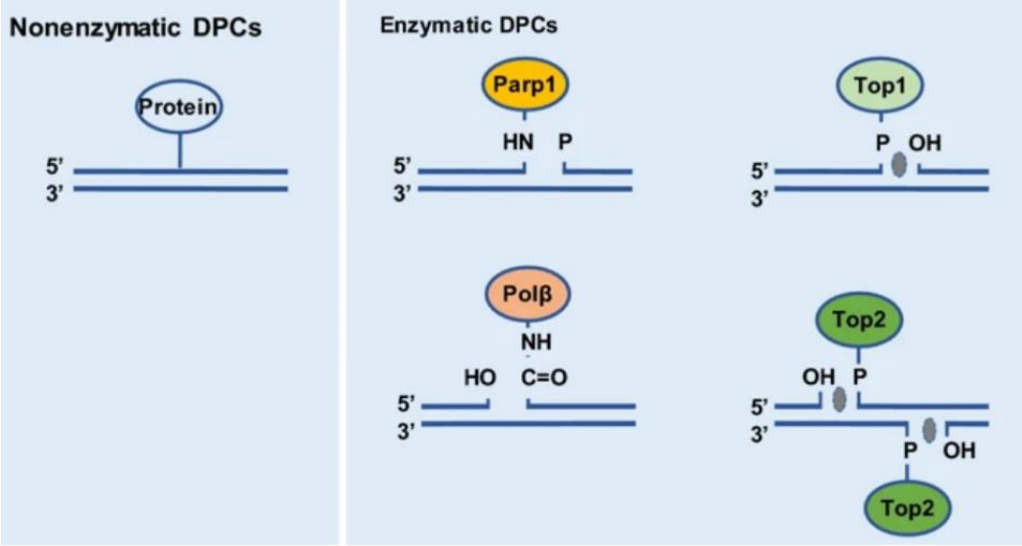


Figure 3. DPCs can be categorized as non-enzymatic or enzymatic based on the characteristics of the crosslinked proteins and according to whether they are associated with disrupted or undisrupted DNA strands (Zhang, Xiong and Chen, 2020).

DPCs can arise from normal catalytic cycles of various enzymes that transiently form covalent complexes with DNA. In the presence of DPC-inducing agents, these transient covalent complexes can become irreversibly crosslinked with DNA, resulting in enzymatic DPCs. The most prominent enzymatic DPCs are the DNA-TOPs crosslinks. TOP1 introduces single-strand break to reduce the torsional stress of DNA supercoiling. TOP1 is bound to the 3' end of the single-strand break, while the free 5'OH can rotate around the intact DNA strand. When the torsional stress is released, TOP1 catalyzes the repair of the single-strand break and is then released from DNA. However, in the presence of DNA structure distortion, that disturbs the alignment of DNA strands, TOP1-dependent annealing of single-strand breaks can be inhibited, resulting in TOP1 being trapped at the site of the single-strand break and forming a DPC. Similarly, TOP2 can be trapped at the terminal ends of DSBs, because TOP2 induces double-strand breaks (Zhang, Xiong and Chen, 2020). This specific DPCs arise upon treatment with chemotherapeutic drugs such as camptothecin and etoposide, which specifically trap Top1 or Top2 on DNA by inhibiting the re-ligation of breaks, thus preventing replication and transcription (Zhang et al., 2011). They can also arise near abasic sites that misalign DNA strands, preventing re-ligation and trapping Top1 on DNA. Other DNA repair enzymes can also form DPCs. O6-methylguanine-DNA methyltransferase (MGMT), an enzyme that removes alkyl adducts from DNA, forms DPCs when cells are treated with nitrogen mustard. Abasic sites, which are the most common type of spontaneous DNA damage and are formed during base excision repair or DNA demethylation, lead to covalent binding of histones or enzymes due to their intrinsic property to crosslink proteins. Poly(ADP-ribose)polymerase-1

(PARP-1), an enzyme involved in DNA damage detection, repair, and chromatin remodeling, covalently binds to abasic sites formed during BER, resulting in the formation of a DPC (Prasad *et al.*, 2014). The chemotherapeutic agent 5aza-2'-deoxycitidine (5-azadC) covalenty trappes various DNA methyltransferases (DNMTs) during methylation of this cytosine analog, resulting in specific DNMT- DPCs. DNA polymerase β, repair enzyme involved in BER, can also be covalently trapped when it acts on certain types of lesions (Klages-Mundt and Li, 2017).

DPC formation can be facilitated by naturally occurring and synthetic compounds. DPC-inducing agents existing in our environment, beside ionizing radiation and UV light, are also various transition metal ions, including chromium and nickel. In addition, there are a variety of endogenous metabolites such as reactive aldehydes, ROS, RNS, and helical DNA modifications that lead to the formation of DPCs in cells (Klages-Mundt and Li, 2017). Anticancer drugs, such as nitrogen mustards, 5-azadC, platinum-based agents, camptothecin, and etoposide lead to the formation of therapeutically-induced DPCs (Ruggiano et al., 2021b). 5-aza-dC, a cytosine analogue that can be incorporated into DNA and acts as a pseudosubstrate for DNMT1, results in the specific DNMT1-DPCs through covalent binding of the enzyme (Stingele, Bellelli and Boulton, 2017). Platinum derivatives, such as cisplatin, carboplatin and oxaliplatin cause non-specific DPCs by crosslinking non-enzymatic proteins with DNA. The chemotherapeutic agents camptothecin and etoposide, which are used to treat various cancers, are used clinically to specifically induce TOP1- and TOP2-DPCs by intercalating into the enzyme–DNA interface.

1.4. DNA-protein crosslink repair (DPCR)

All three components that are part of the DPC can be the targeted for repair. Nucleases can degrade the DNA component, proteases can degrade a protein component, and the covalent bond can be hydrolysed, whereupon the entrapped protein is released from the DNA (Figure 4). Hydrolysis of crosslink can be complicated because the covalent bond is usually hidden within the protein–DNA interface and is not easily accessible to repair enzymes (Stingele, Bellelli and Boulton, 2017).

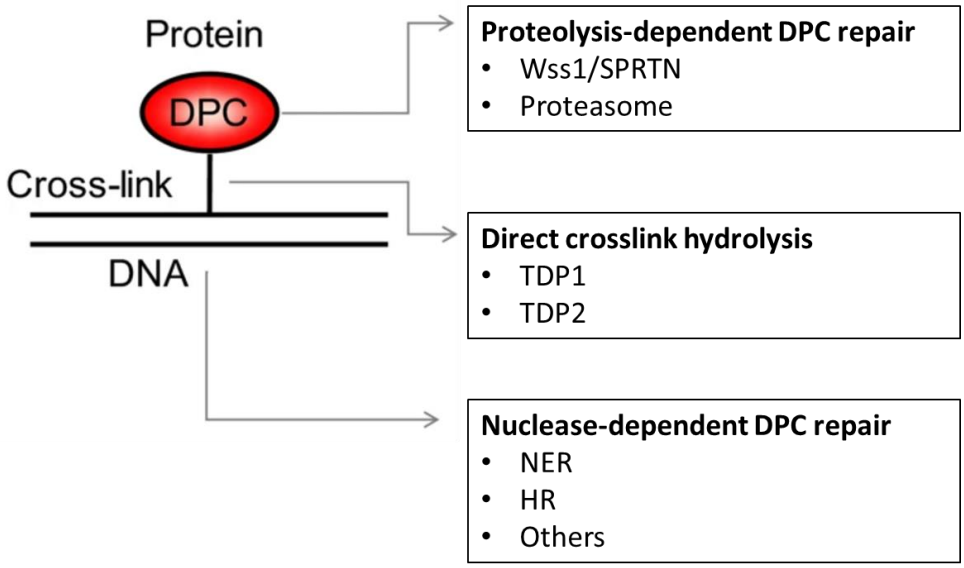


Figure 4. Schematic representation of three DPC components which can be targeted for repair (adapted from Zhang, Xiong and Chen, 2020).

1.4.1. Proteolysis dependent repair

The protein part of a DPC represents a major steric barrier to repair, by preventing direct access of DNA repair factors to the affected bases (Klages-Mundt and Li, 2017). Therefore, the first step in DPC repair involves proteolytic cleavage of the protein part, resulting in a peptide that is covalently bound to the DNA and allows downstream repair factors to access the lesion in to complete repair. Alternatively, the small peptide residues of up to 15 aminoacids can be bypassed by translesion synthesis (TLS) polymerases (Ghodke et al., 2021).

Until recently, it was not known that in eukaryotes there is a specific DPC repair pathway in which replication-coupled DNA-dependent proteases play an important role. The first specific DPC protease discovered was the DNA-dependent metalloprotease Wss1 (Weak Suppressor of Smt3) in the yeast *Saccharomyces cerevisiae*. It was shown that Wss1 cleaves DNA - binding proteins *in vitro* and yeast cells deficient in Wss1 are hypersensitive to DPC-inducing substances, such as formaldehyde (Stingele et al., 2014). Accordingly, studies in *Xenopus* egg extracts confirmed the existence of a proteolytic mechanism for DPC repair that is coupled to replication and is proteasome-independent (Duxin et al., 2014). The identity of the protease in metazoans was recently discovered, when several laboratories reported the existence of a DNA-dependent metalloprotease SPRTN (also known as DVC1)(Lopez-Mosqueda et al., 2016; Vaz et al., 2016; Stingele, Bellelli and Boulton, 2017). Like Wss1, human SPRTN cleaves multiple DNA-associated proteins *in vitro* and its inactivation in *Caenorhabditis elegans*, cultured human cells, and mouse embryonic fibroblasts (MEFs) results in hypersensitivity to

general (formaldehyde) and specific (camptothecin) DPC-inducing agents. The human SPRTN is part of the replisome and removes DPCs upstream of the replication fork. Although DPCs are very heterogeneous type of DNA damage, DPC proteases have the ability to digest different proteins, depending on their size, from smaller ones like histones to larger ones like topoisomerases (Vaz et al., 2016). SPRTN deficiency is associated with accelerated aging and increased risk of cancer development in humans and mice, indicating that repair by DPC proteolysis has major medical importance (Fielden et al., 2018).

Besides SPRTN, a second potential DPC protease may exist in higher eukaryotes. Acidic Repeat Containing (ACRC) protein, also known as GCNA (Germ Cell Nuclear Antigen), was recently discovered to be a protein with SprT domain. Phylogenetically, ACRC is very close to SPRTN and they have a very similar 3D structure of the protease core within the SprT domain. Given the phylogenetic proximity and similarity of their protease cores, ACRC may be proteolytically active and play a role in DPC repair (Fielden et al., 2018). Recently, other enzymes such as FAM111A and B, as well as DDI1 and DDI2 have emerged as potential DPC proteases. They have protease domains with conserved catalytic residues and they are catalytically active. FAM111A and FAM111B are serine proteases. While the function of FAM111B is still unknown, FAM111A is a PCNA interactor and has recently been implicated in the repair of TOP-1 and PARP-1 DPCs. The DDI1 and DDI2 proteins are aspartic proteases that interact with ubiquitin and the proteasome. Recently, the *S.cerevisiae* homolog Ddi1 was shown to be involved, together with the Wss1 protease, in the repair of DPCs induced by formaldehyde and camptothecin. However, direct involvement of these proteases in DPC repair has not yet been proven. Given the heterogeneity of crosslinked proteins, the existence of different DPC proteases with overlapping functions is not surprising (Ruggiano and Ramadan, 2021).

In addition to proteases, alternative pathway for DPC size reduction is a proteasome-mediated degradation, after which peptide residues are removed by NER or other pathways (Duan et al., 2021). Several enzymatic DPCs have been shown to be targeted for proteosomal degradation, such as Top1 and Top2. Proteasome-mediated degradation of the DPC protein moiety allows processing of a DNA-peptide adduct by the enzymes Tdp1 or Tdp2 (Pommier et al., 2006). Consequently, several experiments in mammalian cell lines have shown that proteasome inhibition prevents efficient repair of some general DPCs, such as those induced by formaldehyde. Proteasome activity not only improves cell survival, but is also important for replication and cell cycle recovery after exposure to formaldehyde. Based on these data, it is certain that the protein residues of the specified DPCs can be degraded by the proteasome (Klages-Mundt and Li, 2017).

1.4.2. Nuclease dependent repair

In eukaryotes, nucleases involved in NER and HR participate in DPC repair by cleaving DNA near the site where a DPC is formed. This was first shown in bacteria, where NER removes DPCs smaller than 16 kDa, whereas HR repairs bulky DPCs. This assumption was also

confirmed in mammalian cells, where NER was shown to remove DPCs smaller than 16 kDa *in vitro* (Chesner and Campbell, 2018). In addition, Nakano et al (2007) showed that mammalian NER repairs DPCs with a size limit of 8-10 kDa. The role of the NER in repairing DPCs larger than 10 kDa is still unknown (Nakano et al., 2007). However, proteolytic degradation of crosslinked protein is thought to be necessary before repair by the NER machinery. The activity of both NER and HR is cell cycle-dependent, so it is speculated that NER repairs DPCs before S phase, while Wss1 or HR remove the remaining DPCs in S phase (Fielden et al., 2018). While the role of NER in DPC repair is favorable to cells, HR requires DSB formation, which can be potentially harmful to cells (Vaz, Popovic and Ramadan, 2017).

The MRN complex, composed of Mre11, Rad50, and Nbs1 protein, is the major complex for DSB repair by HR. Interestingly, it was recently shown that deficiency of Mre11 does not affect DPC removal in human cells, suggesting that HR cannot repair the majority of different DPCs (Vaz, Popovic and Ramadan, 2017). Endonucleases that cleave the DNA around a DPC, such as XPF-ERCC1, the MRN complex, and CtBP-interacting protein (CtIP), are involved in repairing Top1- and 2- DPCs (Fielden et al., 2018). MRE11 has been shown to play a role in DPC repair in human cells that is independent of its function in HR. Mre11 removes TOP2-DPCs indirectly by cleaving DNA 15–20-bp downstream of DPC formation. Similar results from *Xenopus* egg extracts showed that etoposide-induced TOP2-DPC are repaired by the coordinated activity of BRCA1, MRN complex, and CtIP (Aparicio et al., 2016). A study by Hoa and coworkers (2016) also showed that Mre11 plays an important role in maintaining genome integrity by removing TOP2-DPCs and promoting the processing of DSBs for HR. The MRN complex is also required for TOP2-DPC repair in G1 phase, indicating that Mre11 can remove these enzymatic DPCs independently of HR (Hoa et al., 2016).

APE2 is an apurinic/apyrimidinic 3'end unblocking endonuclease involved in BER that reverse blocked 3' ends of DNA that inhibit DNA synthesis (Li, 2008). APE2 acts on a variety of endogenously arising 3' blocking lesions and may have partially redundant activity with TDP1 in repairing TOP1-induced 3' -blocking lesions (Álvarez-quilón et al., 2021). In addition to APE2, there are several enzymes which guard the cells from the accumulation of 3' blocking lesions, including Rnase H2, APE1, and TDP1, which process embedded ribonucleotides, AP sites, and TOP1-DPCs.

1.4.3. Hidrolysis dependent repair

Therapeutically very relevant DPCs are topoisomerase 1 and 2 DPCs, which are fromed when the catalytic cycle of the enzymes is inhibited by anticancer drugs (camptothecin and etoposide) or DNA distortions, making them irreversibly covalently bound to DNA. Both complexes can have pathogenic or cytotoxic consequences by disrupting essential DNA processes and are associated with neurodegenerative diseases. In addition, camptothecin (CTP) and etoposide (ETO), which are frequently used in therapy, kill cancer cells by trapping these complexes and causing DPCs.

Topoisomerase DPCs can be repaired by several pathways. Because they are so common, specialized repair enzymes have evolved to remove them specifically. The phoshodiesterases tyrosyl-DNA phoshodiesterase 1 (TDP1) and tyrosyl-DNA phoshodiesterase 2 (TDP2) hydrolyze the phosphodiester bonds between the DNA and Tyr in the active site of TOP1 or TOP2 (Stingele, Bellelli and Boulton, 2017). Repair of TOP1-DPCs depends on partial proteasomal degradation, presumably by Wss1 (Stingele et al., 2014) in yeast and SPRTN in higher eukaryotes (Vaz et al., 2016). Following proteosomal degradation of TOP1 by Wss1 or SPRTN, the remaining peptide is subsequently removed from DNA by TDP1. TDP1 catalyzes the hydrolysis of the bond between tyrosine in the catalytic center of Top1 and the 3' DNA end of nicked DNA (Pommier et al., 2006). Recent experiments have shown that yeast cells lacking TDP1 are resistant to CPT due to Wss1, whereas CPT is lethal to Wss1 and Tdp1 double deletion mutants (Stingele et al., 2014). After the phosphotyrosyl bond is hydrolyzed, 3' phosphate ends incompatible for simple religation are produced by TDP1. Therefore, further processing of the ends by the bifunctional polynucleotide kinase 3'-phosphatase (PNKP) is required. PNKP removes 3' phosphate and phosphorylates the 5' hydroxyl end, so that the SSB repair machinery can ligate the remaining DNA nick (Stingele, Bellelli and Boulton, 2017).

TOP2 is irreversibly bound to 5' ends of DNA and the specific enzyme TDP2 hydrolyzes these 5'-phosphotyrosyl bonds. How TOP2- DPCs are resolved is still unclear, but proteolytic degradation od Top2 is thought to be required prior to TDP2 processing, as cells lacking SPRTN cannot repair Top2- DPCs (Lopez-Mosqueda et al., 2016). In addition, Vaz et al (2016) showed that depletion of SPRTN in human cell lines significantly increses the amount of Top1, Top2α, histone H3, and histone H4 DPCs (Vaz et al., 2016). *In vitro* experiments with reconstituted TOP2-DPCs showed that TDP2 can remove TOP2-DPCs without upstream proteolysis in cooperation with the SUMO ligase ZATT (ZNF451). Binding of ZATT to TOP2-DPCs facilitates TDP2 hydrolase activity and induces interaction between TDP2 and SUMOylated TOP2, allowing efficient TDP2 recruitment. It is hypothesized that ZATT changes the conformation of TOP2-DPCs to facilitate direct repair by TDP2. These results reveal a SUMO2-mediated pathway for direct removal of TOP2-DPCs catalyzed by ZATT and TDP2 (Schellenberg et al., 2017). However, this mechanism remains to be investigated in cell lines and *in vivo* at the organismal level. TDP2 generates DSB with 5'-phosphorylated ends, suitable for direct ligation, so DSBs generated during TOP2-DPCs repair are usually repaired by NHEJ, although HR can also occur (Stingele, Bellelli and Boulton, 2017).

1.5. Proteases in DPCR

The existence of a specialized replication-dependent DPC repair pathway remained unknown until recently. It has been speculated that cells have a specialized DPC repair pathway, because the NER can excise small DPCs and removal of topoisomerase DPCs requires upstream proteolysis into smaller peptides. Several laboratories have identified the proteolytic DPC repair pathway, with DPC-specific proteases discovered in yeast, *Xenopus*

laevis, *Caenorhabditis elegans*, and mammalian systems (Klages-Mundt and Li, 2017). The laboratory of Stefan Jentsch was the first to identify the DNA-dependent protease Wss1 (weak suppressor of SMT3 protein 1) in yeast, which is involved in TOP1-DPCs repair and promotes cell survival after exposure to formaldehyde. These insights from yeast led to the awareness that a similar enzyme might also exist in metazoans and SPARTAN (SPRTN) or DVC1 was discovered (Lopez-Mosqueda et al., 2016; Vaz et al., 2016). Wss1 and SPRTN proteolyze large DPCs into small peptide residues that are presumably further processed by TDP1, TDP2, and NER or bypassed by translesion DNA synthesis (TLS) (Zhang, Xiong and Chen, 2020).

1.5.1. Wss1 and SPRTN proteases

Both proteases are linked to DNA replication, cleave DNA-binding proteins, require DNA to activate their proteolytic activity, and their deficiency leads to hypersensitization of yeast and human cells to formaldehyde, showing a general role in DPC repair. Despite numerous similarities, SPRTN and Wss1 belong to two evolutionarily distant families that do not share a common ancestor (Vaz, Popovic and Ramadan, 2017). The SPRT family is found in cyanobacteria, bacteria, archaea, plants, and animals, and all metazoans have SPRTN proteases. WLM proteins are found in yeast, fungi, and plants. SPRTN and Wss1 proteases share conserved protease domains, although they are not orthologs, but functional homologs. The greatest similarity between the two proteins is the conserved metalloprotease active site, which is part of the N-terminal SPRT or WLM protease domain (Vaz, Popovic and Ramadan, 2017). Both SPRTN and Wss1 contain a HEXXH active site consisting of two zinc-binding histidines and a glutamic acid, making them members of zincin family of metallopeptidases (Figure 5.) Both proteins have C-terminal tail regions bearing interaction motifs, that is 228 amino acids longer in SPRTN than in Wss1 (Vaz, Popovic and Ramadan, 2017; Fielden et al., 2018).

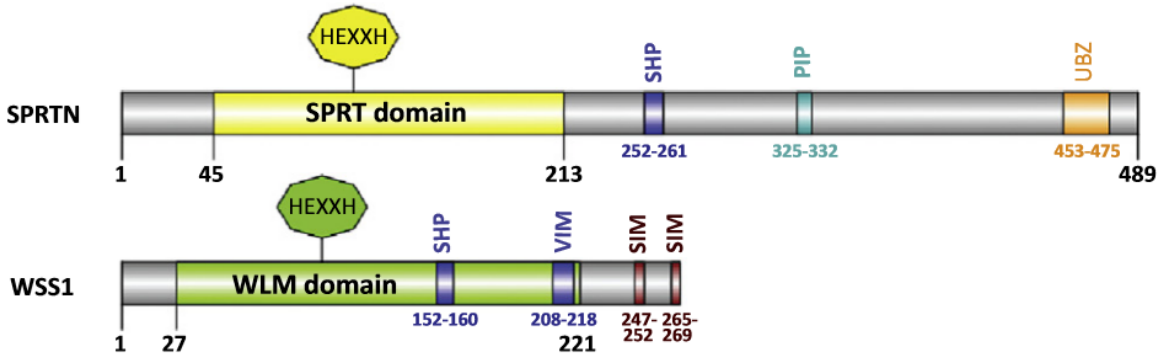


Figure 5. Structure of SPRTN and Wss1 proteases. SPRTN and Wss1 share a HEXXH motif in the active site, and a p97/Cdc48-binding motif (SHP and VIM). SPRTN has a 276 amino acids long C-terminal part containing PCNA-interacting peptide (PIP) and ubiquitin-binding motifs (UBZ). Wss1 has a 48 amino acids long C-terminal arm bearing SUMO interaction motifs (SIM) (Vaz, Popovic and Ramadan, 2017).

In the C-terminal part of both proteins is a SHP motif that allows binding to the ATPase p97 (VCP/Cdc48 in yeast), a segregase that could support proteolysis of DPC intermediates, and motifs for binding ubiquitin (SPRTN) or small ubiquitin modifier (SUMO in Wss1). SPRTN can bind DNA both indirectly, through interaction with the proliferating cell nuclear antigen (PCNA) and the PIP motif, and directly through four DNA binding motifs in the C-terminal region and one in the SprT protease domain (Vaz et al., 2016). Binding of SPRTN to stalled replication forks due to DNA damage was thought to depend on PCNA ubiquitylation by the Rad18 ubiquitin ligase, while other results showed that it is independent of PCNA modification. However, all reports agree that SPRTN and Wss1 act at the replication fork and that both are essential for continuation of replication, suggesting that they remove DPCs during replication (Vaz, Popovic and Ramadan, 2017). The C-terminal portion of SPRTN plays an important role in TLS. DNA damage triggers replication arrest, followed by monoubiquitylation of PCNA, which is recognized by TLS polymerases such as Pol η and Pol ζ that can bypass the DNA lesion. SPRTN plays an important role in the recruitment of Rad18 ligase to chromatin, leading to PCNA ubiquitylation and binding of TLS polymerases to the sites of damage. SPRTN binds ubiquitylated PCNA with both the PIP box and the UBZ domain, and subsequently recruits p97 to sites of DNA damage. The role of p97 segragase is to extract TLS polymerases from chromatin after replication of damaged DNA to increase the risk of introducing mutations due to excessive TLS (Davis et al., 2012). These results demonstrate that SPRTN induces Rad18-mediated PCNA ubiquitylation and TLS (Centore et al., 2012).

In vitro studies have shown that Wss1 and SPRTN have similar proteolytic activity. SPRTN cleaves histones, H2A, H2B, H3, H4, linker histone H1, HMG1,HLTF, Fan1, TOP1, and TOP2 *in vitro* (Stingele, Habermann and Jentsch, 2015; Vaz et al., 2016; Mórocz et al., 2017), while the substrates of Wss1 are Top1, histone H1, high mobility group protein 1, and it also cleaves itself in a DNA-dependent manner (Stingele et al., 2014). In contrast to the similarities found in the *in vitro* studies, the *in vivo* experiments showed differences in DPC removal and sensitivity to DPC-inducing agents. Inactivation of SPRTN leads to a massive increase in total DPCs in human cells, whereas yeast cells lacking Wss1 do not accumulate total DPCs and do not show hypersensitivity to CPT (Stingele et al., 2014). Moreover, cells deficient in SPRTN are sensitive to formaldehyde and unable to repair formaldehyde-induced DPCs, whereas Wss1 is not involved in the repair of DPCs induced by formaldehyde. Although TOP1 is a common substrate for both proteases, there are differences in the repair of TOP1-DPCs. Cells lacking SPRTN accumulate TOP1-DPCs and are sensitive to CPT, whereas depletion of Wss1 in untreated yeast cells has no deleterious effects. Unlike SPRTN in mammals, Wss1 is not an essential gene in yeast (Vaz, Popovic and Ramadan, 2017).

SPRTN is necessary for DPC repair and maintenance genome stability. Mutations in SPRTN lead to premature ageing and tumorigenesis in mice, and knock-out of SPRTN in mice is embryonic lethal (Maskey et al., 2014). SPRTN mutations in humans lead to a rare disorder known as SPARTAN or Ruijs-Aalfs (RJALS) syndrome. At the organismal level, the syndrome is characterised by early-onset hepatocellular carcinoma and premature ageing, while at the

cellular level, cells derived from RJALS patients exhibit an increased frequency of DSBs and DNA replication stress (Lessel et al., 2014). The phenotypes of RJALS patients, namely progeroid syndrome and liver tumours, have been recapitulated in mice carrying a SPRTN mutation (Maskey et al., 2014). Overall, these phenotypes confirm the connection between DPC repair and accelerated ageing and cancer (Fielden et al., 2018). They also show that the active residue (E112) of SPRTN metalloprotease is essential for the repair of endogenous DPCs in human cells. Vaz and coworkers confirmed that the patient mutation SPRTN^{Y117C} affects SPRTN enzymatic activity on all tested substrates. The other pathogenic mutation, SPRTN-ΔC, has autocleavage activity and can proteolyze histones, but is less efficient at processing TOP-DPCs. This result indicates that the C-terminal part of SPRTN may be important for binding specific substrates (Vaz et al., 2016).

To avoid potentially deleterious activity of DPC proteases, there are several regulatory mechanisms that prevent unnecessary cleavage of chromatin-associated proteins. The activity of Wss1 and SPRTN is regulated by the cell cycle, DNA binding, self-cleavage, and post-translational modifications (PTMs) (Fielden et al., 2018). SPRTN is expressed primarily in S and G2 phase, whereas in G1 phase it is degraded by the APC/Cdh1 complex. Its association with the replisome allows DPC degradation before the replication fork runs into DPCs. Analogous cell cycle regulation of Wss1 activity was not observed, although its levels are generally very low. Another similarity between Wss1 and SPRTN is their DNA-dependent activity. SPRTN has multiple DNA-binding motifs, whereas Wss1 has only one and DNA is a link between enzyme and substrate. In both proteases, double-stranded DNA (dsDNA) activates self-cleavage activity, whereas single-stranded DNA (ssDNA) fully activates protease activity and substrate cleavage (Stingele, Bellelli and Boulton, 2017). Self-cleavage is another mechanism of regulation. For both proteases, DNA has been shown to stimulate self-cleavage *in trans,* resulting in cleavage of C-terminal fragments which are realesed from the DNA, while the protease domain remains intact. This could be a mechanism by which active proteases are released from chromatin or by which proteolysis is terminated when DPC is removed. Indeed, it has been shown that the autocleavage products of SPRTN cannot bind DNA. Another way to regulate SPRTN activity is through posttranslational modifications. SPRTN is mono-ubiquitylated in cells and SPRTN is proteolytically inactive while ubiquitin is bound to the ubiquitin-binding domain (UBZ). Induction of DPCs triggers its deubiquitylation, relocalization to chromatin, and activates its proteolytic activity. In addition to ubiquitylation, SPRTN also has numerous SUMOylation sites. Modification by SUMO proteins is increased after proteasomal inhibition and replication stress, so this could be a mechanism to mark damage sites so that DPC proteases can recognize and repair their substrates (Vaz, Popovic and Ramadan, 2017).

However, detailed regulation of DPC proteolytic repair is still unknown. It remains to be explored how SPRTN is regulated, what repair factors operate downstream of SPRTN, and what other factors participate in DPC repair. In addition, it is not yet known which factors act upstream of TDP1 in non-cycling cells, since TDP1 removes the TOP1 peptide after

transcription stalling in postmitotic cells. NER, TLS and HR are thought to be active during S phase, whereas NER and MRE11 are also active in non-dividing cells. DPC repair is also thought to occur in non-proliferative cells such as neurons, where DPCs may interfere with transcription, but it is not yet known how DPC repair is regulated and what factors are involved. However, it is known that SPRTN, as a component of the replisome, is responsible for the DPC proteolysis-based repair pathway during replication (Vaz, Popovic and Ramadan, 2017).

1.5.2. ACRC/GCNA putative protease

ACRC is another possible protease that could be present in higher eukaryotes in addition to SPRTN. ACRC/ GCNA (Germ Cell Nuclear Antigen), belongs to a family of IDR-containing metalloproteases (Carmell et al., 2016) that contain an SprT domain (Figure 6) (Carmell et al., 2016; Vaz, Popovic and Ramadan, 2017). 80–100% of the protein consists of highly disordered regions. Different species have different domains of the protein, because some of them have been lost during evolution, such as the SprT domain, which is present in all metazoans except mice. Analyses of GCNA orthologs in all eukaryotes showed that a GCNA protein has four domains: a large IDR domain, a zinc metalloprotease domain, a C2C2 zinc finger, and a non-canonical two-helix HMG box. Experiments have shown that GCNA mutations in mice impair male fertility, highlighting the importance of the IDR (Carmell et al., 2016). The N-terminal portion of the protein is highly acidic and the C-terminal region is basic and contains nuclear localization signals (Nolte et al., 2001). ACRC is highly expressed in stem cells and germ cells in all eukaryotes, suggesting that it may be involved in protecting the genomic integrity of cells carrying a heritable genome (Carmell et al., 2016).

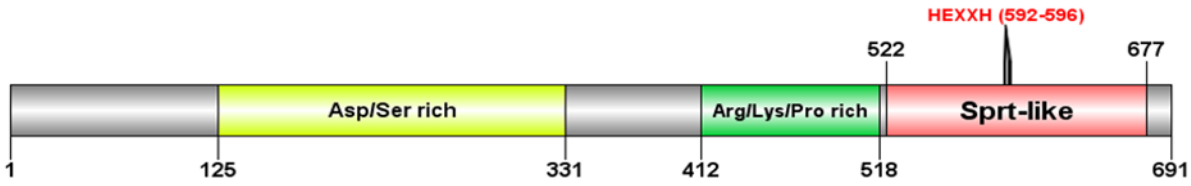


Figure 6. Schematic representation of human ACRC protein.

ACRC and the SPRTN family of proteases are phylogenetically very related (Figure 7). Zebrafish shares one-to-one orthology with human ACRC/GCNA (ACidic Repeat Containg/Germ Cell Nuclear Antigen) when comparing Sprt-domains. The ACRC protein family is evolutionary closer to the SPRTN group than to yeast protease Wss1, a functional ortholog of SPRTN. WLM group represents yeast Wss1 protease and its orthologs in other species (Fielden et al., 2018). Therefore, it is not surprising that ACRC and SPRTN share a very similar 3D structure of the protease active site in the SprT domain. The active site of ACRC is thought to have two α helices, three zinc-binding histidines, and a catalytic glutamate, which together form a HEXXH

motif, present in all zinc dependent metalloproteases. However, it remains to be discovered whether ACRC is a protease involved in DPC repair (Fielden et al., 2018).

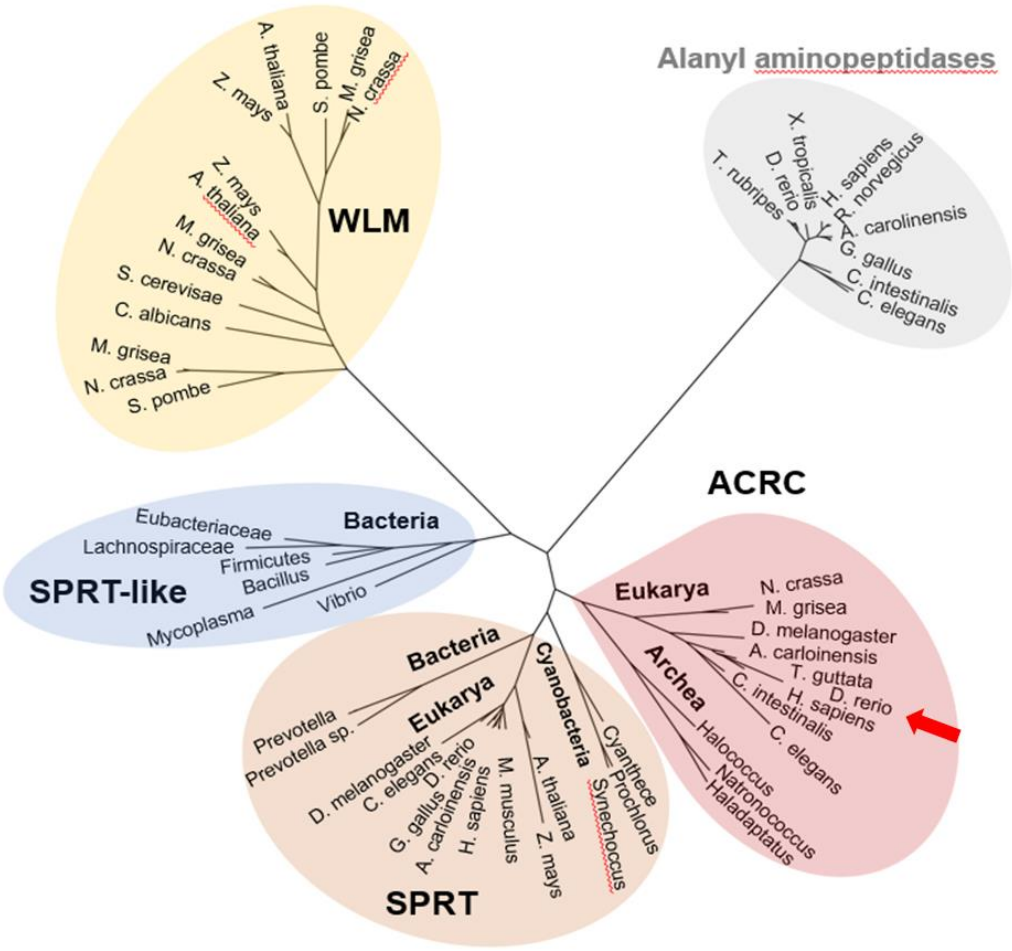


Figure 7. Phylogenetic analysis of Sprt domain in ACRC/GCNA orthologs. ACRC orthologs are found in archea and eukarya, while absent in prokaryotes. Alanyl aminopeptidase family of gluzincins were used as an outgroup. Protein sequences of SprT and WLM domains were aligned using MAFFT and the phylogenetic tree was constructed in PhyML (Fielden et al., 2018).

Several papers report that ACRC is a replicative protease that acts in parallel with SPRTN to promote genome integrity (Dokshin et al., 2020). Germ cells have specialized pathways to protect their genomic integrity as they are exposed to numerous double-strand breaks, recombination events, histone exchange, and chromosome condensation. To explore whether GCNA/ACRC and SPRTN have similar functions, Dokshin et al (2020) mutated GCNA-1 in *C. elegans* and showed that it plays an important role in maintaining germline immortality and that *gcna-1* and *dvc-1* (SPRTN) have partially overlapping functions required for fertility. Treatment of *gcna-1* mutant *C. elegans* embryos with hydroxyurea (HU), an agent that

depletes dNTPs and stalls replication forks, increased embryonic mortality, demonstrating that GCNA-1 is involved in the response to replication stress. In contrast to SPRTN, which is mainly expressed during S and G2 phases of the cell cycle, GCNA-1 is highly expressed during G2/M phases and is located on condensed chromosomes during mitosis. It is suggested that ACRC may be involved in the repair of TOP2-DPCs. Both proteins are abundant in the germline of several species, and in addition to germline-specific functions, TOP2 is also necessary during early embryogenesis and activation of the zygotic genome. TOP2 and GCNA have been shown to interact and colocalize on condensed chromosomes during *C. elegans* embryonic development, suggesting a role for GCNA in TOP2- DPCs repair (Dokshin et al., 2020). Bhargava et al (2020) showed that GCNA deficiency leads to DPC accumulation, genomic instability and mitotic defects in germ cells and embryos of *Drosophila melanogaster, C. elegans*, zebrafish, and human germ cell tumors. The results of this study suggest that GCNA is associated with the replication machinery. This was demonstrated by the increased RPA foci in *Drosophila* cells, microsatellite instability in worms, and GCNA ability to immunoprecipitate components of the MCM complex. These results indicate that GCNA may be involved in regulating replication in germ cells and during embryonic development (Bhargava et al., 2020).

However, another study suggests that ACRC is linked to the post-replicative repair in a SUMO dependant pathway, which acts in parallel with replication-coupled mechanisms for DPC repair. Borgermann et al (2019) demonstrated that ACRC/GCNA-1 family of proteases interacts with SUMO proteins and that GCNA-1 stimulates organismal survival upon DPC formation in *Caenorhabditis elegans* in a pathway associated with SUMOylation (Borgermann et al., 2019). Results of this study showed that both non-specific and specific enzymatic DPCs induce chromatin SUMOylation, indicating a role of SUMOylation in DPC recognition and processing. Chromatin SUMOylation, which was induced as a response to DPCs, was active during interphase, without the need for ongoing DNA replication. Probably cells have both ubiquitin- and SUMO-driven pathways for recognizing and repairing DPCs by mechanisms in which different protease are involved, depending on their cell cycle expression. Therefore, SUMOylation may be included in the processing of DPCs in duplex DNA outside of S phase and in the absence of replication-coupled DPC repair factors. However, the precise role of ACRC in promoting DPC responses in association with SUMOylation is still not established (Borgermann et al., 2019).

1.6. Nucleotide excision repair pathway (NER) and its role in DPCR

1.6.1. GG-NER and TC-NER

NER is a repair pathway for bulky DNA lesions, such as UV-induced pyrimidine dimers and photoproducts, chemical adducts, intrastrand crosslinks, and cyclopurines induced by ROSs (Marteijn et al., 2014). NER deficiency in humans leads to rare syndromes (Xeroderma

Pigmentosum (XP), Cockayne syndrome (CS), and Trichothiodystrophy (TTD)), characterised by skin cancer, neurodegeneration, and ageing.

The NER pathway can be divided into two subpathways, global genome NER (GG-NER) and transcription coupled NER (TC-NER) (Figure 8). In the GG-NER subpathway, the entire genome is examined for incorrect base pairing, whereas TC-NER repairs actively transcribed genes and is activated by RNA polymerase II (RNA Pol II) stalling (Marteijn et al., 2014). Several common proteins are included in both subpathways and they undergo similar repair steps. The exception is the first step, where the DNA lesion is recognised by the XPC–RAD23B and DDB1–DDB2/XPE proteins in GG-NER or by the Cockayne syndrome group A (CSA) and Cockayne syndrome group B (CSB) proteins in TC-NER (Hakem, 2008). After damage recognition, the XPC protein binds to the ssDNA part of the undamaged DNA strand opposite the damage caused by the disrupted base pairing. Binding of XPC is followed by association of the transcription initiation factor IIH (TFIIH) complex, which verifies lesion in the damaged strand and opens the DNA around the lesion with its two DNA helicase subunits XPB and XPD. The XPA protein is also involved in damage verification by binding structurally modified nucleotides in ssDNA. Damage detection and verification is a reversible process that precedes 5′ and 3′ incision and lesion removal to prevent undesirable DNA modifications. The XPF-ERCC1 and XPG endonucleases cut the damaged strand on both sides around the lesion, creating a singlestrand part of 22–30 nucleotides. In this step, the RPA protein protects the undamaged strand from incision and targets the endonucleases to the damaged strand. The final step is DNA synthesis, which fills created gap and ligation. This process is mediated by the Proliferating cell nuclear antigen (PCNA), Replication factor C (RFC), DNA Pol δ, DNA Pol ε or DNA Pol κ, and DNA ligase 1 or XRCC1– DNA ligase 3, depending on the cell cycle phase. DNA Pol ε and DNA ligase 1 are active in replicating cells, whereas DNA Pol δ, DNA Pol κ and XRCC1–DNA ligase 3 are active in nonreplicating cells (Kusakabe et al., 2019).

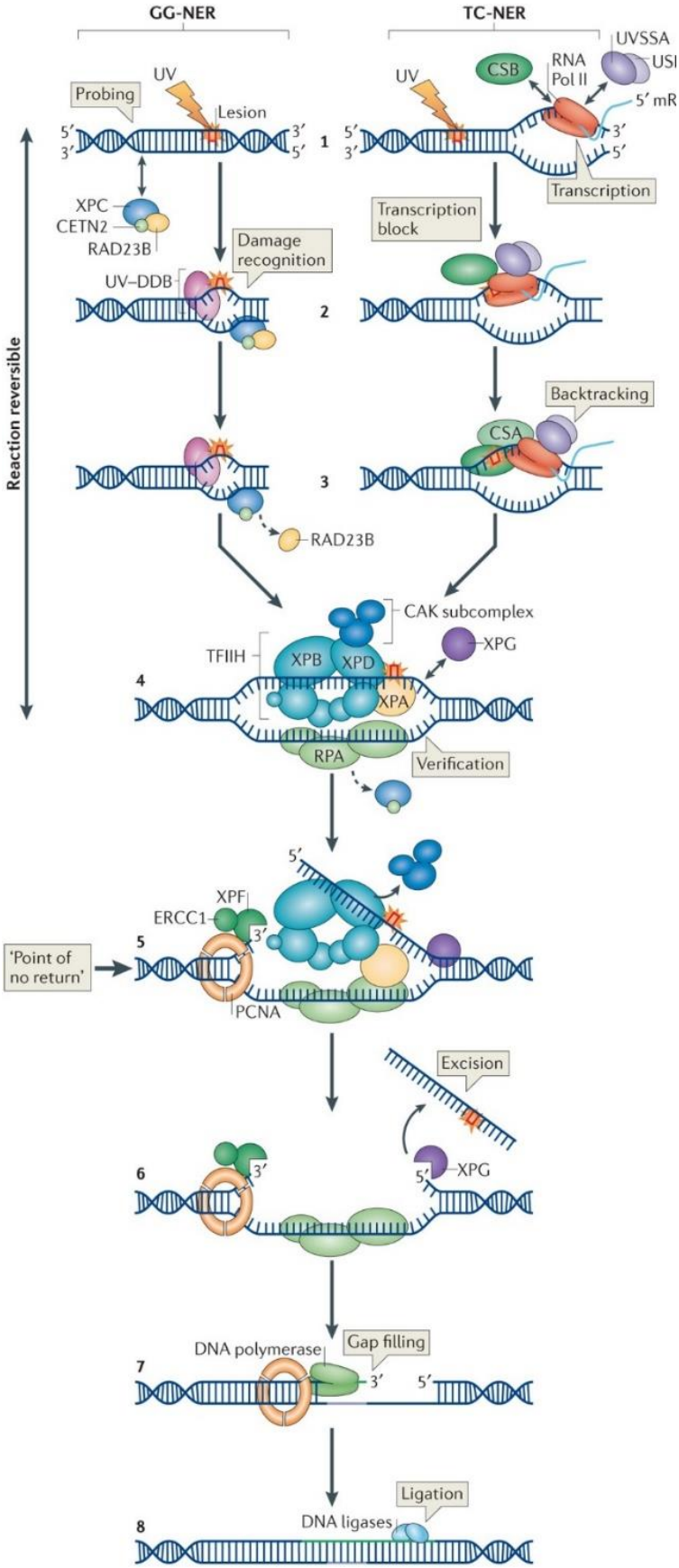


Figure 8. Schematic representation of GG-NER and TC-NER (Marteijn et al., 2014).

The TC-NER subpathway is activated by inhibition of transcript elongation, and thus indirectly detects the presence of a lesion. The TC-NER machinery consist of UV-stimulated scaffold protein A (UVSSA), ubiquitin specific processing protease 7 (USP7), XPA-binding protein 2 (XAB2) and high mobility group nucleosome binding domain containing protein 1 (HMGN1). The CSA and CSB proteins are required for the assembly of the machinery. When RNA Pol II is blocked at the site of a lesion, it likely detaches from the template DNA and recedes to allow recruitment of the TC-NER machinery or even degradates (Duan et al., 2021).

The XPA protein is the main regulator of the NER pathway, as it interacts with almost all NER proteins and plays a crucial role in both subpathways. It binds DNA with high affinity for damaged DNA (Marteijn et al., 2014). It plays a critical role in lesion verification by TFIIH and binds to modified nucleotides in ssDNA. XPA interacts *in vitro* with several key NER repair factors, including ERCC1, XPF, the p32 and p70 subunits of RPA, TFIIH, CSB protein, and the p34 subunit of the TFIIE. Given its preference for damaged DNA and its ability to interact with many NER repair factors, XPA verifies lesions and plays a central role in the recruitment of NER components around the injury. Some of the NER proteins are also known to be involved in other cellular processes (HR-dependent DSB repair, telomere maintenance, DNA replication and transcription), whereas XPA is thought to be specifically involved only in the NER pathway (Schärer, 2013).

1.6.2. NER in DPCR

Many *in vitro* experiments have shown that the NER pathway plays an important role in the DPC repair. Studies showed that the bacterial NER system can repair smaller DPCs of size 12–14 kDa (Ide et al., 2018). In addition, another study showed that UvrABC, the bacterial NER complex, can excise 16 kDa protein covalently attached to DNA. Similarly, NER proteins from *E. coli* are equally efficient at cutting peptides covalently bound to abasic sites (Minko, Zou and Lloyd, 2001). Bacterial NER mutants were also hypersensitive to the DPC and ICL inducer formaldehyde, but not to treatment with 5-azaC, which induces DNMT1- DPCs (Nakano et al., 2007).

The results demonstrating the involvement of the NER machinery in DPC repair in mammalian cells are contradictory. The mammalian NER system has been shown to make damage-specific cuts for smaller DPCs (below 1.5 kDa), whereas cutting is ineffective for larger DPCs of size 16 –37 kDa. Comparison of the incision efficiency as a function of DPC size *in vitro* has shown that the mammalian NER system can repair DPCs with a size limit of approximately 8 –10 kDa, which is much lower than bacterial NER (Nakano et al., 2009). NER may also play a role in repairing specific DPCs. One study has shown that NER-deficient human cells can repair DPCs induced by formaldehyde, whereas repair of DPCs induced by transplatin was inefficient. In addition, NER-deficient cells treated with formaldehyde exhibit chromosomal aberrations, micronuclei, and are sensitive to reactive aldehydes. However, it is not known whether this is due to a defect in the repair of DPCs or ICLs (Stingele, Bellelli and Boulton, 2017), as it has been shown that the NER is also involved in ICL repair. In contrast to the *in tube* experiments,

experiments in mammalian cell lines showed that the NER was unable to repair DPCs of 8 kDa induced by formaldehyde (Nakano et al., 2009). This result proves that the size of the protein is the main factor determining the efficiency of NER-mediated repair. Histones are the most abundant DPCs in cells (Kiianitsa and Maizels, 2013) and because they are larger than 8 kDa, proteolytic size reduction is most likely needed before NER-dependent DPC removal. It is known that proteins larger than 8–10 kDa can block the recruitment of NER repair factors and reduce the incision efficiency of NER nucleases (Zhang, Xiong and Chen, 2020). Indeed, all core histones and linker histone H1 have been shown to be substrates of SPRTN protease (Vaz et al., 2016). SPRTN cleaves disordered N- and C- terminal histone tails, reducing their size (Vaz et al., 2016). In addition, the NER nuclease complex XPF–ERCC1 has been suggested to be involved in the repair of TOP1-DPCs after upstream proteolysis (Zhang et al., 2011). Therefore, pre-processing by proteolytic trimming of proteins or loosening of their structure is likely required prior to activation of NER pathway (Fielden et al., 2018).

1.7. CRISPR/CAS9 genome editing

The Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated (CRISPR/Cas) system is a prokaryotic immune system that incorporates short sequences from plasmids and phages into the host CRISPR locus to generate resistance to foreign genetic elements (Figure 9). The integrated sequences are then transcribed and processed into short RNAs that direct the Cas9 nuclease to degrade the invading nucleic acids (Charpentier and Marraffini, 2014). The CRISPRs locus was first discovered by Ishino and co-workers (1987), who discovered a locus in the *Escherichia coli* genome with repeat sequences and spacers derived from viral or plasmid genomes (Khadempar et al., 2019). Barrangou et al. (2007) demonstrated that the CRISPR/Cas system confers resistance to bacteriophage infection and forms adaptive immunity. Subsequent studies showed that the gene encoding the Cas9 protein, which cuts the target DNA or RNA sequences, is located near the CRISPR locus (Khadempar et al., 2019).

Based on the sequence and structure of Cas proteins, CRISPR/Cas systems can be divided into three types. The type II system is the simplest among them, as a single Cas9 protein is needed to interfere with invading genetic elements, whereas the type I and III systems are not used for genome engineering due to their complexity (Hryhorowicz et al., 2017). Three components are necessary for the functional CRISPR/Cas9 system: Cas9 protein, CRISPR RNA (crRNA), and transactivating crRNA (tracrRNA), which is important for crRNA maturation and formation of the complex with Cas9 protein. Cleavage of the foreign nucleic acid is carried out in three stages: CRISPR acquisition (invading DNA is cleaved by a Cas protein into smaller protospacer fragments, which are then incorporated into the CRISPR locus as a new spacers), crRNA biogenesis (the CRISPR locus is transcribed into a CRISPR RNA), and interference with the invading DNA (the mature crRNA recognises the complementary foreign sequence and

directs the Cas9 nuclease, which then degrades the DNA of the invading phage) (Hryhorowicz et al., 2017).

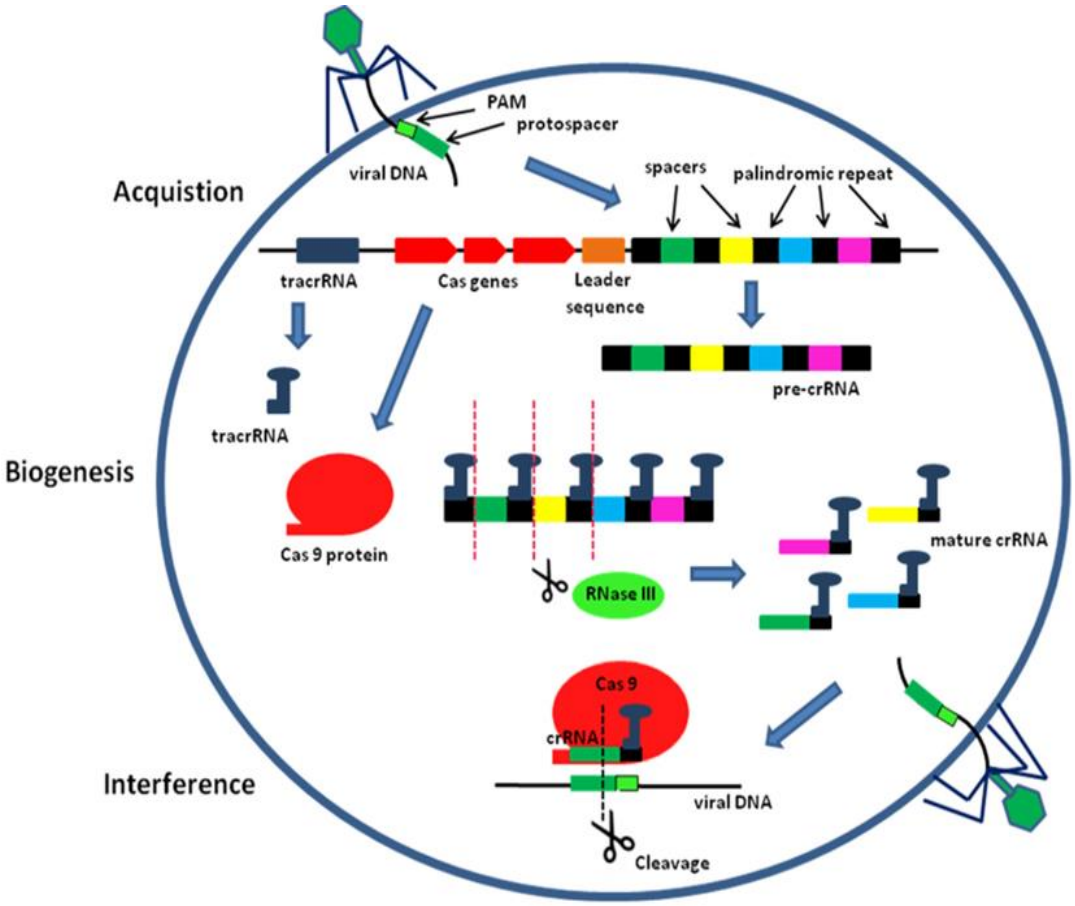


Figure 9. CRISPR/Cas bacterial adaptive immune system functions in three stages: acquisition, crRNA biogenesis, and interference with viral DNA (Hryhorowicz et al., 2017).

Recently, engineered nucleases, such as zinc finger (ZFN), transcription activator-like effector (TALEN) and CRISPR-associated (Cas) nucleases have been used to edit genomes of various organisms. They introduce DSB at the exact location in the genome, which is then repaired by non-homologous end joining (NHEJ) or homology-directed repair (HDR). If homologous template is not provided, the break is repaired by NHEJ, resulting in insertions and/or deletions (indels) due to its error-prone nature. Indels can lead to frameshift mutations that result in premature stop codons and translation termination, transcript degradation, and protein absence. HDR operates if donor template with homology to the target site is present, enabling introduction of target mutations (Ma, Zhang and Huang, 2014). The donor template can be a ssDNA or a dsDNA of various lengths and topologies (linear or circular) (Prill and Dawson, 2020; Yang et al., 2020). Nucleases previously used for genome engineering, such as ZFNs and TALENs, are a complex of the non-specific DNA cleavage domain of the FokI restriction endonuclease and sequence-specific DNA binding domain. When using ZFNs and TALENs, a new DNA-binding domain must be designed for each target site, making these

techniques expensive and time-consuming. In contrast, for the CRISPR/Cas9 system, only appropriate guide RNA needs to be designed to target a new locus. In addition, the Cas9 protein can be introduced together with multiple guide RNAs, allowing different sites in the mammalian genome to be edited simultaneously (Hsieh-Feng and Yang, 2020).

The CRISPR/Cas9 system is widely used for genome engineering in various cell lines and organisms due to its simplicity. In the type II CRISPR system, crRNA and tracrRNA interact to form a duplex that can be replaced by a single guide RNA (sgRNA). The sgRNA directs the Cas9 protein from *Streptococcus pyogenes* (SpyCas9) to the target locus based on complementarity between sgRNA and the target DNA sequence (Ma, Zhang and Huang, 2014). The Cas9 protein contains HNH and RuvC nuclease domains, that introduce DSB (Jinek et al, 2012) 3 bp upstream of the protospacer adjacent motif (PAM) (Figure 10) (Zhu et al., 2019). The PAM sequence is necessary to distinguish the foreign DNA from the host genome, which does not contain PAM sequence. Different Cas proteins recognise different PAM sequences (Hryhorowicz et al., 2017). The most often used Cas protein is from *Streptococcus pyogenes* (SpCas9) which recognises the NGG sequence as PAM (Doetschman and Georgieva, 2017), while for example Cas12a from the *Lachnospiraceae bacterium* has the sequence 5'-TTTV-3' for PAM (Tran et al., 2021).

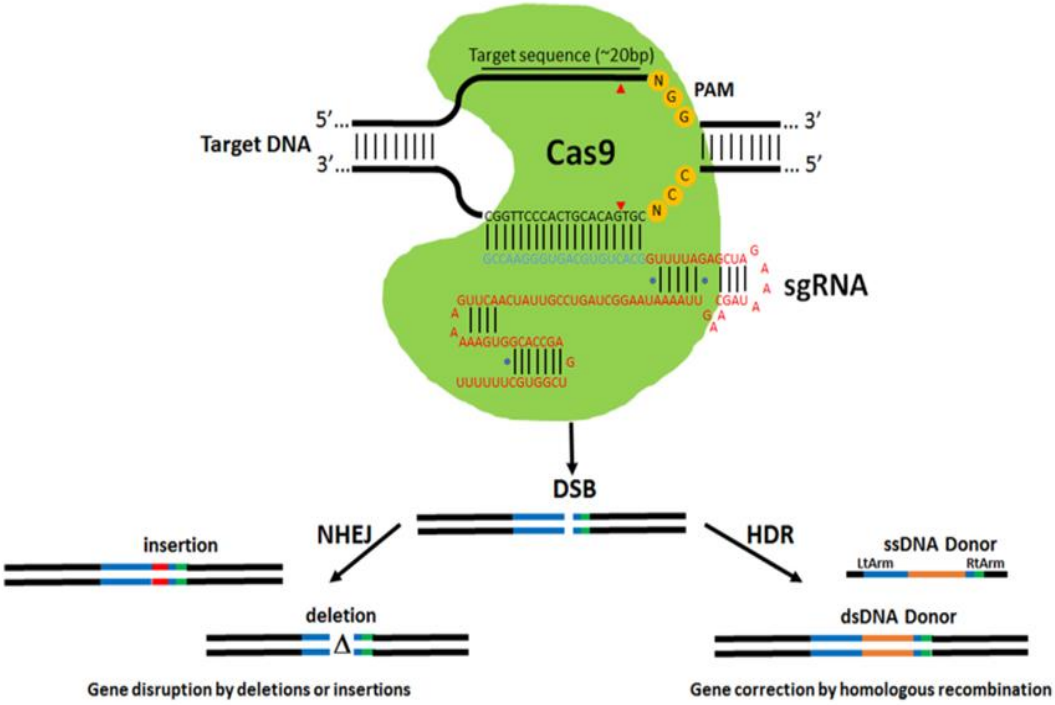


Figure 10. Schematic presentation of the CRISPR/Cas9 system. The Cas9 nuclease (green) is directed to the target sequence by a 20-nt gRNA (blue) and a scaffold (red). The sgRNA binds the target DNA upstream of the PAM sequence NGG (orange circles) and the DSB is introduced 3 bp upstream of the PAM (red triangles). The break is repaired by NHEJ (disruption of the target site by formation of indels) or HDR if a donor template (ssDNA

oligonucleotide with short homology arms (60- 70-bp) or a linear or circular dsDNA plasmid with long homology arms (1 - 3 kb)) is provided (Doetschman and Georgieva, 2017).

Although CRISPR-Cas9 technology is frequently used for genome editing in various biological systems, it is not yet used in medicine due to some unavoidable problems. The main challenge is off-target activity, which is a consequence of base pair mismatches between the target sequence and the gRNA and results in Cas9 introducing DSBs at sites that are partially complementary to the gRNA (Zhang et al., 2016). The amount of CRISPR/Cas9 components plays a critical role in preventing off-target mutations. Reducing the amount of Cas9 considerably reduces off-target effects, but unfortunately also reduces the efficiency of targeted cleavage (Hsu, Lander and Zhang, 2014).

The delivery methods of CRISPR/Cas components can be diverse. Cas9 and sgRNA can be expressed from plasmid or viral vectors as mRNA or DNA. The recombinant Cas9 protein can be transducted into cells or microinjected into zebrafish embryos. CRISPR/Cas9 reagents can also be introduced into cells and zygotes by electroporation. The most prevalent delivery systems for human cells are non-integrating adeno-associated viruses (AAV) and lentiviral vectors. AAVs are challenging due to their low packaging capacity. In addition, if the goal is to introduce a target mutation by HDR, separate AAV vectors must be created that contain Cas9, sgRNA(s), and a donor template (Doetschman and Georgieva, 2017).

Zebrafish was the first vertebrate model to demonstrate that CRISPR/Cas9 can be used *in vivo* for genome editing with up to 50% efficiency (Hwang et al., 2013). Jao et al (2013) showed that CRISPR/Cas9 can efficiently induce biallelic mutations when CRISPR components are injected into zebrafish embryos at the one-cell stage (Jao, Wente and Chen, 2013). Shawn Burgess and co-workers targeted 89 genes with a 99% success rate and a germline transfer rate of ∼28%, which is four to five times higher compared with ZFNs and TALENs (Varshney, Sood and Burgess, 2015). However, the efficiency of introducing mutations into target sites by the CRISPR-Cas9 system rarely reaches 100%, resulting in mosaic mutant cells during zebrafish embryonic development (Gagnon et al., 2014; Burger et al., 2016).

In animal and mammalian cells, NHEJ is the predominant repair pathway, and DSBs induced by Cas9 are only partially repaired by HDR (Symington and Gautier, 2011). Introduction of target mutations is inefficient in non-dividing cells that are in G1 or early S phase of the cell cycle because HDR is active during S/G2 phases. In contrast to HDR, NHEJ is active in all phases of the cell cycle (Pawelczak et al., 2018; Liu et al., 2019). Therefore, in non-proliferating cells, NHEJ is the primary repair pathway that introduces frameshift mutations, generating gene knockouts. However, many mutations remain in-frame, resulting in low knockout efficiency (Guo et al., 2018).

Targeted insertion (knockin) of DNA fragments using HDR is a promising method for determining gene function. Several studies have reported poor efficiency of HDR in genome

editing, especially *in vivo* at the organism level (Shin, Chen and Solnica-Krezel, 2014). HDR efficiency was tried to be improved by using different types of donor templates, stimulating HDR, and suppressing the NHEJ pathway. Results from Nüsslein-Volhard's laboratory showed that a donor template in the form of a plasmid increased HDR efficiency to 46% *in vivo.* It is possible that ssDNA donors form secondary structures that inhibit DNA recombination and thus reduce HDR efficiency. Another limitation of producing knockin lines *in vivo* is the low efficiency of mutation transmission to the next generation via the germline (Zhang, Zhang and Ge, 2018). This challenge can be circumvented by using donor templates with 1 Kb long homology arms, surrounding the DNA insert and flanked by I-SceI meganuclease restriction sites. This knockin approach, using long homology sequences and digestion of the donor template with I-SceI, increases the efficiency of germline transmission (Hoshijima, Jurynec and Grunwald, 2016).

1.8. Zebrafish (*Danio rerio*) as a model organism

The zebrafish *(Danio rerio*) is a small freshwater teleost fish of the family *Cyprinidae* indigenous to South Asia that has been used as a vertebrate model since the 1960s. It initially became a popular laboratory model in developmental biology due to the extrauterine development and optical clarity of the embryos (Rahman Khan and Sulaiman Alhewairini, 2019). The advantages of the zebrafish model also include high fecundity, quick maturation (larval stage is reached up to 72 hours post-fertilization), and relatively easy maintainance and genetic manipulation. Following the development of techniques to facilitate genetic analysis in zebrafish, it has been used to study genetics, gene function, and human disease (Bradford et al., 2017).

Although the zebrafish is a primitive vertebrate, it has many other advantages over other model animals such as the laboratory mouse. For example, ovulation is controlled by light, spawning is frequent, microinjection of fertilized eggs is relatively cheap and easy, the genome is assembled at high resolution, and 70% of genes are conserved between zebrafish and humans, including all genes involved in DNA repair. Zebrafish embryos are transparent and are therefore often used to observe the development of tissues and organs *in vivo*. Their maturation time is 2.5 - 3 months, similar to mice. The F0 modified line is created when an exogenous transgene or mutation is successfully transferred to embryos and passed on to the F1 generation via germline. Fish that transfers transgene or mutation to the next generation through the germline is called the founder fish (Figure 11) (Kong, Cheng and Yu, 2016).

Many genome editing techniques, including knock-in, knock-down, and knock-out, are well developed in zebrafish, making it an excellent model organism for cancer research, the study of vertebrate-specific developmental biology *in vivo*, biomedical studies, and the discovery of molecular mechanisms leading to various diseases of the vascular, hematopoietic, immune and central nervous systems in humans. In addition, zebrafish embryos are used as models for toxicological studies, developmental biology, and carcinogenesis (Kong, Cheng and Yu,

2016). Nevertheless, there are some limitations in using zebrafish as a model organism, including absence of some human organs such as cardiac septum, lung, mammary gland, prostate, limbs, and placenta. However, it is not always easy to generate knockout lines for specific genes due to duplicates of several genes in the zebrafish genome (Lin et al., 2016).

Zebrafish breed in small groups, usually at dawn, because reproduction is strongly controlled by photoperiod. Females produce hundreds of eggs that develop into free-swimming larvae within 4–7 days (Lawrence, 2007). The zebrafish belongs to the teleost lineage in which whole genome duplication occured in the late Devonian, resulting in multiple orthologs for some mammalian genes (Taylor et al., 2003). These duplicated genes may have the same functions, one of them may lose its original function and gain a new function, or they may have undergone subfunctionalization (Bradford et al., 2017).

Zebrafish species do not have X and Y sex chromosomes, but sex is determined by yet unknown gene clusters and the WZ/ZZ sex determination system. Studies have shown that zebrafish chromosomes 3, 4, 5, and 16 contain a putative sex-linked loci (Howe et al., 2013). Indeed, researchers found that wild zebrafish contain a locus in the telomeric region of chromosome 4 that controls sex development, whereas it was absent in most laboratory strains, likely due to continuous breeding and mutations that resulted in the loss of this sex-linked region (Wilson et al., 2014). Domesticated zebrafish are thought to have only a few sex-linked loci instead of one sex-determining gene, and sex is determined by the influence of both genetic factors and the environment. It has been observed that in some fish species the number of individuals in the same space influences the sex ratio, including the paradise fish, *Macropodus opercularus*, some coral reef fish species, *Anguilla anguilla*, the European sea bass, *Dicentrarchus labrax* and the zebrafish. However, a critical reduction in the number of individuals also has some negative effects, such as reduced growth, lower survival, developmental deformities, and changes in reproduction (Ribas et al., 2017; Li et al., 2021).

In the last decade, editing techiques such as ZFNs, TALENs, and CRISPR/Cas9 have been successfully used in zebrafish to mutate target genes to study genetics, developmental biology, toxicology, and drug testing (Li et al., 2021). Compared with ZFNs and TALENs, CRISPR/Cas has many advantageous features, including the simple design of DNA- binding sequences (sgRNAs), relative ease of use, and the ability to mutate several genes simultaneously, making it the most amenable approach for high-throughput mutagenesis. In contrast, ZFNs or TALENs are not suitable for high-throughput mutagenesis because a new DNA-binding domain must be designed for each target gene. In addition, there are many softwares and websites that facilitate the use of the CRISPR/Cas9 system in zebrafish and serve to select sgRNAs with minimal off-target activity, such as CRISPR MultiTargeter, CRISPRdirect, CCTop, CHOPCHOP, sgRNAcas9, and CRISPRscan. Hwang et al (2013) first performed CRISPR/Cas9 gene knock-outs in zebrafish, with rates of somatic mutagenesis ranging from 24 to 59% at 10 loci (Liu et al., 2019). Although the CRISPR/Cas9 system is commonly used for genome editing in zebrafish, and mutagenesis efficiencies typically range from 17% to 90% (Hwang et al., 2013; Jao, Wente and Chen, 2013; Gagnon et al., 2014), high-

throughput methods for generating stable mutants have not yet been developed. Varshney et al (2015) mutated two different loci for 83 genes with a 99% success rate, while the germline transmission rate was 50%. Although on average one out of every two injected fish was a founder, the rates for the different target genes were variable and ranged from 0% to 100%. In the F1 generation, more than 50% of the fish carried a mutation (Varshney, Sood and Burgess, 2015).

The CRISPR/Cas9 knockout method is very efficient in zebrafish, with rates ranging of 75% to 99%, whereas the HDR-mediated knock-in method is less efficient. This is not surprising since the NHEJ pathway is more active during early zebrafish development than HR (Liu et al., 2019). Another challenge related to HDR-dependent mutagenesis is low germline transmission, with only 5 - 15% of injected individuals transmitting the mutation to the next generation (Hoshijima, Jurynec and Grunwald, 2016). Hruscha et al (2013) knock-in a hemagglutinin (HA) tag with short homologous arms (30 - 50 bp) with a succession rate of 3.5%, which is very low for somatic modifications that may not be transmitted through the germline (Hruscha et al., 2013). In addition, studies show that both precise and imprecise knock-in events can occur when single- stranded oligonucleotides with short homology arms (up to 50 bp) are used as donor templates (Hruscha et al., 2013; Hwang et al., 2013). Hoshijima et al (2016) presented a comprehensive and efficient approach for knock-in in zebrafish. In their study, the mutation was transmitted by on average 6.3% of F0 gametes. Efficient knock-in events were observed when plasmid DNA was used as the donor template instead of linear DNA, likely because circular DNA can be injected at higher concentrations and is linearized for integration by Cas9 once injected into embryos (Varshney, Sood and Burgess, 2015).

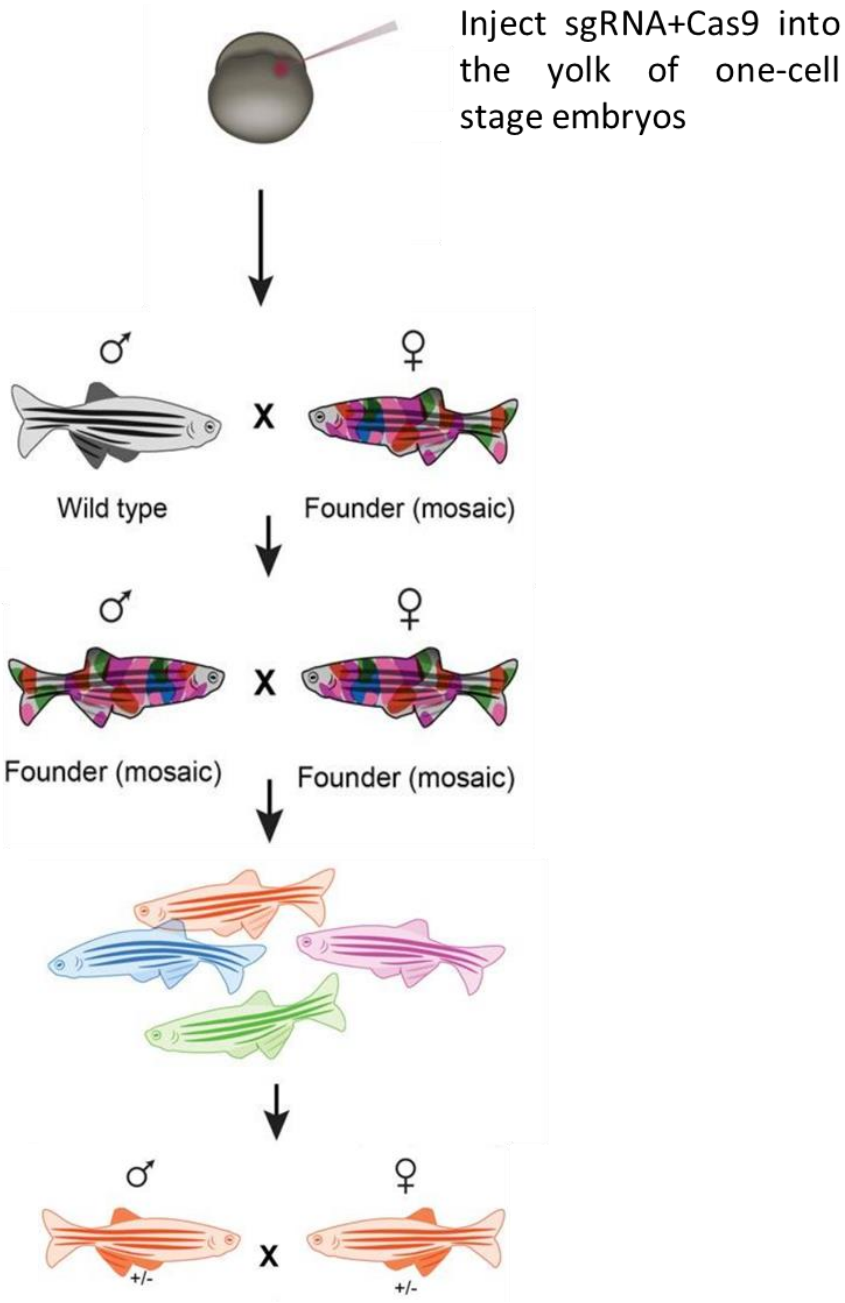


Figure 11. Schematic presentation of CRISPR/Cas9 gene editing in zebrafish. sgRNA and Cas9 protein are injected into the yolk of one-cell stage embryos. The injected embryos are raised and the founder fish is determined by outcrossing with wild type fish. Female and male F0 founder fish are crossed to produce heterozygous F1 generation. Heterozygous F1 mutants are determinated by sequencing and crossed for the F2 generation in which 25% of the offspring are homozygous (adapted from Varshney, Sood and Burgess, 2015).

Morpholino antisense oligomers are used for efficient gene silencing in a number of model organisms, including *Xenopus*, zebrafish, sea urchin, and chicken. Morpholinos are chemically synthesized DNA analogs that are injected into embryos at the one-cell stage, where they

bind complementary target mRNAs and prevent their translation or alter splicing (Nan & Zhang, 2018). Both types of morpholinos inhibit translation of zygotic transcripts, whereas splice-blocking morpholinos are ineffective on maternally inherited transcripts which are already mature (Moulton, 2017).

They were developed as a synthetic derivative of DNA with the deoxyribose ring replaced by a six-membered morpholine ring and the anionic phosphodiester bond replaced by a non-ionic phosphorodiamidate bond. Because of their chemistry, they are stable and are not recognized and degraded by RNAse H. Because of their neutral charge and relatively small size (usually 25 bases), morpholinos are distributed throughout the embryo by diffusion after microinjection. One of the disadvantages of using morpholinos is their off-target activity. However, they offer higher specificity compared to RNAi technology. It has been shown that siRNAs can not only silence the expression of the target gene, but also repress the translation of other genes with partial complementarity, as only 7 bases of complementarity are needed for recognition. Morpholino oligomers require about 15 bases of complementarity with target genes to knock-down gene expression. Morpholinos can be designed to bind at or near the translation initiation site of an mRNA to interfere with translation of specific gene targets. Translation-blocking morpholinos sterically block translation by binding a sequence in the vicinity of the start codon and preventing the 40S ribosomal subunit from scanning the 5'UTR and finding the start codon. They can also be used to disrupt RNA splicing by binding to pre-mRNA splice sites and disrupting mRNA maturation (Bestman and Cline, 2020). Splicing is regulated in eukaryotes by small nuclear ribonucleoproteins (snRNPs) that bind to intron-exon boundaries. Morpholinos targeting intron-exon boundaries can alter splicing, by either preventing splicing and causing intron inclusion or exon excision. Targeting the splice sites of internal exons usually results in exon excision, leading to an mRNA missing the exon with the blocked splice site, while targeting the splice sites of the first or last exon causes intron inclusion, resulting in an mRNA containing the first or last intron. Blocking a splice site can activate a cryptic splice site or lead to a double exon skip, resulting in an mRNA with an unexpected mass. Splice modifications can cause frameshift mutations or inclusion of the intron in the mature mRNA. Any of these outcomes can lead to protein truncation due to the appearance of premature stop codons, inhibition of translation due to the appearance of a miRNA binding site, and nonsense-mediated mRNA decay (Moulton, 2017). The advantages of using morpholinos are stability, nuclease-resistance, efficacy, long-term activity, water-solubility, low toxicity, specificity, and the ease of use. Conventional genome editing methods are well established for zebrafish. Although these methods are efficient, large-scale genetic analysis in vertebrates is costly and time-consuming because many fish must be bred, maintained, and analysed over multiple generations. Morpholinos are commonly used in zebrafish as knock-down tool because of their ease of administration and high efficacy during zebrafish embryonic and larval development. Zebrafish embryogenesis is complete at 50 hpf with nearly all vertebrate organ systems developed. Morpholinos are effective and stable after microinjection into embryos up to 50 hpf, although dilution by cell growth and division reduces their activity and efficacy.

1.9. Summary and aims

In summary, DNA-protein crosslinks (DPCs) are one of the most common forms of DNA damage caused by irreversible covalent binding of proteins to DNA in the presence of cytotoxic, mutagenic and carcinogenic compounds. DPCs present a physical blockage to all essential DNA functions (replication, transcription, recombination), and if left unrepaired, lead to mutations, genomic instability, and / or cell death. DPC repair is specialized DNA repair pathway in eukaryotes in which DNA-dependent proteases associated with replication are involved. The presence of DNA-dependent metalloprotease SPRTN (or DVC1) in humans has recently been discovered. SPRTN is part of the replicome and removes DPCs in front of the replication fork. SPRTN proteolytically cleaves DPCs into smaller DNA-bound peptide residues which are removed by unknown factors acting downstream of SPRTN protease. Removal of smaller DPCs by the Nucleotide Excision Repair patway (NER) has been shown only *in vitro* and in cell lines, however, it is not clear whether NER acts in conjunction with the SPRTN protease. In addition to SPRTN, higher eukaryotes are thought to have other potential DPC protease ACRC. However, it is unknown if ACRC is proteolytically active and whether it has a role in DPC repair.

Considering that majority of knowledge on DPC repair has been obtained using mammalian cell lines, *in vivo* studies are still scarce. In spite of a widely recognized cytotoxic effects and and the risk they pose to genome integrity, pathways involved in their repair and their regulation are poorly understood and studied.

Taking into account the described deficiencies and gaps in knowledge, the main goal of this study was to contribute to understanding of the repair of DPCs at the cellular level and *in vivo* at level of the organism using the zebrafish (*Danio rerio*) as a model and CRISPR/Cas genome editing. Furthermore, the aim was characterization of Xpa and Acrc proteins in zebrafish, which are assumed to be involved in DPC repair.

To accomplish the described general goal, several specific goals were defined:

1) To establish phylogenetic relationships between zebrafish and human *XPA* proteins

2) To perform conserved synteny analysis in order to determine chromosomal locations of zebrafish *xpa* and *acrc* genes and to determine level of gene order preservation

3) To determine expression profile of *acrc* and *sprtn* genes in adult zebrafish and during zebrafish embryonic development to identify tissues and stages in which these proteins are particularly important

4) To generate ACRC and XPA deficient models

5) To study the role of ACRC and NER pathway in DPC repair and to optimize methods for DPC isolation and detection in zebrafish.

2. Materials and Methods

2.1. Materials

Standard chemicals used in this study are listed in Table 2.1, enzymes in Table 2.2, commercial kits in Table 2.3. and primers used for quantitative PCR (qPCR) and High-Resolution Melting (HRM) analysis in Table 2.4. and 2.5. The sequences of the guide RNAs used to generate zebrafish knockout strains are listed in Table 2.6. The small interfering RNAs (siRNAs) that were used to silence target genes in RPE1 cells are listed in Table 2.7. Antibodies used for Western and slot blotting analysis are listed in Table 2.8. Morpholinos used for XPA and SPRTN silencing in zebrafish embryos are listed in Table 2.9.

Competent DH5α *E. coli* cells (Life Technologies, CA, USA) were used for genotyping by cloning and sequencing. These cells were grown on agar plates (Sigma-Aldrich, Taufkirchen, Germany) or in liquid Luria-Bretani medium (Becton, Dickinson and Company, Sparks, USA) supplemented with 100 µg/ml ampicilin (Sigma-Aldrich, Taufkirchen, Germany).

Human embryonic kidney cells (HEK293T) (ATCC, CRL -1573) were selected because of its short amplification time (< 24 h) and high transfection efficiency (Tom et al, 2008). Human retinal pigment epithelium-1 (RPE-1) cells were used in all experiments for DPC isolation by the RADAR assay. This cell line was selected because it is a non-transformed alternative to cancer cell lines and has low endogenous levels of DNA damage and DPCs. Cells were cultured in DMEM-FBS medium, a Dulbecco's Modified Eagle Medium (DMEM) with high glucose content (Life technologies, CA, USA) and 10% fetal bovine serum (FBS) (Lonza, Basel, Switzerland), at 37°C and an atmosphere containing 5% CO2.

Table 2.1. Standard chemicals used in the study.

Chemical	Source
Acrilamide/bis-acrilamide	Sigma-Aldrich, Germany
APS (ammonium persulfate)	Sigma-Aldrich, Germany
β-mercaptoethanol	Sigma-Aldrich, Germany
BSA (bovine serum albumin)	Carl Roth, Germany
Ethanol	Kemika, Croatia
DharmaFECT tranfection reagent	Dharmacon, USA
Opti-MEM	Gibco, USA
Agarose	Sigma-Aldrich, Germany
DNase/RNase-free water	Invitrogen, USA
SDS	Sigma-Aldrich, Germany
TEMED	Sigma-Aldrich, Germany
GeneRuler DNA ladder mix	ThermoFischer Scientific, USA
Precision plus protein ladder	Bio-Rad Laboratories, USA
Clarity Western ECL Substrate	Bio-Rad Laboratories, USA
Trypsin-EDTA	Sigma-Aldrich, Germany

Table 2.2. Enzymes used in the study.

Enzyme	Source
Proteinase K	ThermoFischer Scientific, USA
Cas9 endonuclease	New England Biolabs, USA
Phusion polymerase	New England Biolabs, USA
Power SYBR Green PCR Master mix	Applied Biosystems, USA
MeltDoctor HRM Master mix	Applied Biosystems, USA
Reverse transcriptase	New England Biolabs, USA
T4 DNA ligase	New England Biolabs, USA

Table 2.3. Commercial kits used in the study.

Kit	Source
Monarch Total RNA Miniprep kit	New England Biolabs, USA
Monarch RNA Cleanup Kit	New England Biolabs, USA
ProtoScript II First Strand cDNA Synthesis kit	New England Biolabs, USA
Monarch PCR & DNA Cleanup Kit	New England Biolabs, USA
Zyppy Plasmid Miniprep Kit	Zymo research, USA
MEGAshortscript T7 Transcription Kit	ThermoFischer Scientific, USA
MeltDoctor HRM Master Mix kit	ThermoFischer Scientific, USA
CloneJET PCR Cloning Kit	ThermoFischer Scientific, USA
Power SYBR Green PCR Master Mix	ThermoFischer Scientific, USA
Quant-iT 1X dsDNA HS Assay	ThermoFischer Scientific, USA
ProteoSilver Silver Stain Kit	Sigma-Aldrich, Germany
In-fusion cloning kit	Takara, Japan
Hiscribe SP6 RNA kit	New England Biolabs, USA
HiScribe T7 ARCA mRNA kit	New England Biolabs, USA

Table 2.4. Primers used for qPCR.

Protein name	Primer sequence
DrSPRTN	F 5' ATTCCCTTCAGTGGCAGAGG 3' R 5' GAGGTTCTGGTGGCGCTTTA 3'
DrACRC	F 5' ACCCAAACCACAACGTCCTT 3' R 5' ACTGGCGTGTGGATTACAGG 3'
DrEF1α	F 5' TGATGCCCTTGATGCCATTCT 3' R 5' CACGACCCACAGGTACAGTT 3'
HsSPRTN	F 5' GAGGTGGATGAGTATCGGCG 3' R 5' GGGTTCCCTGTTAGTAGCTCG 3'
HsXPA	F 5' AGCGGGCACTGATGCTG 3' R 5' ACATTAGCCATGCCTCCAGTA 3'
HsACRC	F 5' TGATGATGCTGGTGAGCAGG 3' R 5' TCCTCAGTTGGCAGCTTTCT 3'

HsATP50	F 5' ATTGAAGGTCGCTATGCCACAG 3' R 5' AACAGAAGCAGCCACTTTGGG 3'

Table 2.5. Primers used for High-Resolution Melting (HRM) Analysis.

Protein name	Primer sequence
DrXPA	F 5 'TGAAGTAAAGCACAAGCTAATATCTC 3' R 5' TGTCTTTAAGTAGAGTTTCATGTCTCC 3'
DrACRC	F 5' TGCGGCCACATGGCTCATAAAC 3' R 5' GAAACCATCGGGAGCTCAGGATG 3'

Table 2.6. Guide RNAs used for creating zebrafish gene knockouts.

Protein name	Guide RNA sequence
DrXPA sgRNA	5' GGAGGCCAAGGAGACTAGAG 3'
DrACRC sgRNA 1	5' GGAGCATAAAGCCTCCAGAA 3'
DrACRC sgRNA 2	5' GGCCGCATGACACATTTCA 3'

Table 2.7. Small interfering RNAs (siRNAs) used in the study.

Target gene	Name
HsSPRTN	ON-TARGET plus SPRTN siRNA 1 Silencer Select SPRTN siRNA 2
HsACRC	Silencer Select ACRC siRNA 1 Dharmacon smartpool ACRC siRNA 2
HsXPA	Dharmacon siGENOME XPA siRNA 1 ON-TARGET plus XPA siRNA 2

Table 2.8. Antibodies used for Western blot and slot blot analysis.

Antibody	Host	Producer	Cat.No.
Mouse Anti-ACRC monoclonal primary antibody	Mouse	Sigma-Aldrich, Germany	08235-2E4
Mouse Anti-ds DNA monoclonal primary antibody	Mouse	Santa Cruz Biotechnology, USA	sc-58749
Mouse Anti-Histone H1 monoclonal primary antibody	Mouse	Santa Cruz Biotechnology, USA	sc-8030
Rabbit Histone H3 polyclonal primary antibody	Rabbit	Cell Signaling Technology, USA	CST-9715S
Anti-Mouse IgG polyclonal secondary antibody	Rabbit	Sigma-Aldrich, Germany	A9044
Anti-Rabbit IgG polyclonal secondary antibody	Goat	Sigma-Aldrich, Germany	A0545

Table 2.9. Morpholinos used for zebrafish *xpa* and *sprtn* silencing.

Name	Sequence
sprtn MO	5' AGAGAGGCATATTTAACCAACCTGA 3'
xpa MO	5' AGAGTGAATATGACATACCTGCATT 3'

2.2. Methods

2.2.1. Phylogenetic and syntenic analysis

Nucleotide and protein sequences were retrieved from the following NCBI (http://www.ncbi.nlm.nih.gov/) and ENSEMBL (http://www.ensembl.org/index.html) databases, respectively. Blastx algorithm was used. Sequences were aligned with MUSCLE algorithm (Edgar, 2004) and phylogenetic tree was constructed using Maximum Likelihood method in PhyML 3.0.1 software (Guindon and Gascuel, 2003). Orthology predictions using conserved synteny analysis between zebrafish and humans for genes of interest were made using Genomicus (http://www.genomicus.biologie.ens.fr/genomicus), a conserved synteny browser synchronized with genomes from the Ensembl database (Louis, Muffato and Crollius, 2013).

2.2.2. Analysis of gene expression in adult zebrafish tissues and during embryonic development (qPCR)

Adult zebrafish of both sexes, strain AB, were obtained from the European Zebrafish Resource Center (https://www.ezrc.kit.edu/) at the Karlsruhe Institute of Technology (Germany) and sacrificed for tissue collection by immersion in ice-cold water for 30 min. Three independent pools of different tissues were collected and the tissues were homogenized in RIPA buffer (NaCl 150 mM, EDTA 1 mM, Tris 25 mM, NP-40 0.8%) using a rotor-stator homogenizer at 10,000 rpm for 20s. After lysis, samples were centrifuged at 15,000g for 5 min (4°C) and the supernatant was used. Zebrafish embryos were obtained by crossing male and female wild-type fish and collected at 1, 4, 6, 12, 24, 48, and 72 hours post fertilization (hpf). Total RNA isolation from zebrafish tissues and embryos was performed using the Total RNA Miniprep Kit. RNA was quantified using the Bio-Spec Nano Spectrophotometer (Shimadzu Corporation, Kyoto, Japan) and RNA integrity was determined by agarose gel electrophoresis. The purified total RNA was reverse transcribed (1 µg of total RNA) using the ProtoScript II First Strand cDNA Synthesis Kit. For qPCR, specific primers were designed using Primer Blast software (National Center for Biotechnology Information, Rockville Pike, USA), manually adjusted as needed, and purchased from Macrogen Service (Amsterdam, The Netherlands). Quantification of zebrafish gene expression was performed using the qPCR method of relative quantification. Zebrafish genes were normalized to the housekeeping gene elongation factor 1α *(ef1α)*.

The relative quantification method is described with equation:

MNE = (primer efficiency of housekeeping gene ^Ct value of housekeeping gene) / (primer efficiency of target gene^Ct value of target gene) x 10^6

Expression was classified as very high for MNE*10^6 > 11700 (Ct < 22), high for MNE*10^6 = 2500-11700 (Ct =22-25.9), moderate for MNE*10^6 = 800-2500 (Ct = 26-30), and low for MNE*10^6 < 800 (Ct > 30). The qPCR was performed using ABI PRISM 7000 Sequence Detection System and Power SYBR Green PCR Master Mix. The qPCR reaction mix was prepared to a final volume of 10 µL and contained: 5 µL of the Power SYBR Green Master Mix, 0.5 µL of each

primer, 1 µL of template (10 ng/sample), and 3 µL ultrapure water (Molecular Bioproducts, San Diego, CA, USA). After initial denaturation at 95˚C for 10 min, 40 amplification cycles were performed with denaturation at 95˚C/15 s, annealing and elongation at 60˚C/1 min, all together followed by melting curve analysis. Data were analysed using GraphPad Prism software version 5.00.

2.2.3. Creation of gene knockouts using CRISPR/Cas9 system

To investigate the role of the NER pathway and ACRC protease in DPC repair, mutant zebrafish strains were generated using CRISPR/Cas9 genome editing. We have identified a putative active site (E451) in zebrafish that is targeted for the production of the enzymatically inactive Acrc protein. The single-guide RNAs (sgRNAs) specific for the *xpa* and *acrc* genes were identified by CRISPRscan (http://www.crisprscan.org/), and those with a high CRISPRscan score and no off-targets were selected. After selecting the guide with the best parameters, the guide sequence was inserted between two scaffold sequences containing the T7 promoter (GGATCCTAATACGACTCACTATAG) and the sequence important for annealing the rest of the guide (GTTTTAGAGCTAGAA). The DNA template for *in vitro* transcription of the sgRNA was generated using phusion PCR by preparing the following reaction mixture: 5 µL 5x phusion HF buffer (New England Biolabs, USA), 0.5 µL 10 mM dNTPs, 0.5 µL 1 µM IVT-FW primer (forward primer for PCR amplification of annealed guide DNA before *in vitro* transcription, contains the T7 promoter), 0.5 µL 1 µM IVT-REV primer (reverse primer for PCR amplification of annealed guide DNA), 0.25 µL 100 µM IVT-sgRNA primer (sequence depends on guide sequence), 0.25 µL IVT-SCAFFOLD primer (constant oligonucleotide with tracrRNA sequence for amplification of guide DNA), 0.25 µL phusion polymerase (2U/µL), and 17.75 µL nuclease-free water. The reaction mixture was incubated at 95°C for 2 minutes, followed by 30 cycles of 95°C/10 s, 57°C/10 s, 72°C/10 s, and 72°C for 2 min. The presence of the DNA template was confirmed by assaying 200 ng of the PCR product on agarose gel. The expected product size is 127 bp. The remainder of the PCR mixture was purified using the DNA/PCR clean up kit. The DNA was transcribed into RNA using the MEGAshortscript T7 Transcription Kit by incubating 150 nM PCR amplicon at 37°C/4 h. Finally, the product was purified by RNA clean up kit and the efficiency of *in vitro* transcription reaction was confirmed on 2% agarose gel. After synthesis of the sgRNAs, they were injected in complex with the Cas9 protein into the yolk of zebrafish embryos at the one-cell stage to generate a mutant Xpa and Acrc strains. For microinjection into zebrafish embryos, 1 µL of phenol red (0.05%), gRNA (180 ng/µL), Cas9 protein (600 ng/µL), and KCl (300 mM) were mixed. The mixture was introduced into a thin needle and 1 nL of the mixture was injected using a microscope. The generated mutated strain of zebrafish was reared and crossed with wild-type individuals after reaching sexual maturity to identify the individual that passes the acquired mutation to the offspring (founder). After 48h, 15 embryos obtained by crossing the injected F0 generation with wild-type individuals were collected for DNA extraction and genotyping of the F0 generation. As a control, 5 WT embryos at the same developmental stage were used. Genomic DNA was isolated using embryo digestion buffer (Tris-Cl 10 mM, 50 mM KCl, Tween-20 30%) and proteinase K (0.2

mg/mL). Genotyping was performed using the StepOnePlus™ Real-Time PCR System (ThermoFischer Scientific, MA, USA) and High-Resolution Melting Analysis software (ThermoFischer Scientific, MA, USA). The presence of the acquired mutation was also confirmed by sequencing. Individuals obtained from different breeds and carrying a mutation in the target gene were crossed with each other to obtain an F1 generation. Individuals of the F1 generation were genotyped by analyzing genomic DNA obtained by lysis of tail fin tissue by PCR and sequencing to find homozygous male and female for the target mutation. Homozygous individuals were then crossed to obtain homozygous F2 progeny, in which the contribution of the studied proteins in the repair of DPCs was analyzed.

2.2.4. Genotyping using High-Resolution Melting (HRM) Analysis

HRM analysis is a post-PCR analysis method used to identify genetic variations in nucleic acid sequences. HRM analysis begins with a PCR reaction of the target region in the presence of a fluorescent dye that binds dsDNA. After amplification, a high-resolution melting step occurs in which the dsDNA dissociates into a single strand and the fluorescence changes according to release of the dye. As the temperature increases, denaturation of the double-stranded DNA occurs and the dye is released, resulting in a fluorescence decrease. The greatest decrease in fluorescence is observed near the melting temperature (T_m) of the PCR product. The T_m is a temperature at which 50% of the DNA is double-stranded and 50% is single-stranded (melted), and it depends on the properties of the PCR product (GC content, length, and sequence). The result of the analysis is a melting curve profile specific for the amplicon, which allows mutation screening, genotyping, methylation and other applications. This method was used to genotype the injected zebrafish embryos as well as F1 and F2 generation adults to identify founders. Primers were designed using Primer- BLAST software (NCBI, National Center for Biotechnology Information) to give a 100 - 200 bp product around the gRNA target site in genomic DNA. The reaction mixture consisted of 5 µL commercial MeltDoctor™ HRM Master Mix Kit, 0.6 µL forward and reverse primer (5 µM), 0.5 µL genomic DNA, and 3.3 µL ultrapure water. The 96-well microplate containing the prepared reaction mixtures was centrifuged (5 min, 4°C, 500 g). The reaction was incubated at 95°C/10 min followed by 40 cycles of 95°C/15 s, 60°C /1 min and 95°C/10 s, 60°C/1 min, 95°C/15s (High Resolution Melting), 60°C/ 15 s using StepOnePlus™ Real-Time PCR System. Results were analyzed using High Resolution Melt Software v3.2.

2.2.5. Genotyping using cloning and sequencing

Individuals that were positive in HRM analysis were also sequenced to confirm the presence of the acquired mutation. The commercial CloneJET PCR Cloning Kit was used for the cloning and sequencing experiment. The first step was amplification of the region of interest using Phusion® High-Fidelity DNA Polymerase the primers used in the HRM analysis. The reaction mixture was prepared by mixing 5 µL 5xPhusion HF buffer, 0.5 µL 10 mM dNTPs, 1.25 µL 10 µM forward and reverse primers, 1 µL genomic DNA, 0.25 µL Phusion DNA polymerase, and

15.75 µL ultrapure water. The reaction mixture was incubated at 98°C/30 s, followed by 35 cycles of 98°C/10 s, annealing for 30 s (temperature depends on the primers used), 72°C/15 s, and 72°C/10 min. Phusion DNA polymerase generates a blunt-ended PCR product that can be used directly for ligation into a linearized vector. The efficiency of the PCR reaction was confirmed by agarose gel electrophoresis, and DNA amplicons were purified from the PCR mixture using the DNA/PCR clean up kit. pJET1.2/blunt is a linearized cloning vector that accepts inserts from 6 bp to 10 kb. This vector contains a lethal gene that is disrupted by ligation of a DNA insert into the cloning site. As a result, only cells with recombinant plasmids can replicate. The vector contains an ampicillin resistance gene, so only transformed cells can survive in the presence of this antibiotic. The recommended amount of PCR product for the ligation reaction depends on the length of the PCR product. The reaction mixture was prepared by mixing 5 ng of the PCR product (size 100 pb), 10 µL of 2x reaction buffer, 1 µL of linearized blunt cloning vector pJET1.2, ultrapure water, and 1 µL of T4 DNA ligase. The mixture was incubated at 22°C for 30 min and then used to transform bacterial competent *Escherichia coli* DH5α cells. 1 µL of the ligation mixture was added to 25 µL of the bacterial cell suspension, and the cells were incubated on ice for 30 min. Cells were then subjected to heat shock (42°C/ 30 s) in a thermoblock and incubated on ice for 2 min. After incubation, 475 µL of LB medium heated to 37°C was added and the cells were placed in an incubator shaker (37°C/1 h, 225 rpm). 250 µL of the bacterial cell suspension was added to heated (37°C) agar plates containing ampicillin and the plates were incubated overnight at 37°C. The next day, bacterial colonies were transferred to the tubes containing LB medium with ampicillin using a plastic micropipette. The tubes were incubated overnight in a shaker incubator (37°C, 225 rpm). The next day, plasmid DNA was isolated from the obtained bacterial cultures in suspension using the commercial Zyppy™ Plasmid Miniprep Kit. The isolated plasmids with integrated DNA amplicon were sequenced using the EZ-seq DNA sequencing service from Macrogen (Amsterdam, The Netherlands). According to the instructions of the above service, 5 µL of a 100 ng /µl isolated plasmid sample was mixed with 5 µL of a 10 µM sequencing primer (pJET1.2 Forward Sequencing Primer: 5' CGACTCACTATAGGGAGAGCGGC 3'). The results were analyzed using BioEdit Sequence Alignment Editor software.

2.2.6. Cell culture

Cells were cultured in DMEM-FBS medium, a Dulbecco's Modified Eagle Medium (DMEM) with high glucose content (Life technologies, CA, USA) and 10% fetal bovine serum (FBS) (Lonza, Basel, Switzerland), at 37°C and an atmosphere containing 5% CO2. Cells were passaged twice weekly, by detaching from the surface with a trypsin-EDTA solution (Sigma-Aldrich, Taufkirchen, Germany) heated to 37°C. After several minutes of incubation in trypsin-EDTA, the reaction was stopped by adding four times the volume of DMEM-FBS. A portion of the cell suspension was returned to the 25 or 75 cm^2 culture flask for cell culture maintenance, while the remainder was used for experiments.

To maintain the stock of cells of early passages, a portion of the cells was frozen in liquid nitrogen (-196°C). Cells were detached from the surface using trypsin-EDTA solution as

described above and collected for freezing by centrifugation at 1100 × g for 5 minutes, resuspended in 900 µL DMEM-FBS, and transferred to freezing vials. 100 µL of DMSO was added to the cell suspension (to protect against the formation of water crystals during freezing) and the cells were briefly incubated on ice until transferred to a rack at -80°C to ensure slow freezing. After 3 days, the vials containing the cells were transferred to liquid nitrogen (-196°C), where they were kept until the next thaw.

The cells were thawed by immersing the vial in a water bath heated to 37°C. They were then transferred to a 75 cm^2 culture flask in 20 mL of the previously warmed DMEM-FBS. The next day, the medium was replaced to completely remove the DMSO.

2.2.7. Gene silencing by transfection of small interfering RNAs

To silence targeted gene expression, HEK293T or RPE1 cells were transfected with specific small interfering RNA (siRNA) using the commercially available reagent DharmaFECT (Dharmacon, USA). Cells (4.2×10^5) were seeded in a 10-cm cell culture dish in 7 mL of antibiotic-free DMEM-FBS so that confluence was 60-80% on the day of transfection. Solution A was prepared by mixing 7 µL siRNA and 553 µL opti- MEM, and solution B was prepared by mixing 19 µL Dharmafect transfection reagent and 541 µL opti- MEM. 1120 µL of the transfection mixture (solution A + solution B) prepared according to the manufacturer's instructions and described above was added to the cells. The final siRNA concentration was 10 nM. Cells were incubated at 37°C and 5% CO_2 for 48 or 72 h. After 48 or 72 h, cells were treated with 10 mM formaldehyde in ice-cold DMEM for 20 min to induce general DPCs and collected by trypsinization for DPC isolation by RADAR assay. In addition, a small aliquot of the cells was collected to check the efficiency of silencing by qPCR. Total RNA isolation from RPE1 or HEK293T cells, in which target gene was silenced by siRNA transfection, was performed using the Total RNA Miniprep Kit. RNA was quantified using the Bio-Spec Nano Spectrophotometer (Shimadzu Corporation, Kyoto, Japan) and RNA integrity was determined by agarose gel electrophoresis. The purified total RNA was reverse transcribed (1 µg of total RNA) using the ProtoScript II First Strand cDNA Synthesis Kit. For qPCR, specific primers were designed in Primer blast software (National Center for Biotechnology Information, Rockville Pike, USA), manually adjusted as needed, and purchased from Macrogen Service (Amsterdam, The Netherlands). Quantification of human gene expression was performed using the qPCR method of relative quantification. Human genes were normalized to the housekeeping gene *ATP50*.

2.2.8. Morpholino-mediated gene silencing in zebrafish embyros

All morpholinos were ordered from Genetools and used as described in Nasevicius and Ekker (2000). Both morpholinos (MO) targeting zebrafish *sprtn* and *xpa* were designed to alter splicing. 1nL of each MO solution (300 µM) diluted in 0.3M KCl and 0.015% phenol red was injected into 1-4 cell stage WT embyos and the resulting morphant phenotypes were then observed daily until 5 dpf. Pools of 5-15 embryos were collected at different stages for verification of splice-blocking efficiency via PCR on cDNA, or for RADAR assay. Embryos were

collected in eppendorf tubes, E3 medium was removed and samples for cDNA were dry-frozen. For DPC isolation by RADAR assay, embryos were deyolked and then dry-frozen. For the verification of splice-blocking efficiency of morpholino, RNA was extracted from pools of 5 embryos using the Total RNA Miniprep kit and cDNA was synthesized *in vitro* using the Protoscript II kit. PCR reactions were performed on 5 ng cDNA samples using a forward primer that binds to the exon upstream of the targeted exon, and a reverse primer binding the exon downstream of the targeted exon. Amplified DNA fragments were separated using 1% agarose gel electrophoresis. A change in the size of the amplicon derived from morpholino-injected embryos indicated successful splice-blocking activity of the morpholino.

2.2.9. Isolation of DPCs using KCl/SDS precipitation method

Detection of DPCs was performed using the KCl/SDS precipitation assay according to the patent of Zhitkovich and Costa (1992), adapted from the protocol described by Mórocz et al. (2017) (Mórocz et al., 2017). The method uses harsh treatments (2% SDS, heat at 65°C) to dissociate non-covalent and covalent DNA-protein bonds and selectively precipitate stable DNA-protein complexes by adding potassium chloride (KCl). KCl and SDS form an insoluble precipitate obtained by slow centrifugation. Since SDS binds proteins but not DNA, addition of KCl leads to precipitation of DNA containing a crosslinked proteins. Therefore, it is easy to separate the protein-free DNA that remains in the supernatant from the protein-bound DNA that is precipitated with the protein-bound SDS when the cation is changed from Na to K. The amount of DNA in the pellet represents the DNA-protein crosslinks (Zhitkovich and Costa, 1992). This method was used for DPC isolation from ACRC deficient zebrafish embryos treated with DPC-inducing agent formaldehyde. Zebrafish embryos were lysed in 400 µL lysis buffer (2% SDS; 20 mM Tris, pH 7.5) and then incubated for 15 min at room temperature. Samples were then placed in liquid nitrogen and incubated overnight. The next day, samples were thawed in a thermoblock (55°C/15 min) and then sonicated. Proteins were then precipitated by adding 400 µL KCl buffer (200 mM KCl; 20 mM Tris, pH 7.5) followed by incubation on ice for 5 min. The precipitated proteins were pelleted by centrifugation at 15,000 g for 5 min (4°C), and the supernatant was kept for quantification of soluble DNA. The pellet was washed three times by adding 400 µL KCl buffer, followed by incubation at 55°C for 5 min, 5 min on ice, and centrifugation at 15,000 g at 4°C for 5 min. After washing, each pellet was resuspended in 400 µL KCl buffer containing proteinase K (0.2 mg/mL) and incubated at 55°C for 1h. After incubation, BSA (1.25 mg/mL) was added and samples were incubated on ice for 5 min. After centrifugation at 15,000 g at 4°C for 5 min, the supernatant contained the crosslinked DNA. From both the soluble and crosslinked DNA, 50 µL of the sample was taken for treatment with RNase A (0.2 mg/mL) for 30 min at 37°C. Soluble DNA and crosslinked DNA were quantified using the Quant-iT™ 1X dsDNA HS assay and the amount of DPCs was calculated as the ratio of crosslinked DNA to total DNA (soluble + crosslinked).

2.2.10. Isolation and detection of DPCs by RADAR (Rapid Approach to DNA Adduct Recovery) assay

Total DPCs were isolated and detected in human cells and zebrafish embryos using a modified rapid approach to DNA adduct recovery (RADAR) assay (Kiianitsa and Maizels, 2013). Cells or zebrafish embryos were lysed in 1 or 1.5 mL of DPCs lysis buffer (6M guanidinium isothiocyanate, 10 mM Tris-HCl (pH 6.8), 20 mM EDTA, 4% Triton X-100, 1% N-lauroylsarcosine sodium, 5% β-mercaptoethanol) equilibrated at 55°C. DNA was precipitated by adding an equal volume of 100% ethanol followed by centrifugation at 10 000 x g for 5-10 min at 4°C. The DNA pellet was washed three times in wash buffer (20 mM Tris HCl pH 7.5, 50 mM NaCl, 1 mM EDTA, 50% ethanol) and dried for 30 min at 45°C in a thermoblock. The DNA was dissolved in 1 mL of 8 mM NaOH. A small portion of the recovered DNA (20 µL) was digested with 20 µg/mL proteinase K in proteinase K buffer (50 mM Tris pH 8.0, 100 mM NaCl, 0.5% SDS) in a total volume of 150 µL for 2 h at 55°C. DNA concentration was determined using a Quant-iT™ 1x dsDNA HS assay according to the manufacturer's instructions. Normalized amounts of dsDNA containing the DPCs were digested with 0.2 µL benzonase (25 U/ µL) in 5x benzonase buffer (250 mM Tris pH 7.9, 10 mM MgCl2) for 1 h at 37°C. Samples were frozen in liquid nitrogen and lyophilized overnight at 4°C. The next day, samples were dissolved in 50 µL SDS loading buffer (4M urea, 62.5 mM Tris-HCl pH =6.8, 1 mM EDTA, 2% SDS) and equilibrated at 65°C. DPCs were resolved with SDS-PAGE 5-18% gradient gel and visualized with ProteoSilver™ Silver Stain Kit as recommended by the manufacturer.

2.2.11. Western blot analysis

Zebrafish embryos or tail fins from adults were lysed in RIPA buffer (NaCl 150 mM, EDTA 1 mM, Tris 25 mM, SDS 0.5%) with the protease and phosphatase inhibitors (Sigma-Aldrich, Taufkirchen, Germany) for 30 min on ice. After the lysis, samples were briefly sonicated and centrifuged at 1,000 x g for 10 min at 4°C and the supernatants were used. Protein concentration was measured using the Bradford assay (Bradford, 1976). For this purpose, the Bradford reagent was prepared by dissolving 100 mg Coomassie Brilliant Blue G250 (Sigma-Aldrich, Taufkirchen, Germany) in 50 mL ethanol (96%; Kemika, Zagreb). Then, 100 mL of phosphorous acid was added to the ethanol solution. Finally, 850 mL mQ water was added to the mixture and filtered through a membrane filter (pore diameter 0.2 µm; TPP Techno Plastic Products AG, Switzerland). Western blot analysis was performed using the Mini- PROTEAN 3 Cell electrophoresis chamber (Bio-Rad Laboratories, CA, USA) for polyacrylamide gel electrophoresis together with the Multiphor ll Electrophoresis System (Pharmacia LKB Biotechnology, Uppsala, Sweden) for wet transfer onto a polyvinylidene difluoride membrane (Millipore, MA, USA). Five micrograms of protein per lane were separated by electrophoresis in a sodium dodecyl sulphate-polyacrylamide gel with a gradient of 5-18%. Protein size was determined using Precision plus protein ladder. Proteins were then transferred to polyvinylidene difluoride membrane (Millipore, MA, US) by wet blotting. Blocking was performed in 5% non-fat dry milk diluted in TBST for 2 h. Membranes were then washed in TBST buffer (20 mM Tris, 150 mM NaCl, 0.1% Tween) for 5 min and incubated overnight with

the primary antibody at 4°C in 2.5% BSA/TBST. Rabbit anti-mouse and goat anti-rabbit IgG peroxidase (1:10 000) were used as secondary antibodies (Sigma-Aldrich, Taufkirchen, Germany). Proteins were visualised by using Clarity Western ECL Substrate and chemiluminescence.

2.2.12. Slot blot analysis

After DPC isolation by RADAR assay from human cells, specific DPCs were detected by slot-blot analysis. It is a simplification of the Western blot method, with the difference that the detected proteins are not first separated by gel electrophoresis. Slot blot analysis was performed using the Bio-dot microfiltration device (Bio-Rad Laboratories, CA, USA). A nylon membrane was used for dsDNA blotting because it is positively charged and therefore enhances the binding of negatively charged DNA molecules, and a PVDF membrane was used for proteins. 200 µL of each sample was loaded and samples were vacuumed using a compressor (700 mbar). For dsDNA detection, DNA was crosslinked to the membrane by irradiation with UV light from the transiluminator for 5 min. Membranes were blocked for 2 h in 5% non-fat dry milk diluted in TBST, washed in TBST buffer (20 mM Tris, 150 mM NaCl, 0.1% Tween) for 5 min, and incubated overnight in the primary antibody at 4°C in 2.5% BSA/TBST. Rabbit anti-mouse and goat anti-rabbit IgG peroxidase (1:10 000) were used as secondary antibodies (Sigma-Aldrich, Taufkirchen, Germany). Proteins were visualised by using Clarity Western ECL Substrate and chemiluminescence.

2.2.13. Rescue experiments

To perform rescue experiments, the WT coding sequence of zebrafish *acrc* gene was amplified by PCR (F 5'-AGAGGATCTGCTCGAGATGGATCCTGGTACTTTATCACT-3', R 5'-TCACTATAGTTCTAGATCAACTTTGACTGAGACGAGTCT-3') and cloned into the multi-cloning site of the pCS2+HisMyc vector (between XhoI and XbaI) using the In-fusion cloning kit and verified by sequencing. The vector contained a promoter (SP6), a multicloning site, a polyA tail and a unique restriction site after the polyA tag (KpnI). RNA was synthesized using purified, KpnI-linearized plasmid by performing an *in vitro* transcription with SP6 using the Hiscribe SP6 RNA kit and a cap analog from the HiScribe T7 ARCA mRNA kit for improved mRNA stability. The resulting RNA was purified using the RNA cleanup kit and injected into 1-2 cell stage *acrc* mutant embryos or into control WT embryos (1nL of the mRNA solution containing 250 ng/µL RNA in 0.3M KCl and 0.015% phenol red). To determine the functionally relevant amino acids and domains of Acrc, deletion constructs were cloned by inverse PCR based on the original pCS2+HisMyc-DrAcrc plasmid. RNAs were then synthesized *in vitro* and injected as described above.

3. Results

3.1. Phylogenetic analysis of XPA proteins

Human and zebrafish share one-to-one orthology for XPA protein (Figure 12). XPA is conserved throughout the vertebrate and invertebrate lineage, with distant orthologs found in yeast and green algae (Figure 12). This level of conservation indicates the importance of XPA function in the cell. Considering that XPA is specifically involved in the NER pathway (Lehmann et al., 2017) highlights the relevance of the NER pathway in the repair of DNA damage.

Figure 12. Phylogenetic analysis of XPA proteins. XPA orthologs in vertebrates are shown in blue (fish are indicated in dark blue and other vertebrates in light blue), invertebrates are shown in colorless, yeast and fungi in orange, and green algae in green. The full length protein sequences were aligned using the MAFFT (Multiple Alignment using Fast Fourier Transform) alignment algorithm. The phylogenetic tree was constructed using the Maximum Likelihood method.

3.2. Synteny analysis of human and zebrafish *xpa* and *acrc* genes

Conserved synteny analysis was performed in order to determine chromosomal locations of zebrafish *xpa* and *acrc* genes and to determine level of gene order preservation. Zebrafish *xpa* is syntenic to the human *XPA* gene. In humans, the *XPA* gene is located on chromosome 9, whereas in zebrafish, the gene is located on chromosome 1 (Figure 13). XPA gene surrounding of genes which show synteny includes *B4GALT1, SPINK4, GNE, TDRD7* and *TMOD1* genes which are localized upstream of the *XPA* ortholog in humans and downstream of the *xpa* gene in zebrafish (Figure 13). *nansa* gene is localized downstream of the *XPA* ortholog in both, humans and zebrafish, whereas *slc24a2* is localized upstream of both orthologs (Figure 13). Other genes in the vicinity of human *XPA* show no syntenic relationship to the respective zebrafish orthologs. Some genes were translocated to other chromosomes and separated or were completely lost from the genome during the vertebrate evolution, resulting in loss of synteny (Figure 13).

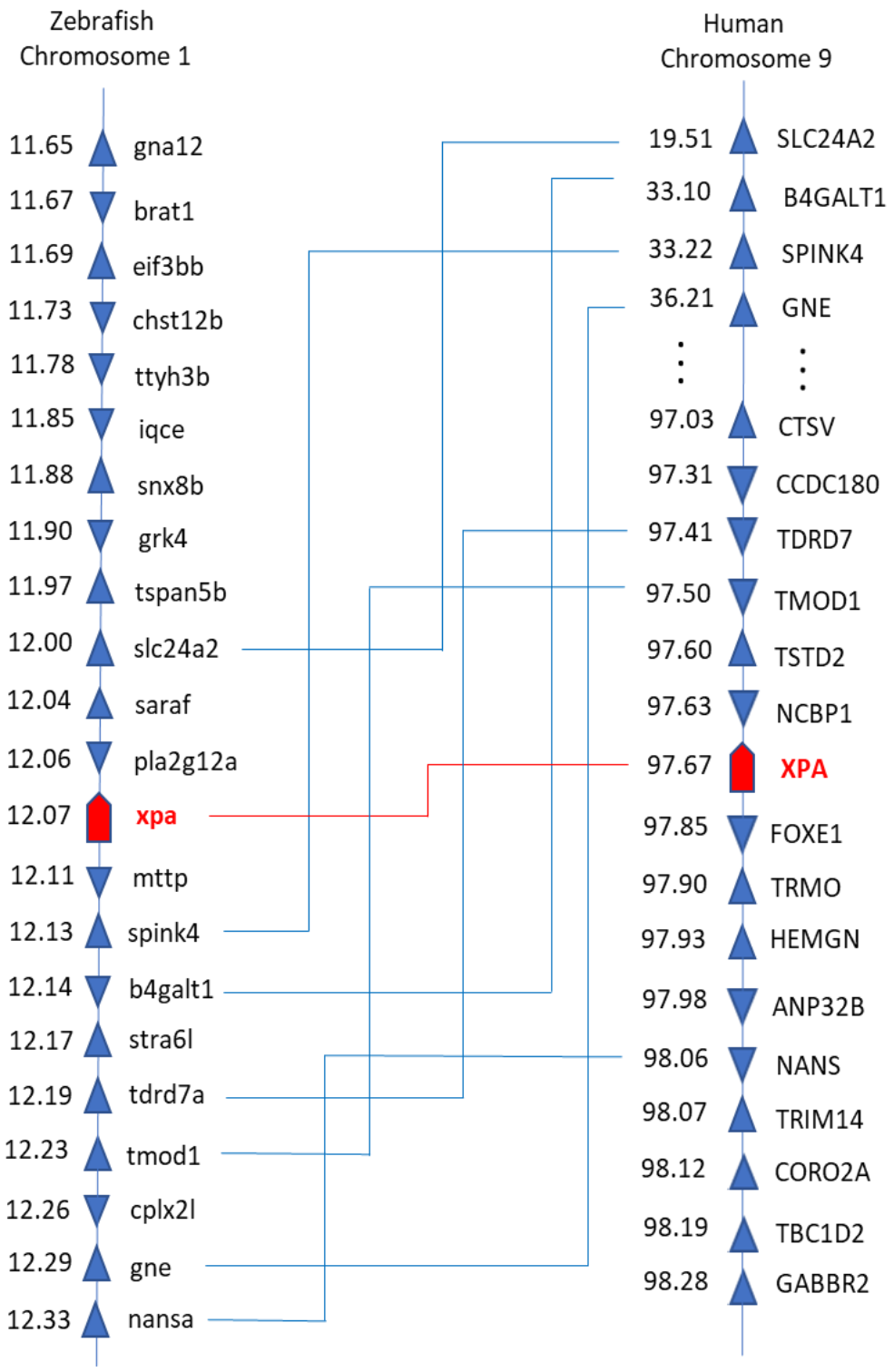


Figure 13. Conserved synteny analysis of zebrafish and human *XPA* genes. Numbers next to the gene names represent megabase pair (Mbp) of gene location on the respective chromosome. Chromosome segments are represented in blue (continuous segments). Orthologs are connected with lines.

Zebrafish *acrc* is syntenic to human *ACRC* gene. Human *ACRC* gene is located on X chromosome, while in zebrafish, the gene is positioned on the chromosome 14 (Figure 14). Syntenic genes in the gene surroundings of human and zebrafish *ACRC* orthologs include *THOC2, PLS3, RAB9B, ZNF185, NSDHL* and *CETN2* genes which are localized upstream of the *ACRC*, while *OGT* gene is localized downstream in both species. Human *NLGN3a* is syntenic to zebrafish *nlgn3a* and is localized upstream of the zebrafish *acrc* and downstream of human *ACRC* (Figure 14). Other genes in the surrounding of zebrafish *acrc* show no syntenic relationship to the human orthologs (Figure 14) due to their translocation or complete loss from the genome.

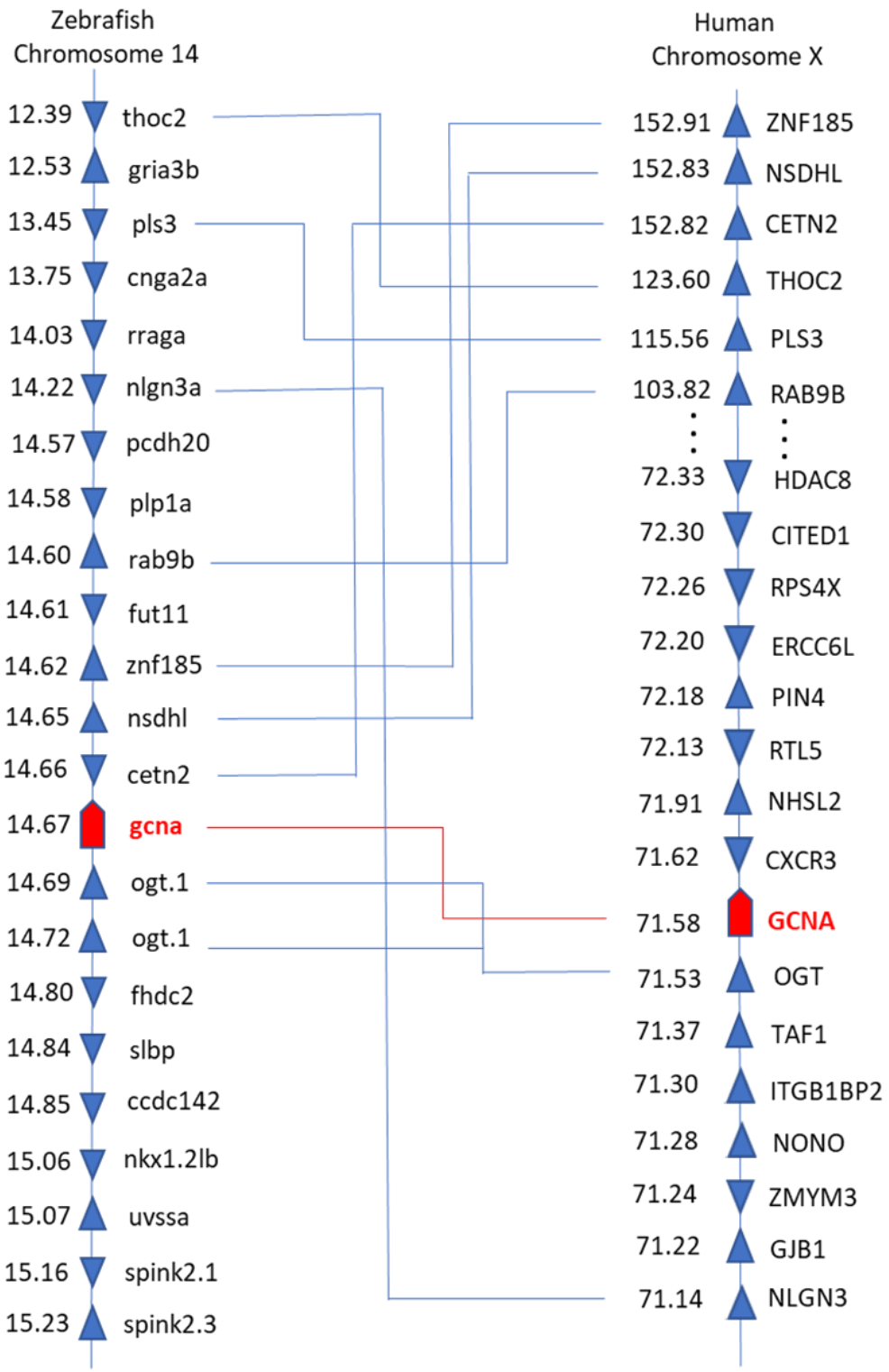
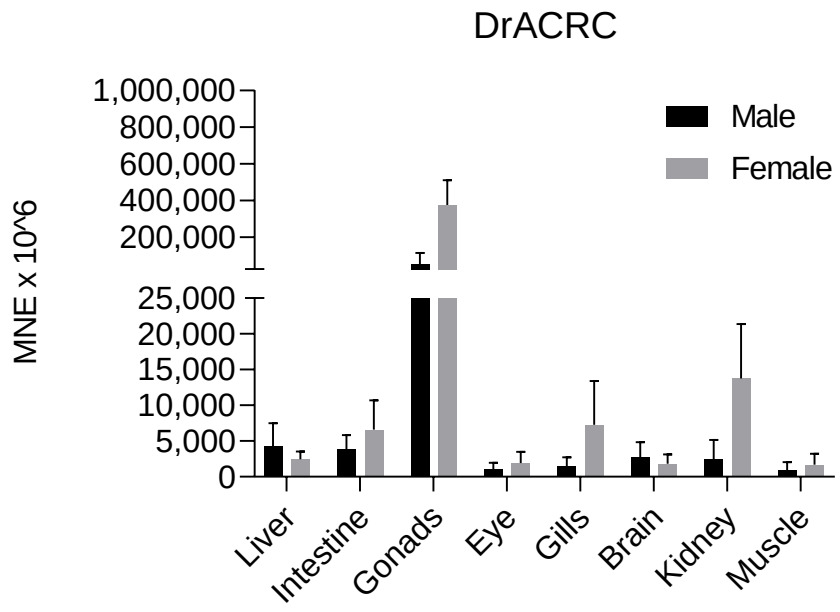


Figure 14. Conserved synteny analysis of zebrafish and human *GCNA/ACRC* genes. Numbers next to the gene names represent megabase pair (Mbp) of gene location on the chromosome. Chromosome segments are represented with blue (continuous segments). Orthologs are connected with lines.

3.3. Expression analysis of *acrc* in zebrafish embryos and adults

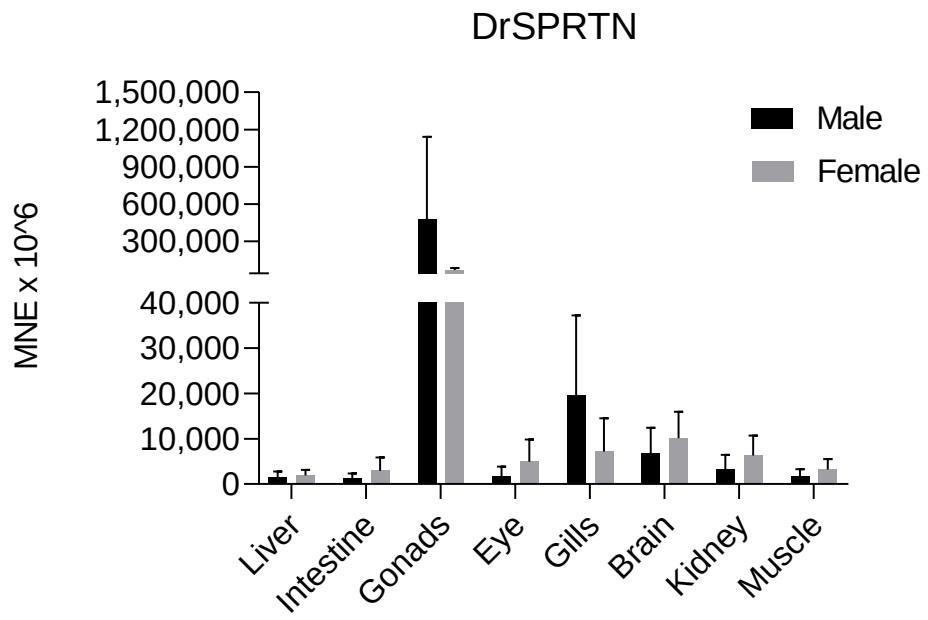
To perform functional characterization of Acrc in zebrafish and to compare it with Sprtn protease due to possible redundant role in DPC repair, the tissue specific expression profile of both genes was determined. The results of expression analysis of ACRC and SPRTN in adult zebrafish tissues of both sexes and during embryonic development are shown in Figure 15, while gene expression levels determined by arbitrary treshold values are listed in Table 3.1. ACRC shows very high expression in zebrafish gonads, while expression in other tissues is high or moderate. Of all male tissues analysed, ACRC is most highly expressed in testis (Figure 15 A, MNE*10^6= 52878±61449), where expression is very high. After testis, expression is highest in liver (Figure 15 A, MNE*10^6= 52878±3172), followed by intestine (Figure 15 A, MNE*10^6= 3842±1988) and brain (Figure 15 A, MNE*10^6= 2704±2128), where expression is high and reduced by 12-, 13- and 19-fold, respectively, compared with testis. In other male tissues, ACRC is moderately expressed (eye, gills, kidney and muscle). After brain, expression is highest in kidney (Figure 15 A, MNE*10^6= 2451±2721), followed by gills (Figure 15 A, MNE*10^6= 1525±1191), eye (Figure 15 A, MNE*10^6= 991±974) and muscle (Figure 15 A, MNE*10^6= 929±1134), where expression is reduced 21-, 34-, 53-, and 56-fold compared with testis. A similar expression pattern is observed in female tissues. ACRC is most highly expressed in ovaries (Figure 15 A, MNE*10^6= 376190±134313), followed by kidney (Figure 15 A, MNE*10^6= 13769±7585), where expression is also very high, although reduced 27-fold compared with ovaries. In the gills (Figure 15 A, MNE*10^6= 7274±6122) and intestine (Figure 15 A, MNE*10^6= 6581±4125), ACRC expression is high and reduced 51- and 57-fold compared with the ovaries. In other female tissues, ACRC is moderately expressed. After intestine, expression is highest in liver (Figure 15 A, MNE*10^6= 2439±1117), followed by eye (Figure 15 A, MNE*10^6= 1917±1571), brain (Figure 15 A, MNE*10^6= 1808±1342) and muscle (Figure 15 A, MNE*10^6= 1676±1536) in which 154-fold, 196-fold, 208-fold, and 224-fold reductions in expression were observed compared with ovaries.

(A)



To compare the expression levels of the two proteases, SPRTN expression was also analyzed in adult zebrafish tissues and during embryonic development. Like ACRC, SPRTN shows very high expression in zebrafish gonads, while expression in other tissues is high or moderate. In male tissues, SPRTN shows very high expression in testis (Figure 15 B, MNE*10^6= 474367±667802) and gills (Figure 15 B, MNE*10^6= 19461±17792). The expression of SPRTN in gills is 24-fold lower compared with that in testis. Expression of SPRTN is high in brain (Figure 15 B, MNE*10^6= 6748±5657) and kidney (Figure 15 B, MNE*10^6= 3135±3320), where 70- and 151-fold reduced expression was observed compared to testis. In other male tissues, SPRTN is moderately expressed (muscle, liver and intestine). After kidney, expression is highest in muscle (Figure 15 B, MNE*10^6= 1628±1647), followed by liver (Figure 15 B, MNE*10^6= 1575±1164) and intestine (Figure 15 B, MNE*10^6= 1238±1075), where expression is reduced 291-fold, 301-fold, and 383-fold compared with testis. In female tissues, SPRTN is most highly expressed in the ovaries (Figure 15 B, MNE*10^6= 66669±17790), where expression is very high. In all other female tissues, SPRTN is highly expressed, except in the liver, where expression is moderate. After ovaries, SPRTN is highly expressed in brain (Figure 15 B, MNE*10^6= 10130±5819), followed by gills (Figure 4 B, MNE*10^6= 7201±7305), kidney (Figure 4 B, MNE*10^6= 6310±4413), eye (Figure 15 B, MNE*10^6= 4969±4875), muscle (Figure 15 B, MNE*10^6= 3139±2310) and intestine (Figure 15 B, MNE*10^6= 2953±2924) in which 6-, 9-, 10-, 13-, 20-, and 22-fold reduced expression was observed compared with ovaries. In liver, expression is moderate (Figure 15 B, MNE*10^6= 1897±1203) and reduced 35-fold in compared with ovaries.

(B)



We also compared the expression of ACRC and SPRTN during zebrafish development. ACRC is expressed 2-5 times more than SPRTN between 6 and 48 hours post fertilization (hpf) (Figure 15 C). Zebrafish embryonic genes are transcriptionally silent until the 2.7 hpf (hours post fertilization) when embryos rely on maternal transcripts. After 4 hpf, zygotic transcription is established and reaches its maximum after 6 hpf. Up to 6 hpf, both zygotic and maternal transcripts can be detected (Laue et al, 2019). The maternal transcripts of both proteases are very highly expressed. ACRC shows the highest expression 1 hpf, which are likely maternal ACRC transcripts. ACRC expression is very high 1 hpf (Figure 15 C, MNE*10^6= 73123±77063) and 4 hpf (Figure 15 C, MNE*10^6= 29363±17938). At 6 hpf, expression is high and decreased 7-fold compared with 1 hpf (Figure 15 C, MNE*10^6= 9663±6682). At 12 hpf, expression increases again, probably because ACRC transcription from zygotic genome is established, and remains very high at 12 hpf (Figure 15 C, MNE*10^6= 13773±14308) and 24 hpf (Figure 15 C, MNE*10^6= 13419±7153). At 48hpf, ACRC expression is again reduced by 1.5 -fold compared with previous time point but remains high at both 48 hpf (Figure 15 C, MNE*10^6= 8771±6033) and 72 hpf (Figure 15 C, MNE*10^6= 4401±1366).

Similar to ACRC, SPRTN also shows the highest expression 1 hpf, which corresponds to maternal SPRTN transcripts. SPRTN expression is very high at 1 hpf (Figure 15 C, MNE*10^6= 16902±17336) followed by lower expression in other time points. However, expression is high at 4 hpf (Figure 15 C, MNE*10^6= 5992±5026), 6 hpf (Figure 15 C, MNE*10^6= 3121±1880), and 12 hpf (Figure 15 C, MNE*10^6= 3646±2716), although reduced by 2.8-fold, 5.4-fold and 4.6-fold compared with 1 hpf. Between 6 hpf and 12 hpf, SPRTN expression is increased 1.2-fold, which may be due to native expression of SPRTN from the zygotic genome. At 24 hpf, SPRTN

expression is again reduced by 1.5 -fold compared with the previous time point and is moderate (Figure 15 C, MNE*10^6= 2493±1808), but at 48 hpf (Figure 15 C, MNE*10^6= 2875±3222) and 72 hpf (Figure 15 C, MNE*10^6= 3809±2024) it increases to levels corresponding to high expression.

(C)

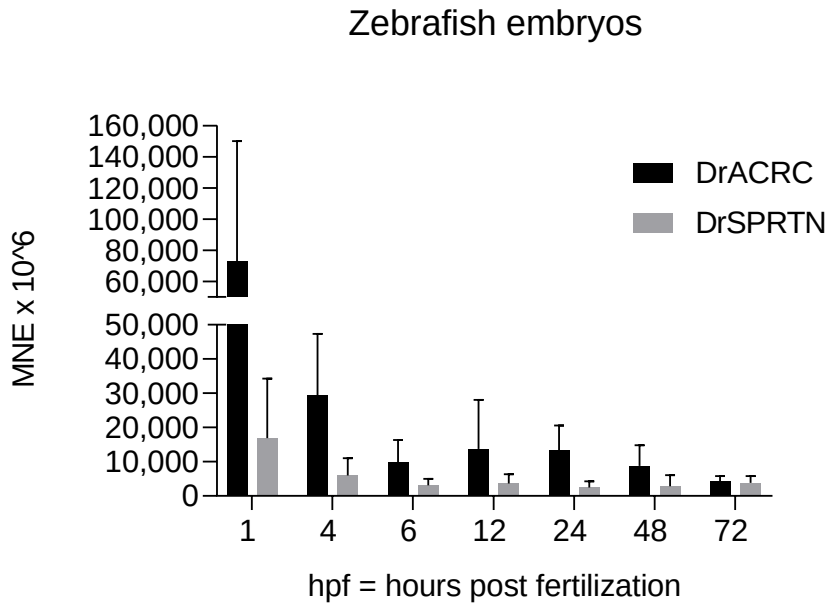


Figure 15. Expression patterns of (A) ACRC and (B) SPRTN in tissues of adult zebrafish quantified with qRT-PCR. C) Expression pattern of ACRC and SPRTN in zebrafish embryos quantified with qRT-PCR. Results of three independent experiments for females and males (three pools of organs collected from 5 individuals) and embryos (three pools of 30 embryos collected in indicated hour post fertilization) are given, while the kidney expression data results from two pools of 7 individuals. Data represent MNE (Mean Normalized Expression) ± SD normalized to the houskeeping gene *elongation factor 1α* (*ef1α*).

Table 3.1. Gene expression levels determined by arbitrary treshold values following previously published criteria (Lončar et al., 2016).

Expression	MNE value	Ct value
Very high	MNEx10^6 > 11700	< 22
High	MNEx10^6=2500-11700	22-25.9
Moderate	MNEx10^6=800-2500	26-30
Low	MNEx10^6 < 800	> 30

3.4. Creation of Xpa deficient zebrafish strain

To investigate the role of the NER pathway in DPC repair, a mutant zebrafish strain was generated using CRISPR/Cas9 genome editing (Figure 16). sgRNA targeting exon 6 and binding to position 5638-5656 on genomic DNA of zebrafish *xpa* gene on chromosome 1 (CDS position 610-628) was designed. sgRNA was injected in complex with the Cas9 protein into the yolk of zebrafish embryos at the one-cell stage to generate a mutant strain containing a premature stop codon within the Xpa protein sequence. A frameshift mutation was introduced into the DNA-binding domain of the Xpa protein, resulting in a premature stop codon and impairment of the NER pathway. The generated mutant strain of zebrafish was reared and crossed with wild-type individuals after reaching sexual maturity to identify the individuals that harbour germline mutations (founders). High-resolution melting (HRM) analysis was used to distinguish embryos carring a mutation from those without mutations. Embryos that were HRM positive (T_m different from WT embryos) were sequenced to determine the type of mutation. F0 adults carrying a mutation in the *xpa* gene were crossed with each other to obtain an F1 generation. F1 adults carrying the desired mutation in the *xpa* gene were identified by genotyping using tail fin tissue for the purpose of finding homozygous individuals (male and female) for the target mutation (Figure 16).

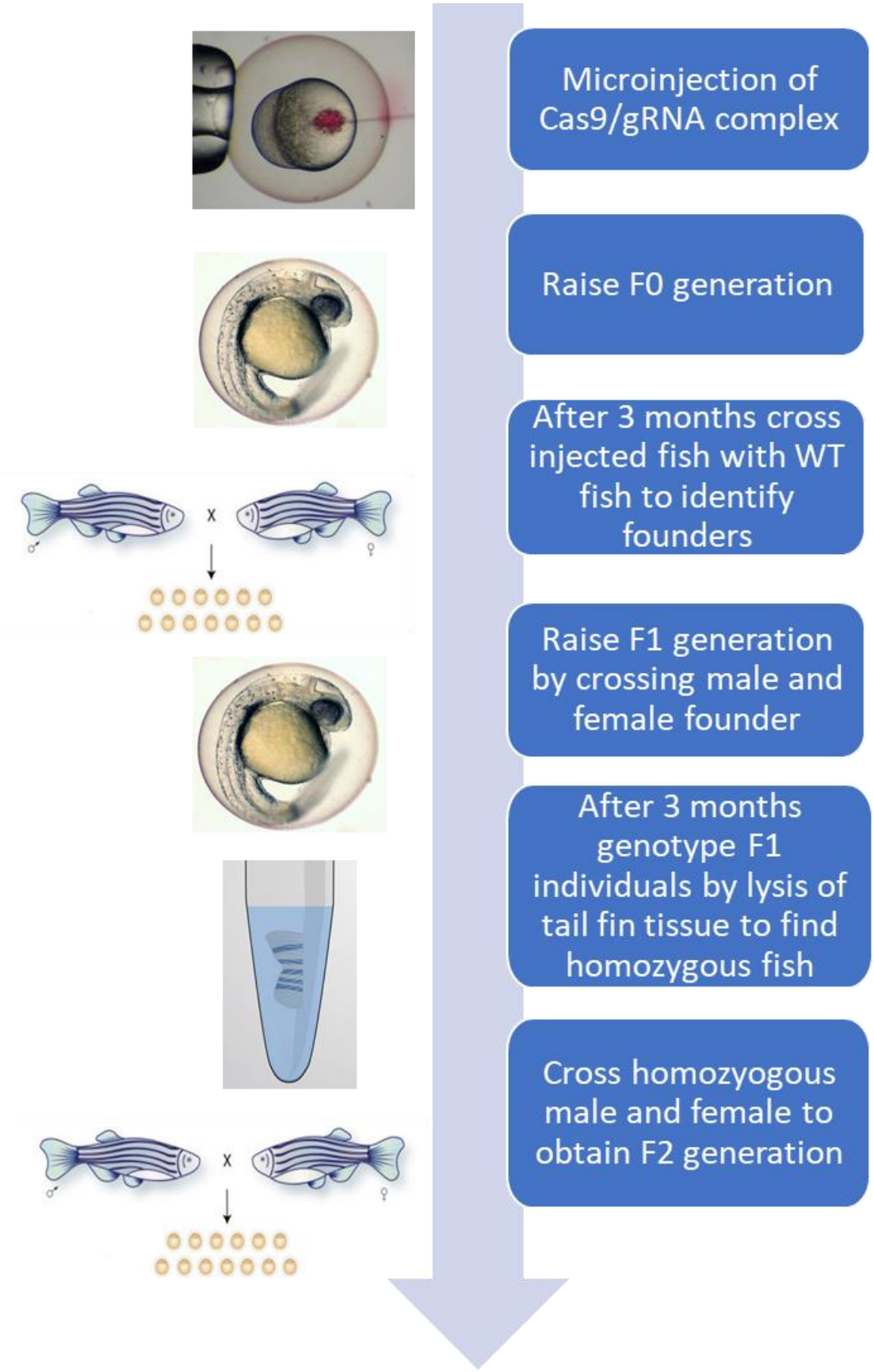
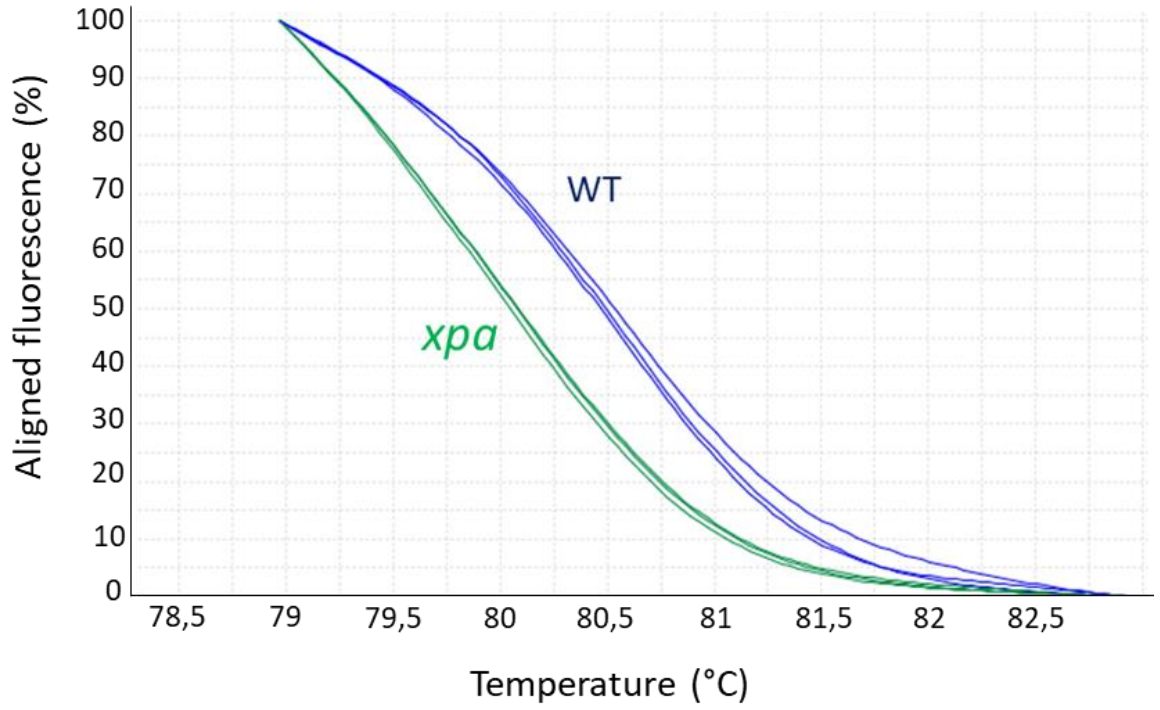


Figure 16. Schematic representation of Xpa and Acrc mutant zebrafish strain generation.

3.4.1. Genotyping of Xpa F0 generation

Embryos obtained by crossing F0 *xpa* mutant male fish and WT female fish were screened for the presence of mutations by HRM analysis to identify a male individual which transmits the target mutation to the next generation (founder). Out of 10 embryos randomly collected from crossing F0 *xpa* mutant male fish and WT female fish, melting curves analysed by HRM software showed that 7 embryos were potentially WT and 3 embryos potentially carried a mutation. PCR amplicons of target gene sequence analysed by HRM showed significantly different melting temperatures (T_m) from WT embryos (Figure 17 A and Table 3.2). The presence of target mutation was further confirmed by cloning and sequencing of two HRM positive embryos. In embryo 1 and 2 (E1, E2) a 19 nt deletion in exon 6 at position 5637-5655 of the *xpa* gene was detected (Figure 17 B). A 19 nt deletion in the F0 *xpa* mutant male fish results in a frameshift mutation and premature stop codon at position 222 of the Xpa protein (Figure 6 B). This founder was therefore used to obtain the F1 generation.

(A)



(B)

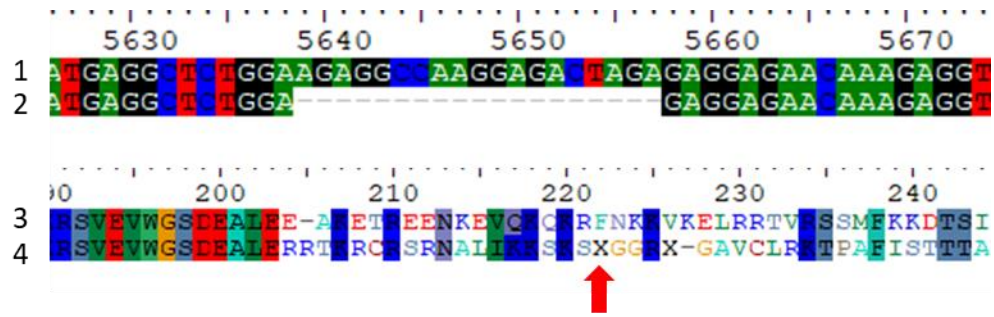


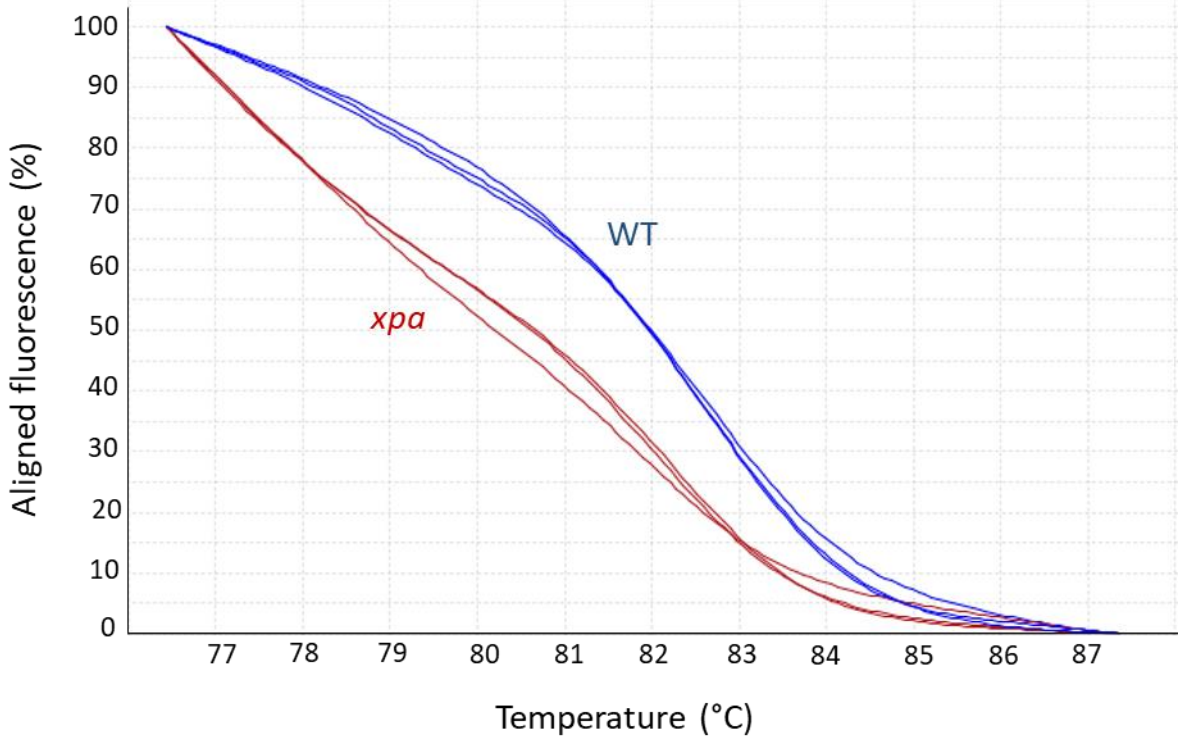
Figure 17. A) Genotyping of F0 *xpa* mutant male fish using HRM analysis. Aligned melt curves determined by HRM analysis are shown for WT embryos and progeny from *xpa* mutant fish which was crossed with WT fish. The samples that the software recognized as different genomic variants are marked in blue (WT) or green (*xpa* mutation). (B) Genotyping of F0 *xpa* mutant male fish by cloning and sequencing. Aligned nucleotide sequence (above image) and amino acid sequence (bottom image) for the sequenced DNA amplicon of the E1 embryo sample are shown. The first row of the above figure (1) shows the consensus nucleotide sequence of the *xpa* zebrafish gene, and the second row (2) shows the nucleotide sequence of the sequenced DNA amplicon. The first row of the figure below (3) shows the consensus amino acid sequence of the zebrafish Xpa protein, and the second row (4) shows the translated amino acid sequence from the sequenced DNA amplicon. The 19 nt deletion on the genomic DNA corresponds to a frameshift mutation and a premature stop codon at position 222 (red arrow). The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_956765.1).

Table 3.2. Melting temperatures (T_m) determined by HRM analysis in high-resolution melting software for WT and *xpa* mutant embryos. T_m - melting temperature, WT - wild type embryos, *xpa* E1, E2 and E3 - embryos obtained by crossing the *xpa* mutant F0 male with a wild-type female.

Sample	**T_m (°C)**
WT1	80,8
WT2	80,8
WT3	80,8
xpa E1	80,2
xpa E2	80,2
xpa E3	80,2

Embryos obtained by crossing F0 *xpa* mutant female fish and WT male fish were screened for the presence of mutations by HRM analysis to identify female founder. *Xpa* mutant F0 female had three HRM positive embryos that differed in T_m from WT embryos (Figure 18 A and Table 3.3). Out of 10 embryos collected from crossing *xpa* F0 female and WT male, melting curves analysed by HRM software showed that 7 embryos were potentially WT and 3 embryos potentially carried a mutation. The presence of mutations was confirmed by cloning and sequencing. We have sequenced embryo 1, 2 and 3 (E1, E2, E3). A 33 nt deletion in exon 6 at position 5629-5661 in the genomic DNA was detected in embryos 1 and 3, which results in 11 amino acids deletion (Figure 18 B). In embryo 2, a 30 nt deletion at position 5630 – 5659 in the genomic DNA was detected, which results in 10 amino acids deletion (Figure 18 B). Neither of this mutations results in a premature stop codon and frameshift mutation in the XPA protein sequence (Figure 18 B). However, this individual was crossed with F0 *xpa* mutant male fish to obtain the F1 generation. Breeding of F0 *xpa* mutant male fish with other F0 mutant female fish, carrying a frameshift mutation was unsuccessful for an unknown reason (there was no production of offspring).

(A)



(B)

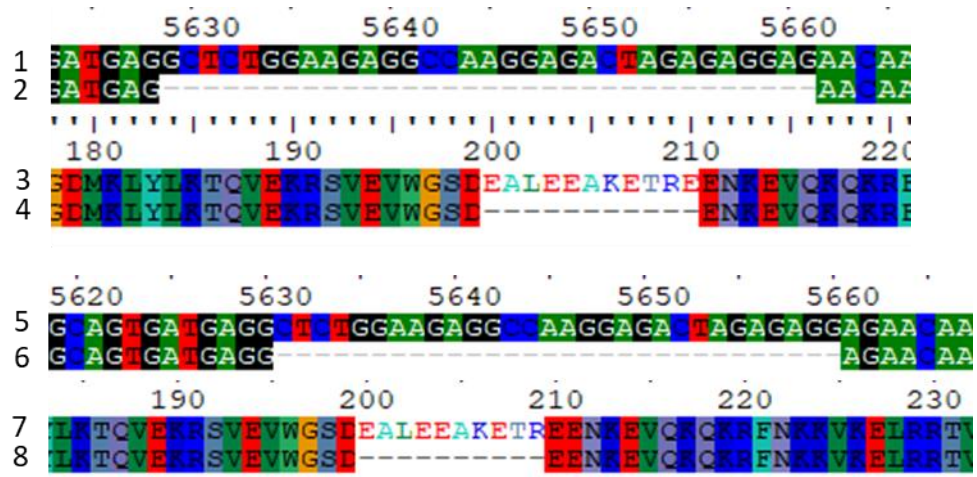


Figure 18. A) Genotyping of F0 *xpa* mutant female fish using HRM analysis. Aligned melt curves are shown determined by HRM analysis in high-resolution melting software for WT embryos and progeny from *xpa* mutant female fish crossed with WT male. The samples that the software recognized as different genomic variants are marked in blue (WT) or red (*xpa* mutation). (B) Genotyping of F0 *xpa* mutant female fish by cloning and sequencing. Aligned nucleotide sequence (above images) and amino acid sequence (bottom images) for the sequenced DNA amplicon of the E1, E2 and E3 embryo sample are shown. The first row of the figures above (1 and 5) shows the consensus nucleotide sequence of the *xpa* zebrafish gene, and the second row (2 and 6) shows the nucleotide sequence of the sequenced DNA amplicon. The first row of the figures below (3 and 7) shows the consensus amino acid sequence of the zebrafish Xpa protein, and the second row (4 and 8) shows the translated amino acid sequence from the sequenced DNA amplicon. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_956765.1).

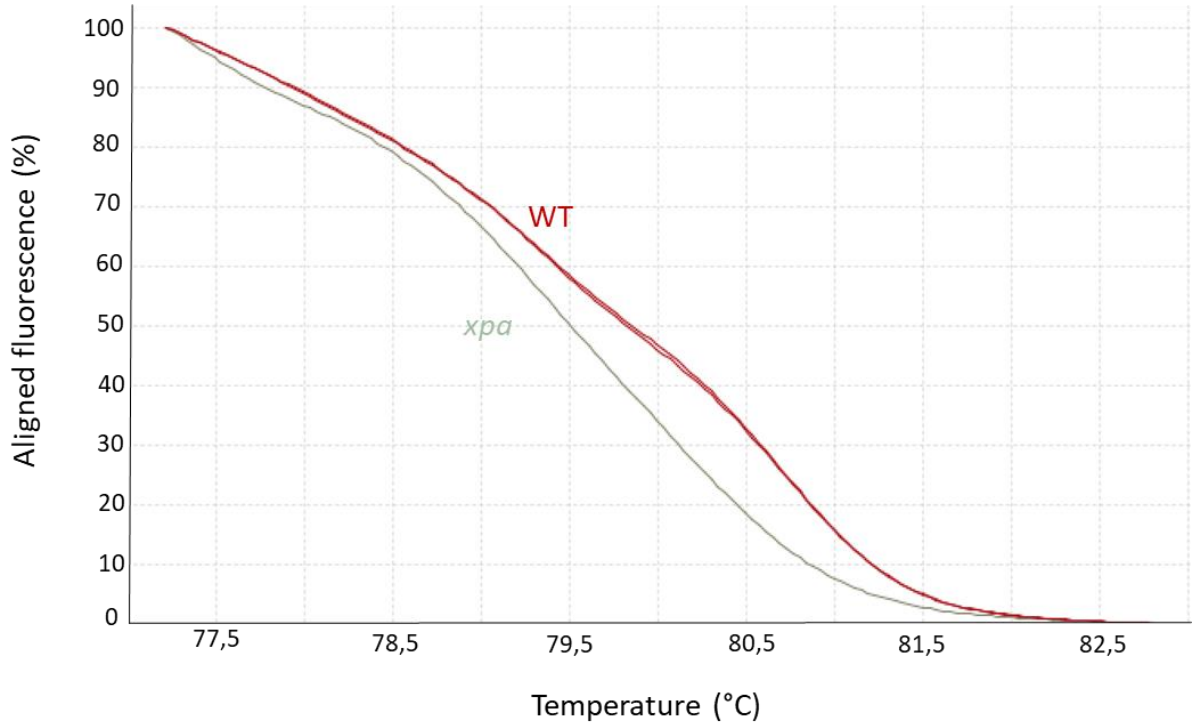
Table 3.3. Melting temperatures (°C) determined by HRM analysis for WT and *xpa* mutant embryos. T_m - melting temperature, WT - wild type embryo sample, *xpa* E1, E2 and E3 - embryos obtained by crossing the mutant *xpa* F0 female with a wild-type male.

Sample	**T_m (°C)**
WT1	80,8
WT2	80,8
WT3	80,8
xpa E1	80,6
xpa E2	78,1
xpa E3	80,5

3.4.2. Genotyping of Xpa F1 generation

The F1 generation was created by crossing F0 *xpa* mutant male fish and F0 *xpa* mutant female fish without frameshift mutation in XPA protein. After 3 months when F1 generation reached sexual maturity, individuals were screened for the presence of *xpa* gene mutations by HRM analysis. The F1 female fish were genotyped by analysis of DNA from fin tail tissue. Melting curves analysed by HRM software showed that out of 20 analysed fish, 15 were potentially WT and 5 potentially carried a mutation. Aligned melt curves are shown for the one F1 female that was HRM positive with a T_m different from that of a DNA from fin tail tissue of the WT fish (Figure 19 A and Table 3.4). The presence of a mutation in the *xpa* gene in this individual was confirmed by cloning and sequencing. Sequencing of the DNA from fin tail tissue of F1 *xpa* mutant female fish showed the presence of a 19 nt deletion at position 5637-5655 in the genomic DNA, resulting in a frameshift mutation and premature stop codon at position 222 at a protein level (Figure 19 B). Same mutation was observed in the previous generation and is the one we wanted to select because it disrupts Xpa protein.

(A)



(B)

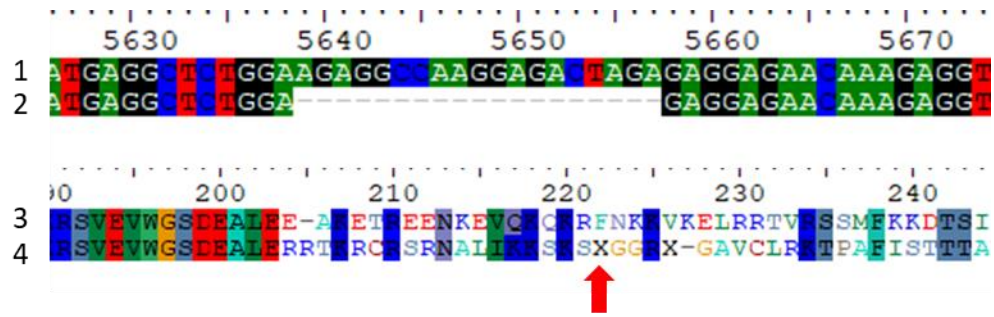


Figure 19. (A) Genotyping of F1 *xpa* mutant female fish by HRM analysis. Aligned melt curves determined by HRM analysis are shown for DNA obtained by lysis of tail fin tissue from WT and F1 *xpa* mutant female fish. The samples that the program recognized as different genomic variants are marked in red (WT) or grey (*xpa* mutation). (B) Genotyping of F1 *xpa* mutant female fish by cloning and sequencing. The 19 nt deletion on the genomic DNA corresponds to a frameshift mutation and a premature stop codon at position 222 (red arrow). Aligned nucleotide sequence (above image) and amino acid sequence (bottom image) for the sequenced DNA isolated from tail fin tissue of F1 *xpa* female fish are shown. The first row of the above figure (1) shows the consensus nucleotide sequence of the *xpa* zebrafish gene, and the second row (2) shows the nucleotide sequence of the sequenced DNA amplicon. The first row of the figure below (3) shows the consensus amino acid sequence of the zebrafish Xpa protein, and the second row (4) shows the translated amino acid sequence from the sequenced DNA amplicon. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_956765.1).

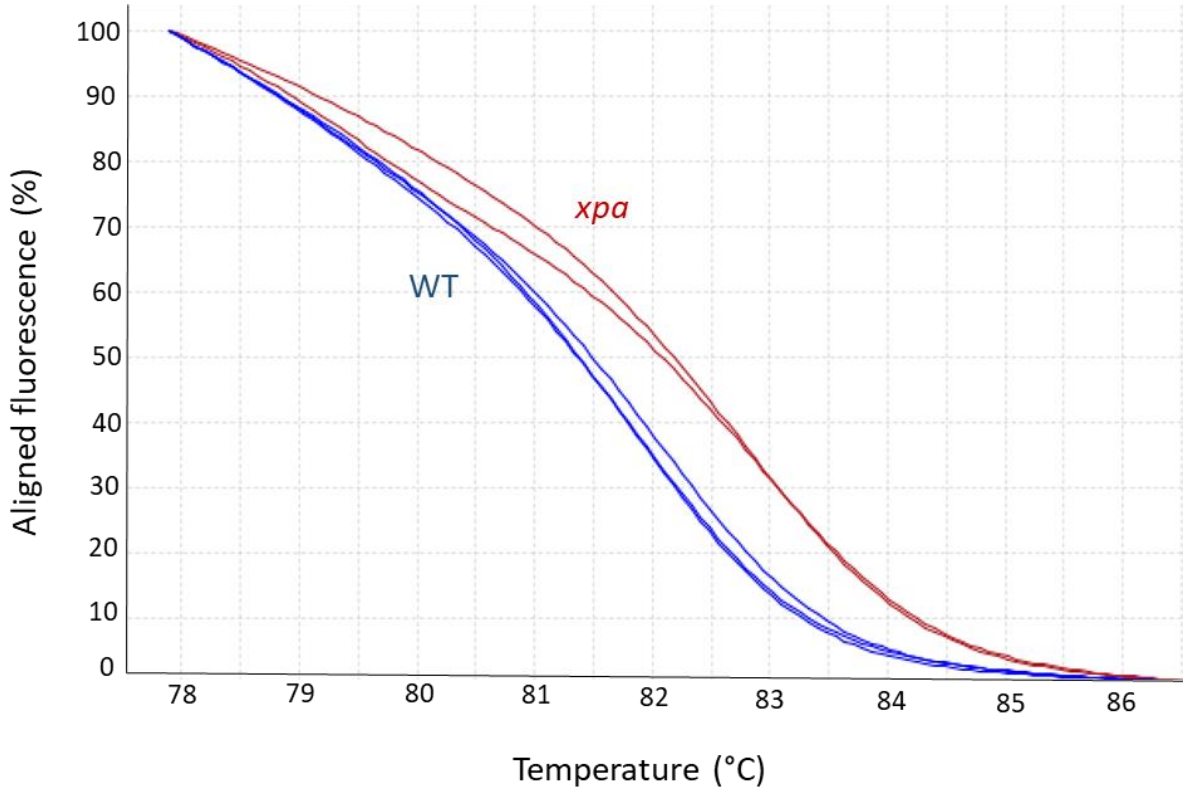
Table 3.4. Melting temperatures (°C) determined by HRM analysis in high-resolution melting software for DNA obtained by lysis of tail fin tissue from WT and F1 *xpa* mutant fish. T_m - melting temperature, WT FC1 and 2 – DNA isolated from tail fin tissue of WT fish, *xpa* FC1 - DNA isolated from tail fin tissue of F1 *xpa* mutant female fish.

Sample	**T_m (°C)**
WT FC1	81,2
WT FC2	81,1
xpa FC1	80,1

Embryos obtained by crossing F1 *xpa* mutant male fish with WT female fish were screened for the presence of mutations by HRM analysis to identify F1 male mutant fish. F1 *xpa* mutant male fish had two HRM positive embryos whose T_m is different from WT embryos (Figure 20 A and Table 3.5). Out of 10 embryos collected from crossing *xpa* F1 male with WT female,

melting curves analysed by HRM software showed that 8 embryos were WT and 2 embryos carried a mutation. The presence of mutations was confirmed by cloning and sequencing. We have sequenced embryo 1 (E1) in which a 19 nt deletion at position 5637-5655 in the genomic DNA was detected, resulting in a frameshift mutation and premature stop codon at position 222 at a protein level (Figure 20 B).

(A)



(B)

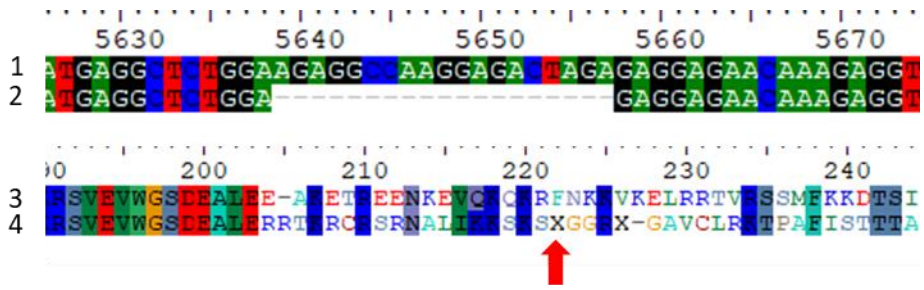


Figure 20. A) Genotyping of F1 *xpa* mutant male fish by HRM analysis. Aligned melt curves determined by HRM analysis are shown for WT embryos and progeny from F1 *xpa* mutant male fish crossed with WT female fish. The samples that the software recognized as different genomic variants are marked in blue (WT) or red (*xpa* mutation). (B) Genotyping of F1 *xpa* mutant male fish by cloning and sequencing. Aligned nucleotide sequence (above image) and

amino acid sequence (bottom image) for the sequenced DNA from the E1 embryo sample are shown. The first row of the above figure (1) shows the consensus nucleotide sequence of the *xpa* zebrafish gene, and the second row (2) shows the nucleotide sequence of the sequenced DNA sample. The first row of the figure below (3) shows the consensus amino acid sequence of the zebrafish Xpa protein, and the second row (4) shows the translated amino acid sequence from the sequenced DNA amplicon. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_956765.1).

Table 3.5. Melting temperatures (°C) determined by HRM analysis in high-resolution melting software for WT and *xpa* mutant embryos. T_m - melting temperature, WT - wild type embryo sample, *xpa* E1 and E2 - embryos obtained by crossing F1 *xpa* mutant male fish with wild-type female fish.

Sample	T_m (°C)
WT1	80,7
WT2	80,7
WT3	80,6
xpa E1	81,1
xpa E2	81,1

3.5. Morpholino-mediated XPA silencing in zebrafish embryos

To silence *xpa* gene expression in zebrafish embryos, a splice-blocking morpholino was designed to target the exon 4-intron 4 boundary of the zebrafish *xpa* gene (Figure 21). By binding to the *xpa* pre-mRNA, it prevents splicing, thereby most likely leading to a frameshift and Xpa loss-of-function (Cecile Otten, IRB, unpublished results).

Figure 21. A schematic of the *xpa* morpholino binding position in genomic DNA. The morpholino binds at position 2886-2910 in genomic DNA, which corresponds to the boundary between exon 4 and intron 4.

The efficiency of morpholino was tested by PCR reaction on cDNA derived from pools of embryos 2 days post fertilization (Cecile Otten, IRB, unpublished results). A WT amplicon (610 bp) was present in the non-injected WT sample, whereas it was completely absent in the *xpa* morphant sample, where only a truncated fragment (504 bp) resulting from the exon 4 skipping was present. This indicated that the *xpa* morpholino is 100% efficient (Figure 22). Importantly, *xpa* morphant embryos showed no obvious phenotypic changes and 98% of morphants were similar to WT embryos (Figure 23).

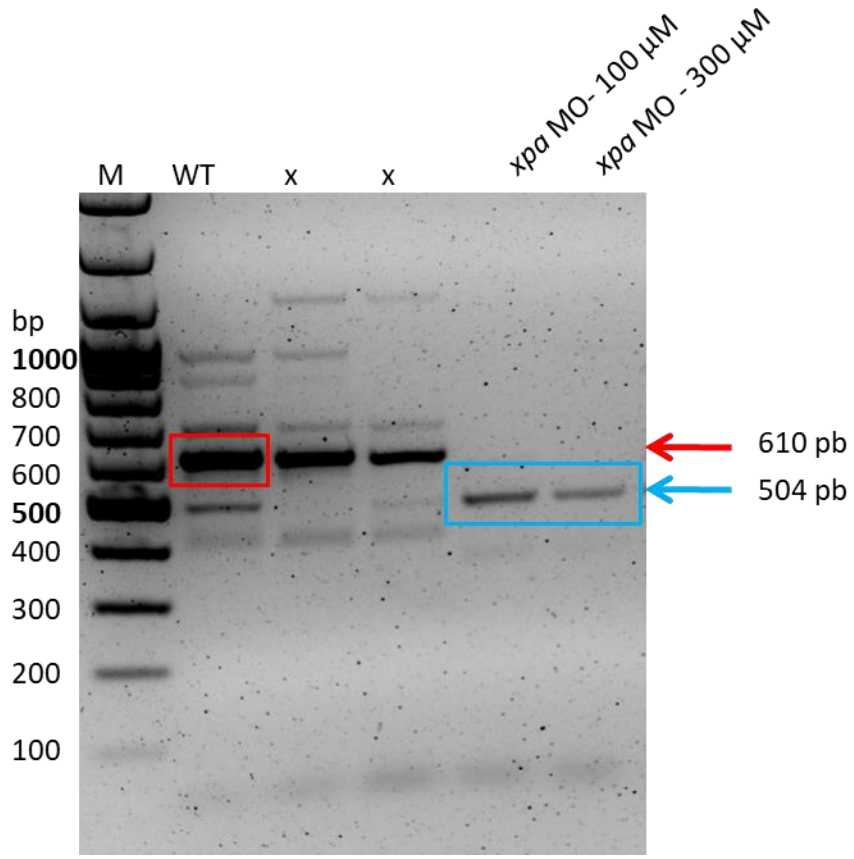


Figure 22. PCR reaction performed with cDNA from WT zebrafish embryos and embryos injected with 100 µM and 300 µM *xpa* morpholino. A specific amplicon at 610 pb is detected in WT sample, while it is completely absent in embryos injected with *xpa* morpholino. A shorter fragment (504 pb) corresponding to the truncated amplicon is detected in embryos injected with *xpa* morpholino. M- marker, WT – non-injected WT embryos, *xpa* MO - 100 and 300 µM – embryos injected with increasing concentrations of *xpa* morpholino.

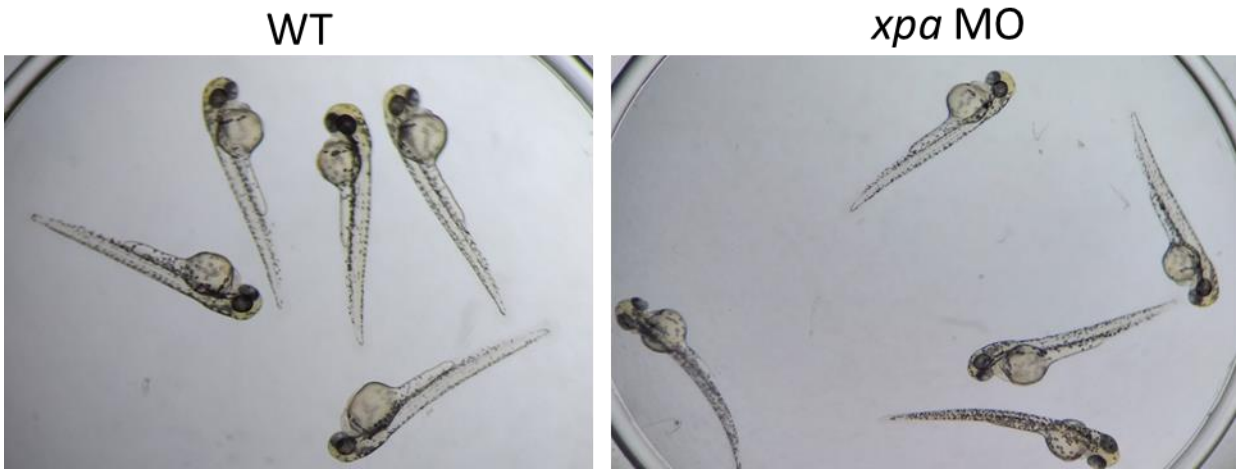


Figure 23. Representative pictures of zebrafish WT embryos and embryos injected with *xpa* morpholino (2 days post fertilization, n = 58).

3.6. Morpholino-mediated SPRTN silencing in zebrafish embryos

To silence *sprtn* in zebrafish embryos, the splice-blocking morpholino was engineered to inhibit mRNA splicing, resulting in frameshift mutation and degradation of Sprtn protein. Morpholino binds to position 2896-2920 in genomic DNA, which corresponds to the boundary between exon 2 and intron 2 (Figure 24) (Cecile Otten, IRB, unpublished results).

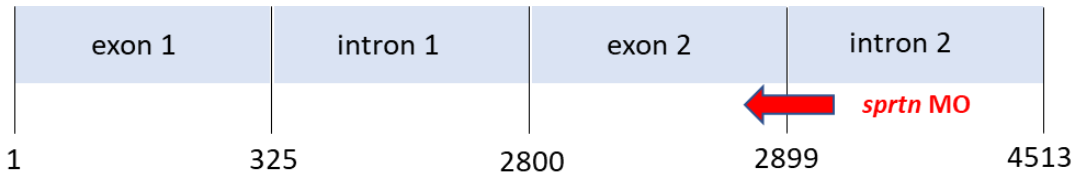


Figure 24. A schematic of the *sprtn* morpholino binding position in genomic DNA. The morpholino binds at position 2896-2920 in SPRTN genomic DNA, which corresponds to the boundary between exon 2 and intron 2.

The efficiency of morpholino was tested by PCR reaction (Cecile Otten, IRB, unpublished results). Wild-type zebrafish embryos were injected with *sprtn* morpholino (300 µM) and lysed. RNA was isolated from the injected embryos and transcribed into cDNA, which was used as a template for PCR reaction. The products of PCR reaction were run on 1% agarose gel electrophoresis. In non-injected WT embryos, a WT amplicon (439 pb) is present, whereas in embryos injected with *sprtn* morpholino WT amplicon is reduced and the truncated fragment (339 pb) resulting from exon skipping is dominant, indicating that *sprtn* morpholino

is 79% efficient (Figure 25). Embryos injected with *sprtn* morpholino had shorter body and mild tail curvature in comparison to WT embryos in 36% of embryos (n = 58) (Figure 26).

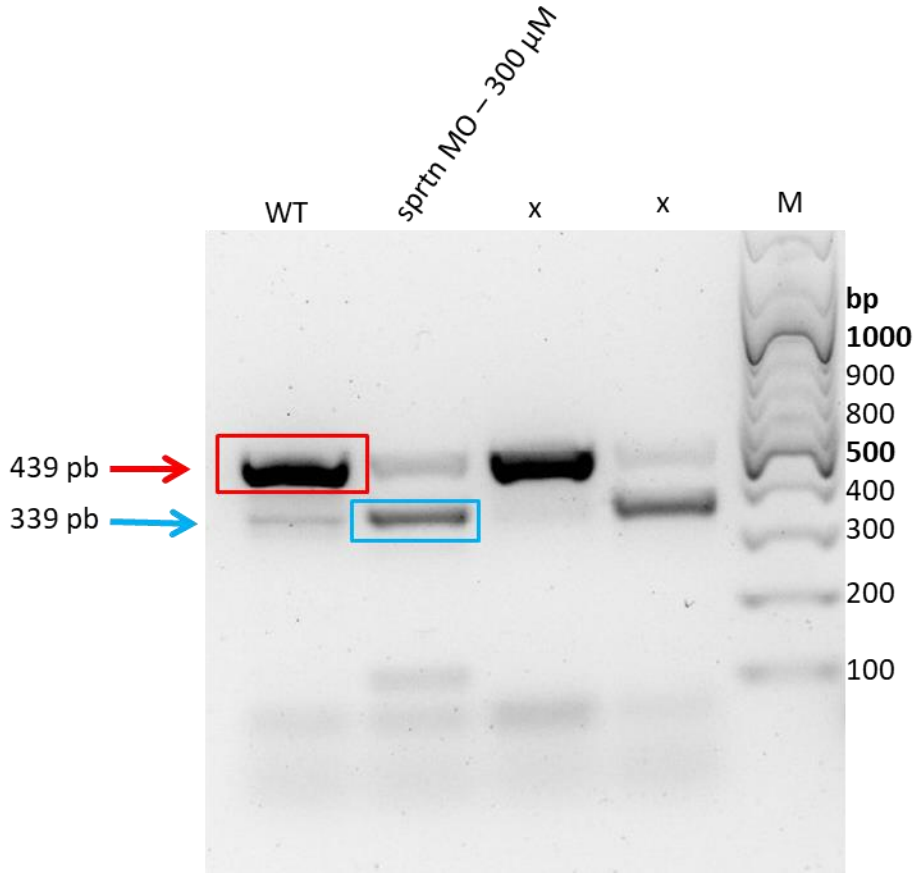


Figure 25. PCR reaction using cDNA from WT zebrafish embryos and embryos injected with 300 µM *sprtn* morpholino. In the WT sample, a specific amplicon is detected at 439 pb. In sample injected with *sprtn* morpholino, WT amplicon was reduced and a shorter fragment at 339 pb became dominant. M- marker, WT – non-injected WT embryos, *sprtn* MO - 300 µM – embryos injected with 300 µM *sprtn* morpholino.

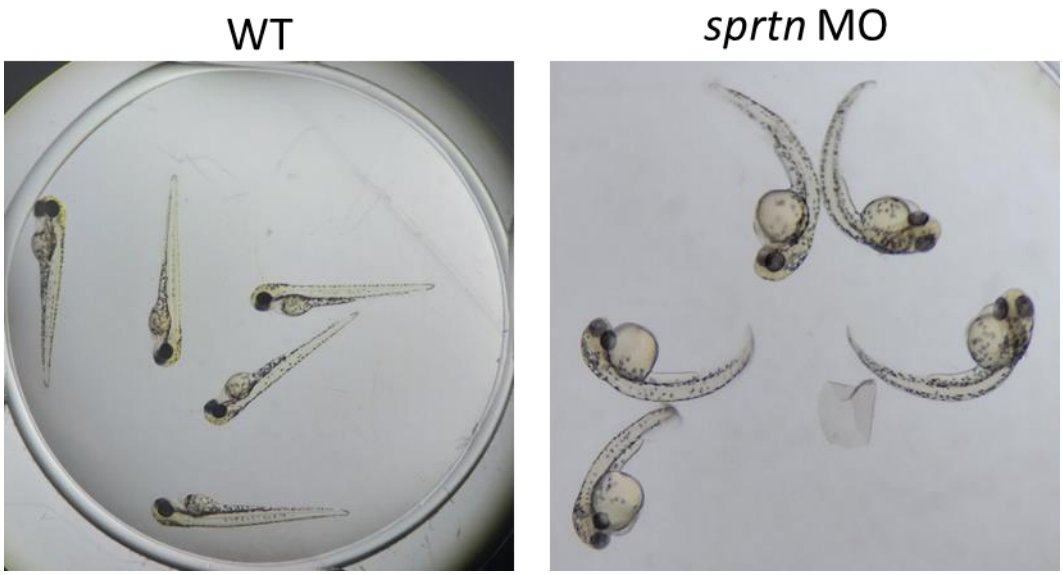
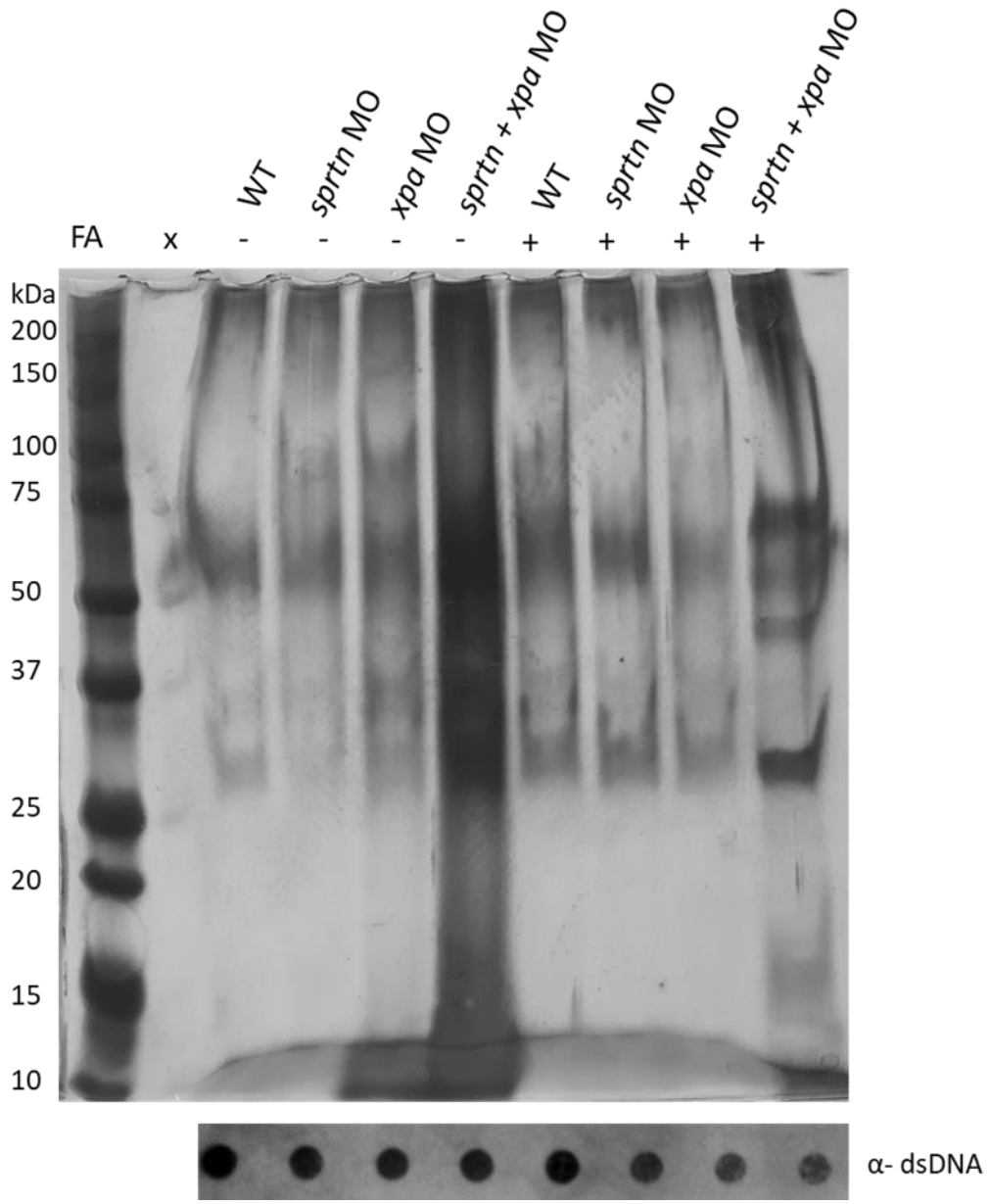


Figure 26. Representative pictures of zebrafish WT embryos and embryos injected with *sprtn* morpholino (2 days post fertilization, n = 58).

3.7. DPC levels in zebrafish embryos after morpholino-mediated *xpa* and *sprtn* silencing

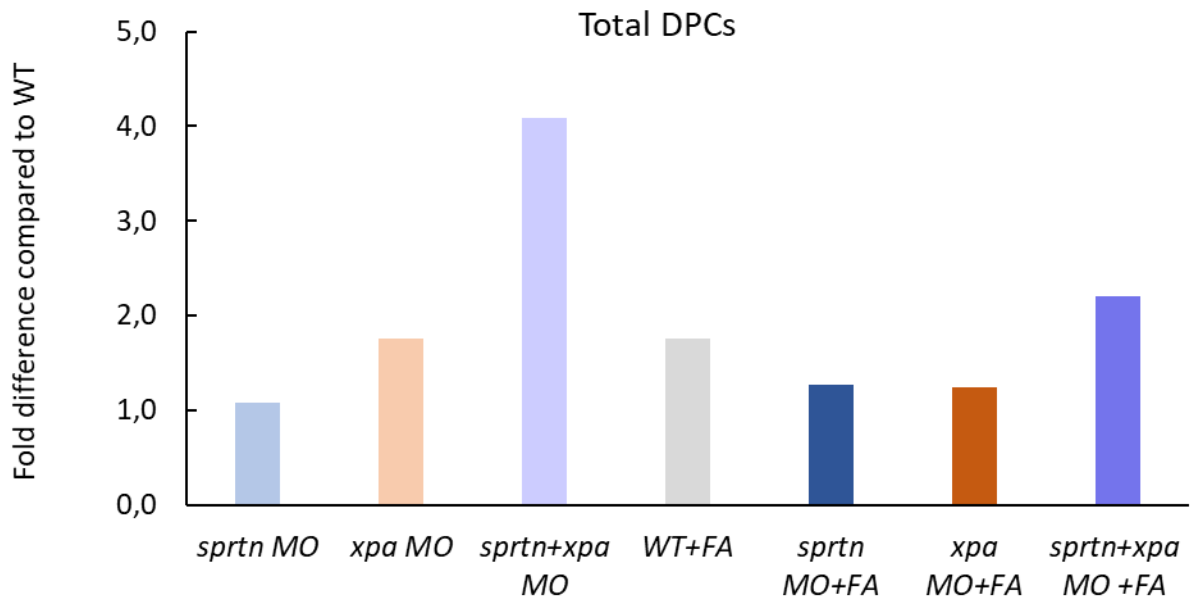
Zebrafish embryos were injected simultaneously with *sprtn* and *xpa* morpholino (300 µM) and collected 2 days post fertilization. DPCs were isolated using the RADAR assay, separated using SDS-PAGE and visualized with silver staining (Figure 27 A). DPC isolates were quantified for total DNA amount to ensure equal amount of DNA for DPC analysis. Double-stranded DNA (dsDNA) was detected with a specific antibody to ensure loading of equal amount of DNA for each sample and to compare levels of proteins crosslinked to DNA (Figure 27 A). It is known that selection of DPC repair pathway often depends on size of crosslinked proteins, therefore ImageJ software was used to quantify total, high (HMW), medium (MMW) and low molecular weight (LMW) DPCs. The threshold values for the analysis of DPC size are shown in Table 3.6.

(A)


FA
x
WT
sprtn MO
xpa MO
sprtn + xpa MO
WT
sprtn MO
xpa MO
sprtn + xpa MO
- - - - + + + +
kDa
200
150
100
75
50
37
25
20
15
10
α- dsDNA

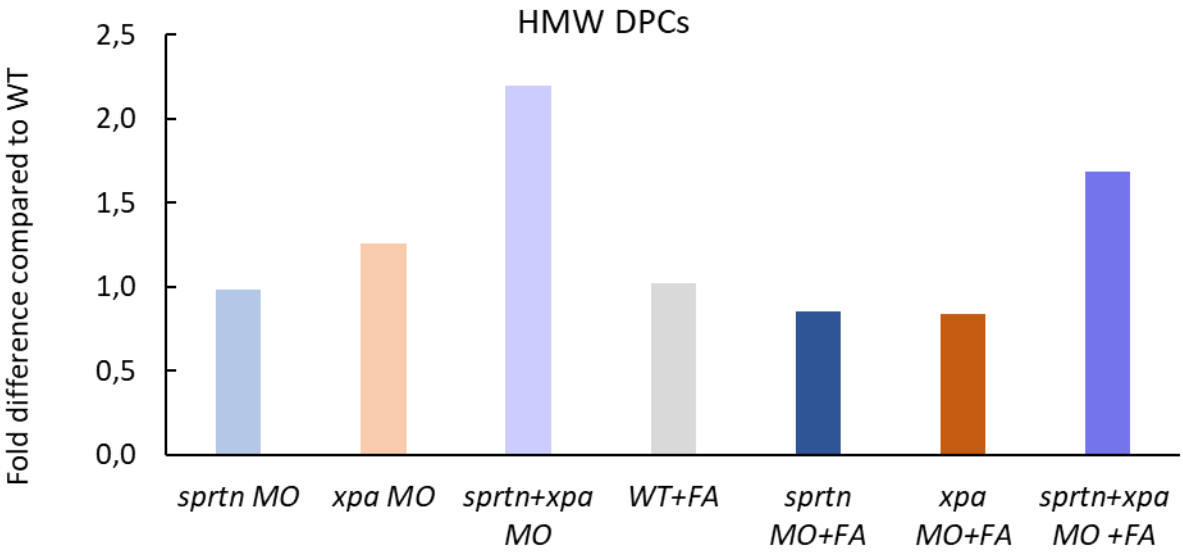
Zebrafish embryos deficient in Xpa protein showed a 1.8-fold increase in accumulation of total DPCs in comparison to WT embryos (Figure 27 B). Attenuation of Sprtn did not cause significant increase in DPC levels (1.1-fold compared with WT). However, in embryos deficient in both Xpa and Sprtn, there is a strong and significant accumulation of total DPCs which increased by 4.1-fold in comparison to WT embryos (Figure 27 B). Embryos deficient in either Xpa or Sprtn and treated with a general DPC inducer formaldehyde showed slight increase in total DPCs (1.2-fold compared to WT), whereas embryos deficient in both proteins and treated with formaldehyde had 2.2-fold more of total DPCs than WT embryos (Figure 27 B).

(B)



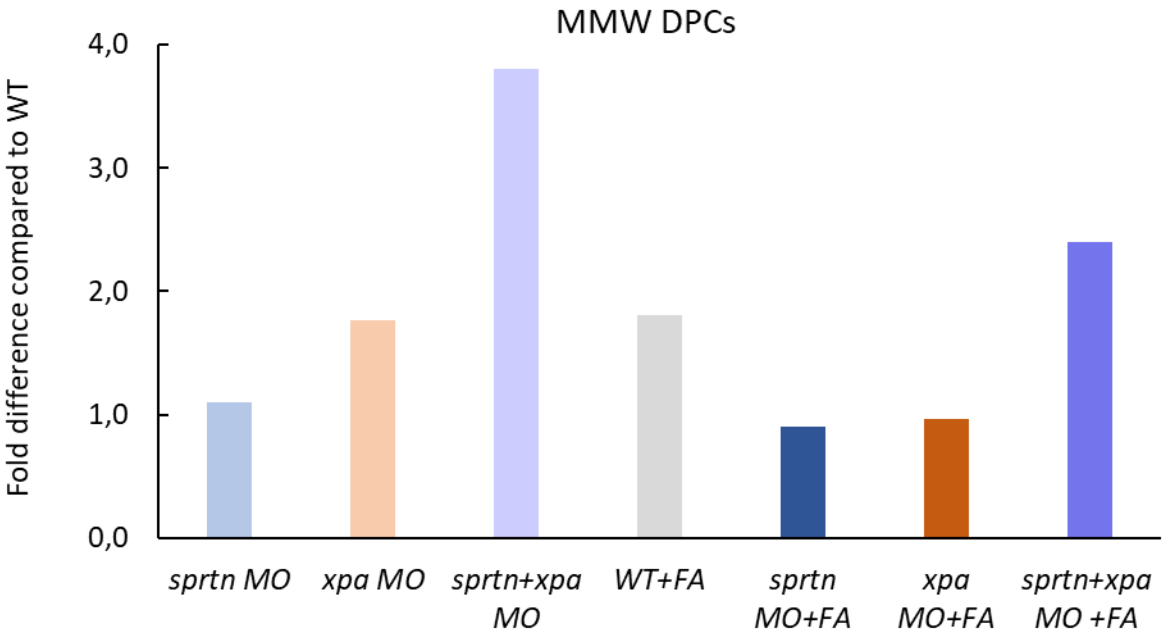
Quantification of HMW DPCs showed that Xpa deficient embryos accumulated more DPCs than Sprtn deficient embryos compared with WT (1.3-fold versus 1-fold) (Figure 27 C). The greatest increase in HMW DPCs is observed in embryos deficient in both Sprtn and Xpa (2.2-fold compared with WT) (Figure 27 C). Sprtn or Xpa deficient embryos treated with formaldehyde have similar amount of HMW DPCs comparable to WT embryos, whereas embryos that are both Sprtn and Xpa deficient and treated with formaldehyde accumulate 1.7-fold more HMW DPCs than WT embryos (Figure 27 C).

(C)



A similar, but more pronounced pattern was observed in the accumulation of MMW DPCs. Xpa deficient embryos have more MMW DPCs than Sprtn deficient embryos (1.8-fold compared with WT), whereas the highest accumulation is observed in embryos that are both Sprtn and Xpa deficient (3.8-fold compared with WT) (Figure 27 D). Sprtn or Xpa deficient embryos treated with formaldehyde accumulated similar amount of MMW DPCs compared with WT embryos, whereas embryos that are both Sprtn and Xpa deficient and treated with formaldehyde accumulated 2.4-fold more MMW DPCs than WT embryos (Figure 27 D).

(D)



The greatest difference beetwen Sprtn and Xpa is observed in the accumulation of LMW DPCs. Xpa deficient embryos have 3.1 times more LMW DPCs than Sprtn deficient ones. A large increase of 9.5-fold in LMW DPCs was observed in embryos deficient in both Sprtn and Xpa (compared with WT) (Figure 27 E). Formaldehyde causes a similar accumulation of LMW DPCs in Sprtn and Xpa deficient embryos, whereas it causes a 3.7-fold higher accumulation in embryos deficient in both Xpa and Sprtn (Figure 27 E). In summary, Xpa deficient embryos accumulate more DPCs than SPRTN deficient, especially LMW DPCs, while double morphants showed the highest level of all types of DPCs among analyzed samples, also with the highest level of LMW DPCs.

(E)

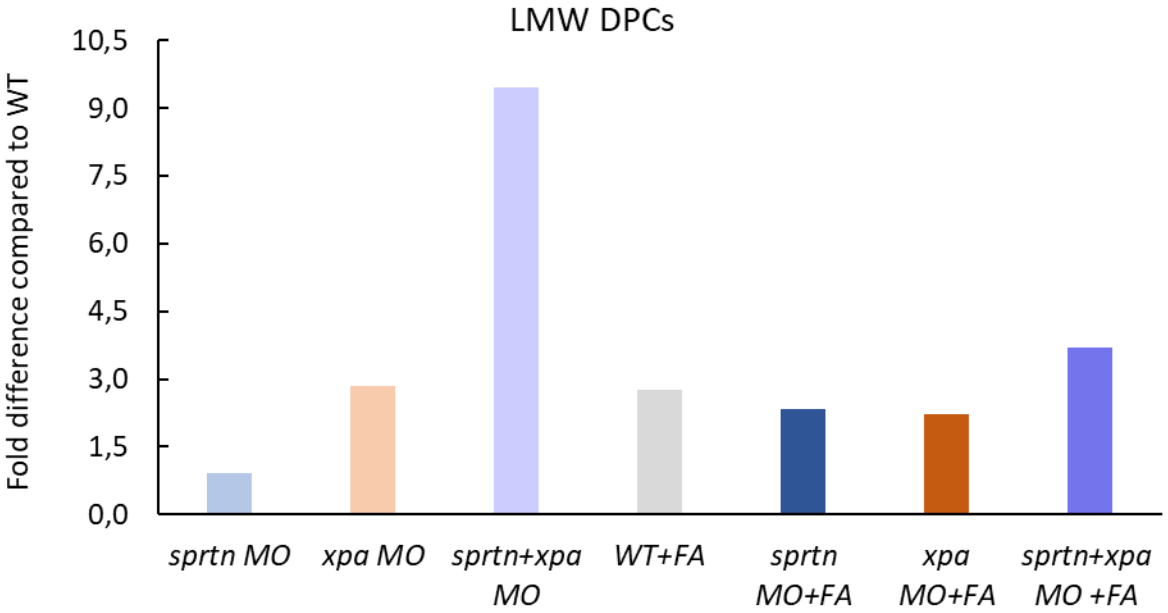


Figure 27. (A) Accumulation of total DPCs in zebrafish embryos after morpholino-mediated SPRTN and XPA silencing and exposure to formaldehyde (1mM, 1h). Cellular DPCs were isolated using the RADAR assay, followed by SDS-PAGE separation and visualization by silver staining. As a loading control, double-stranded DNA (dsDNA) was detected with a specific antibody by slot blot analysis. Quantification of (B) total DPCs, (C) high molecular weight (HMW) DPCs, (D) medium molecular weight (MMW) DPCs and (E) low molecular weight (LMW) DPCs from (A) using ImageJ software.

Table 3.6. DPC sizes for high (HMW), medium (MMW) and low molecular weight (LMW).

DPCs	Size (kDa)
HMW	151 - 250
MMW	41 - 150
LMW	5 - 40

3.8. Creation of a zebrafish Acrc mutant strain carrying enzymatic mutation (ΔEMCH)

To study the role of ACRC in DPC repair, two different zebrafish strains were created using CRISPR/Cas9 genome editing. sgRNA 1 targeting exon 12 and binding to position 16347-16365 on genomic DNA of zebrafish *acrc* gene on chromosome 14 (CDS position 1350-1368) was used with the aim to create Acrc strain carrying enzymatic mutation. sgRNA was designed to target predicted active site residue E451 within the conserved active site in zebrafish Acrc. sgRNA 1 was injected in complex with the Cas9 protein into the yolk of one-cell stage zebrafish embryos. Cas9 protein introduces DNA double-strand break at targeted position. During DSB repair by non-homologous end joining (NHEJ), insertion or deletion of different nucleotide sequences may occur, resulting in different mutations in different embryos and adults. The generated mutated strain of zebrafish was reared and crossed with wild-type individuals after reaching sexual maturity to identify the individuals that passes the acquired mutation to the offspring (founders). High-resolution melting (HRM) analysis was used to distinguish embryos carring a mutation from those without mutations. Embryos that were HRM positive (T_m different from WT embryos) were sequenced to determine the type of mutation. Individuals carrying a mutation in the *acrc* gene were crossed with each other to obtain an F1 generation. Individuals in the F1 generation were genotyped by analyzing genomic DNA from tail fin tissue using PCR and sequencing for the purpose of finding homozygous individuals (male and female) for the target mutation. Homozygous individuals were crossed to obtain homozygous F2 progeny (Figure 16). At the genomic level, we induced a 12 nt deletion at positions 16349 - 16360 corresponding to the mutation in ACRC protease core: a deletion of the catalytic glutamate E451 and the following three amino acids including the Zn-bearing histidine H454, which together form the HEXXH motif (Figures 28 and 29).

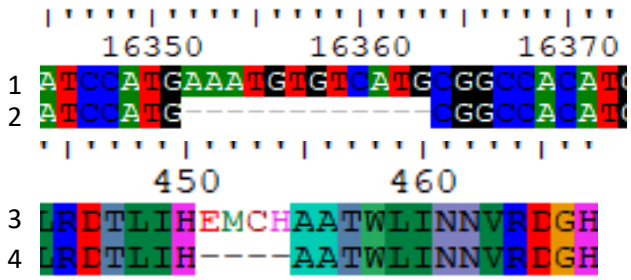


Figure 28. ACRC mutation on genomic and protein level in Acrc mutant strain carrying enzymatic mutation. The 12 nt deletion on the genomic DNA corresponds to deletion of catalytic glutamate E451 and following three amino acids in the protease core including Zn-bearing histidine H454 which together form HEXXH motif. The first row of the above figure (1) shows the consensus nucleotide sequence of the *acrc* zebrafish gene, and the second row (2) shows 12 nt deletion in genomic DNA. The first row of the figure below (3) shows the consensus amino acid sequence of the zebrafish ACRC protein, and the second row (4) shows the translated amino acid sequence resulting from 12 nt deletion. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_ 001013591.1).

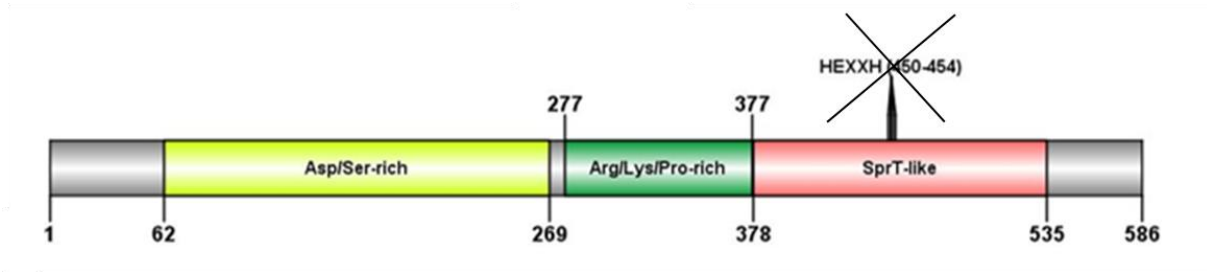


Figure 29. Topology of the zebrafish Acrc protein with the indicated enzymatic ΔEMCH mutation. Acrc protein has a large N-terminal intrinsically disordered region (IDR), encompassing Asp/Ser-rich and Arg/Lys/Pro-rich regions and an SprT-like domain in the C-terminal part of the protein containing the HEXXH protease core.

3.8.1. Phenotyping of Acrc mutant strain carrying enzymatic mutation (ΔEMCH)

The progeny of Acrc heterozygous mutant mothers carrying enzymatic mutation are viable without any apparent phenotypes, while progeny of homozygous mothers die within the first 12 hours post fertilization (Figure 30) (Cecile Otten, unpublished results). This phenotype is similar to the recently reported phenotypes of the Acrc mutant zebrafish strains (Bhargava et al, 2020), which would imply that deletion of catalytic glutamate E451 and proteolytic inactivity of Acrc are responsible for the observed phenotype.

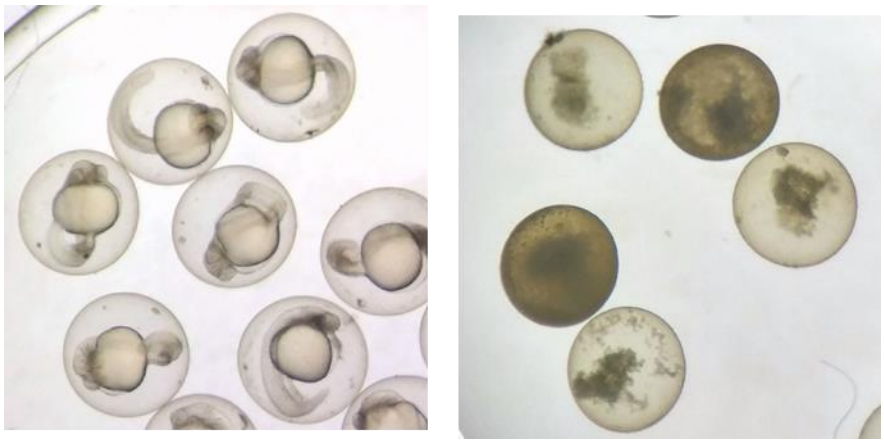
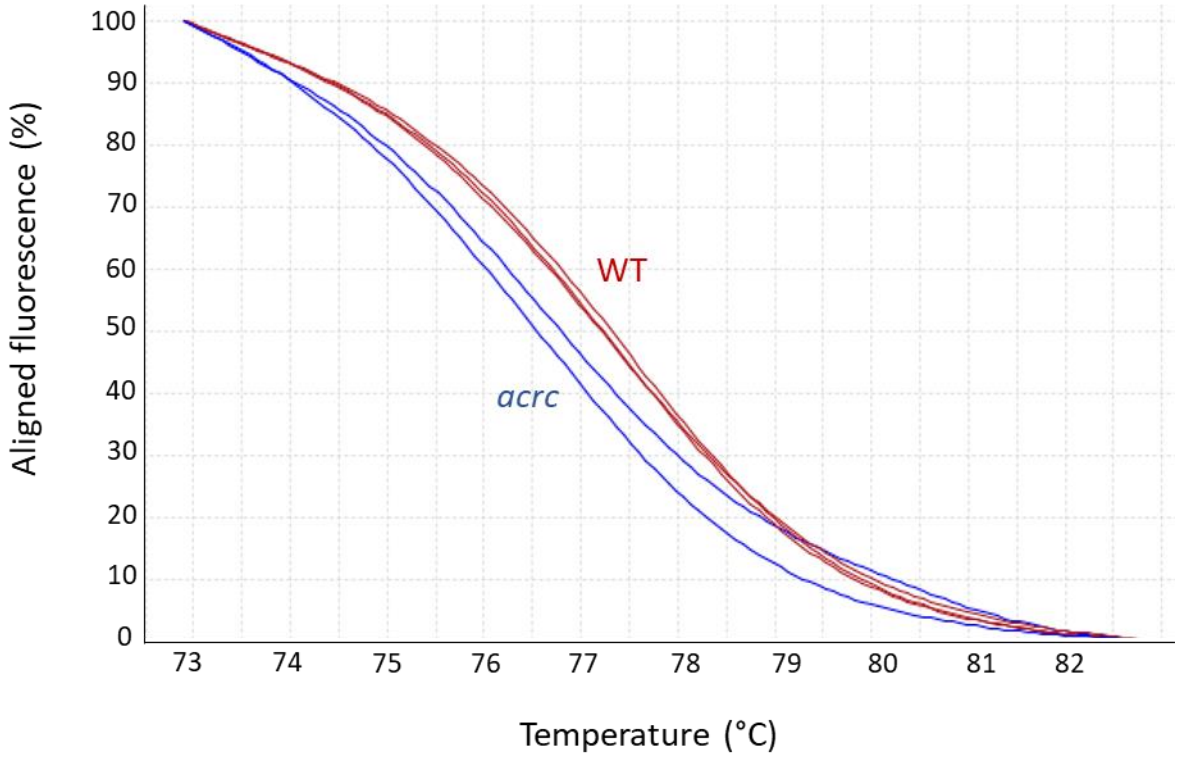


Figure 30. Representative pictures of embryos heterozygous for the Acrc enzymatic mutation, derived from heterozygous mutant females (left) or homozygous females (right) crossed with WT (1 day post fertilization, n = 62). Embryos obtained from heterozygous mutant mothers are viable, show no phenotypic changes, and 97% of embryos are similar to WT embryos. Progeny from the homozygous mutant mothers die 12 hours post fertilization (Cecile Otten, unpublished results).

3.8.2. Genotyping of Acrc mutant strain carrying enzymatic mutation (ΔEMCH) F0 generation

Embryos obtained by crossing F0 *acrc* mutant male fish and WT female fish were screened for the presence of mutations by HRM analysis to identify male founder. *Acrc* mutant F0 male had two HRM positive embryos that differed in T_m from WT embryos (Figure 31 A and Table 3.7). Out of 10 embryos collected from crossing F0 *acrc* mutant male fish and WT female fish, melting curves analysed by HRM software showed that 8 embryos were potentially WT and 2 embryos potentially carried a mutation. The presence of mutations was also confirmed by cloning and sequencing. We have sequenced embryo 1 and 2 (E1, E2) in which a 4 nt deletion in exon 12 at position 16355-16358 in the genomic DNA was detected (Figure 31 B). A 4 nt deletion in the F0 *acrc* mutant male fish results in a mutation of E451catalytic glutamate to lysine (E451K) and the following three histidine residues, which are in SPRTN important for Zn binding, to glycine, leucine and proline. Also, a premature stop codon is present at position 463 at a protein level (Figure 31 B). This male was crossed with WT female to obtain the F1 generation.

(A)



(B)

Figure 31. (A) Genotyping of F0 *acrc* mutant male fish using HRM analysis. Aligned melt curves are shown determined by HRM analysis in high-resolution melting software for WT embryos and progeny from *acrc*-injected fish crossed with WT fish. The samples that the software recognized as different genomic variants are marked in red (WT) or blue (*acrc* mutation). (B) Genotyping of F0 *acrc* mutant male fish by cloning and sequencing. The 4 nt deletion on the genomic DNA corresponds to a frameshift mutation and a premature stop codon at position 463 (red arrow). Aligned nucleotide sequence (above image) and amino acid sequence (bottom image) for the sequenced DNA amplicon of the E1 and E2 embryo sample are shown. The first row of the above figure (1) shows the consensus nucleotide sequence of the *acrc* zebrafish gene, and the second row (2) shows the nucleotide sequence of the sequenced DNA amplicons. The first row of the figure below (3) shows the consensus amino acid sequence of the zebrafish ACRC protein, and the second row (4) shows the translated amino acid sequence from the sequenced DNA amplicons. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_ 001013591.1).

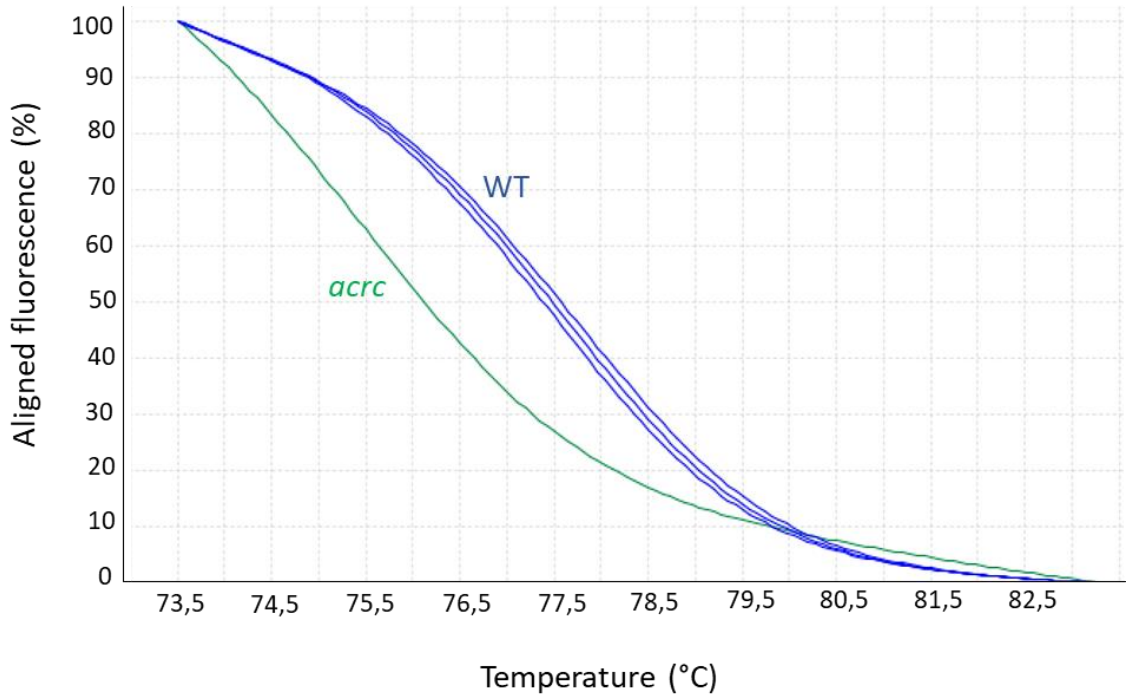
Table 3.7. Melting temperatures (°C) determined by HRM analysis in high-resolution melting software for WT and acrc mutant embryos. T_m - melting temperature, WT - wild type embryo sample, *acrc* E1 and E2 - embryos obtained by crossing the injected *acrc* F0 mutant male fish with a wild-type female fish.

Sample	**T_m (°C)**
WT1	75,5
WT2	75,5
WT3	75,4
acrc E1	75,0
acrc E2	75,0

3.8.3. Genotyping of Acrc mutant strain carrying enzymatic mutation (ΔEMCH) F1 generation

The F1 generation was created by crossing F0 *acrc* mutant male fish and WT female fish. After 3 months, individuals of the F1 generation were screened for the presence of mutations by HRM analysis. The F1 *acrc* mutant male fish was crossed with WT female and their progeny was genotyped. The F1 *acrc* mutant male fish had one HRM positive embryo that differed in T_m from WT embryos (Figure 32 A and Table 3.8). Out of 10 embryos collected from crossing F1 *acrc* mutant male fish and WT female, melting curves analysed by HRM software showed that 9 embryos were potentially WT and 1 embryo potentially carried a mutation. The presence of mutation in this individual was also confirmed by cloning and sequencing. Sequencing of the embryo E1 showed the presence of a 12 nt deletion at position 16349-16360 in the genomic DNA, which results in a deletion of catalytic glutamate E451 and following three amino acids in the protease core including Zn-bearing histidine H454 (Figure 32 B).

(A)



(B)

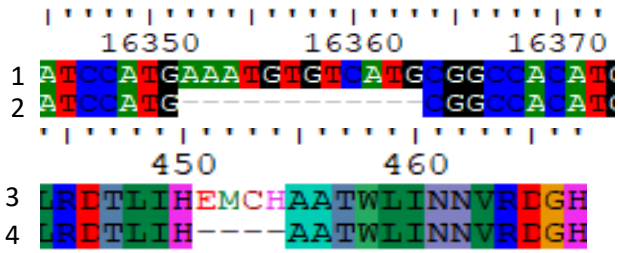


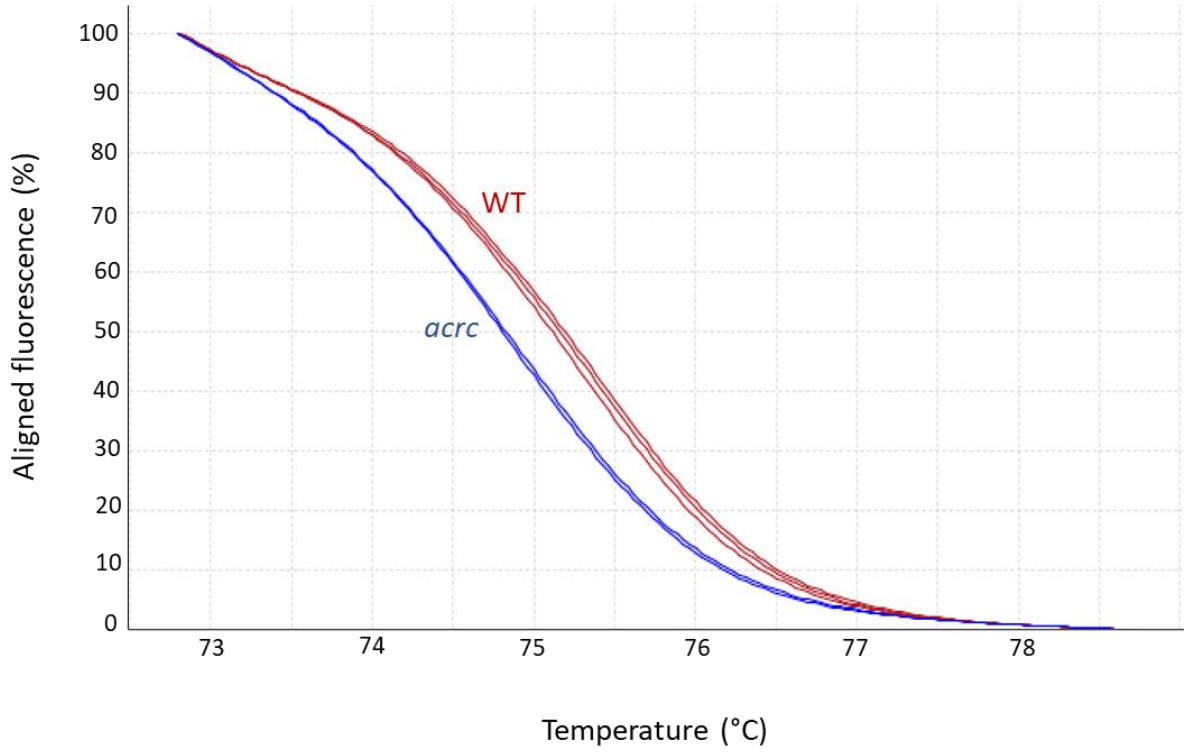
Figure 32. (A) Genotyping of F1 *acrc* mutant male fish by HRM analysis. Aligned melt curves are shown determined by HRM analysis in high-resolution melting software for WT embryos and progeny from F1 *acrc* mutant male fish crossed with WT female fish. The samples that the software recognized as different genomic variants are marked in blue (WT) or green (*acrc* mutation). (B) Genotyping of F1 *acrc* mutant male fish by cloning and sequencing. The 12 nt deletion on the genomic DNA corresponds to a deletion of catalytic glutamate E451 and following three amino acids in the protease core including Zn-bearing histidine H454. Aligned nucleotide sequence (above image) and amino acid sequence (bottom image) for the sequenced DNA amplicon of the E1 embryo sample are shown. The first row of the above figure (1) shows the consensus nucleotide sequence of the *acrc* zebrafish gene, and the second row (2) shows the nucleotide sequence of the sequenced DNA amplicon. The first row of the figure below (3) shows the consensus amino acid sequence of the zebrafish ACRC protein, and the second row (4) shows the translated amino acid sequence from the sequenced DNA amplicon. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_ 001013591.1).

Table 3.8. Melting temperatures (°C) determined by HRM analysis in high-resolution melting software for WT and *acrc* mutant embryos. T_m - melting temperature, WT - wild type embryo sample, *acrc* E1 - embryo obtained by crossing the F1 *acrc* mutant male fish with a wild-type female fish.

Sample	**T_m (°C)**
WT1	75,5
WT2	75,6
WT3	75,6
acrc E1	75,2

The F1 *acrc* mutant female fish were genotyped by analysis of DNA from fin tail tissue. Melting curves analysed by HRM software showed that out of 8 analysed samples, 6 females were potentially WT and 2 potentially carried a mutation. Aligned melt curves are shown for two F1 *acrc* females that were HRM positive with a T_m different from that of DNA from fin tail tissue of the WT fish (Figure 33 A and Table 3.9). The presence of mutation in this two individuals was also confirmed by cloning and sequencing. Sequencing of the DNA from fin tail tissue of F1 *acrc* mutant female fish showed the presence of a 12 nt deletion at position 16349-16360 in the genomic DNA, which results in a deletion of catalytic glutamate E451 and following three amino acids in the protease core including Zn-bearing histidine H454 (Figure 33 B). F1 *acrc* mutant FC1 female fish was crossed with F1 *acrc* mutant male fish to obtain F2 generation.

(A)



(B)

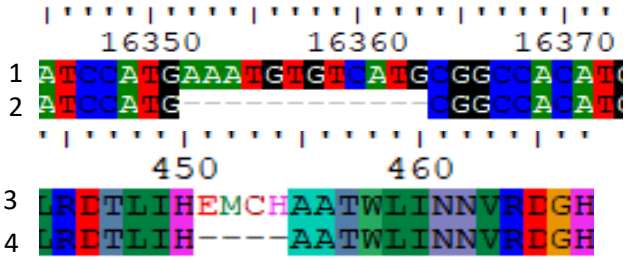


Figure 33. (A) Genotyping of F1 *acrc* mutant female fish by HRM analysis. Aligned melt curves are shown determined by HRM analysis in high-resolution melting software for DNA obtained by lysis of tail fin tissue from WT and F1 *acrc* mutant female fish. The samples that the program recognized as different genomic variants are marked in red (WT) or blue (*acrc* mutation). (B) Genotyping of F1 *acrc* mutant female fish by cloning and sequencing. The 12 nt deletion on the genomic DNA corresponds to a deletion of catalytic glutamate E451 and following three amino acids in the protease core including Zn-bearing histidine H454. Aligned nucleotide sequence (above image) and amino acid sequence (bottom image) for the sequenced DNA isolated from tail fin tissue of F1 *acrc* mutant female fish are shown. The first row of the above figure (1) shows the consensus nucleotide sequence of the *acrc* zebrafish gene, and the second row (2) shows the nucleotide sequence of the sequenced DNA amplicon. The first row of the figure below (3) shows the consensus amino acid sequence of the zebrafish ACRC protein, and the second row (4) shows the translated amino acid sequence

from the sequenced DNA amplicon. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_ 001013591.1).

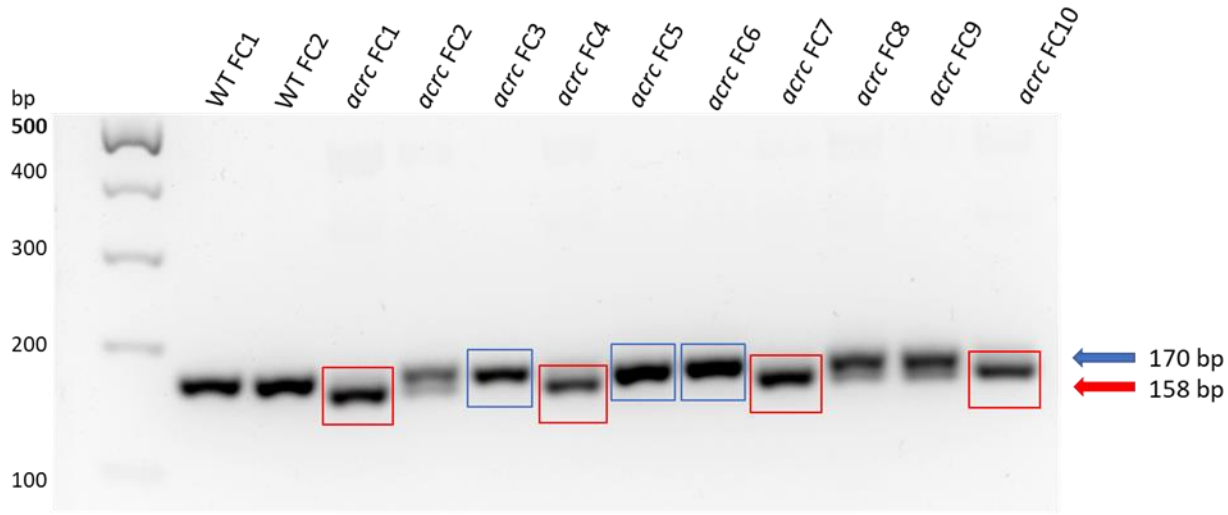
Table 3.9. Melting temperatures (°C) determined by HRM analysis in high-resolution melting software for WT and *acrc* mutant embryos. T_m - melting temperature, WT - wild type embryo sample, *acrc* FC1 and FC2 - DNA isolated from tail fin tissue of F1 *acrc* mutant female fish.

Sample	**T_m (°C)**
WT1	75,3
WT2	75,3
WT3	75,3
acrc FC1	74,9
acrc FC2	74,9

3.8.4. Genotyping of Acrc mutant strain carrying enzymatic mutation (ΔEMCH) F2 generation

F1 *acrc* mutant FC1 female fish was crossed with F1 *acrc* mutant male fish to obtain F2 generation. Individuals from F2 generation were genotyped by analysis of DNA from fin tail tissue to find homozygous male and female for crossing and obtaining 100% homozygous F3 generation. Since by HRM analysis we could not distinguish heterozygotes from homozygotes, we decided to perform PCR reaction on genomic DNA isolated from fin tail tissue of individuals from F2 generation to differentiate WT fish (WT band at 170 pb), homozygous fish (mutant band at 158 pb) and heterozygous fish (1 band at 170 pb and one at 158 pb). Products of the PCR reaction were run on 2% agarose gel. Out of 10 tested individuals, 3 fish were WT, 3 were heterozygous and 4 were homozygous (Figure 34 A). The presence of mutation in this four homozygous individuals was also confirmed by cloning and sequencing (Marin Kutnjak, unpublished results). Sequencing of the DNA from fin tail tissue of four homozygous F2 fish showed the presence of a 12 nt deletion at position 16349-16360 in the genomic DNA, resulting in a deletion of catalytic glutamate E451 and following three amino acids in the protease core including Zn-bearing histidine H454 (Figure 34 B). One homozygous male and female were selected for crossing to obtain homozygous F3 generation ($ACRC^{\Delta EMCH/\Delta EMCH}$).

(A)



(B)

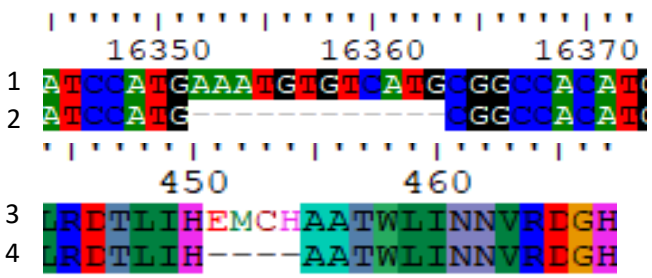


Figure 34. A) PCR reaction on DNA isolated from fin tail tissue of individuals from F2 generation used to differentiate WT, homozygous and heterozygous fish. B) Genotyping of homozygous F2 *acrc* mutant fish by cloning and sequencing. The 12 nt deletion on the genomic DNA corresponds to a deletion of catalytic glutamate E451 and following three amino acids in the protease core including Zn-bearing histidine H454. Aligned nucleotide sequence (above image) and amino acid sequence (bottom image) for the sequenced DNA isolated from tail fin tissue of F2 *acrc* mutant homozygous fish are shown. The first row of the above figure (1) shows the consensus nucleotide sequence of the *acrc* zebrafish gene, and the second row (2) shows the nucleotide sequence of the sequenced DNA amplicons. The first row of the figure below (3) shows the consensus amino acid sequence of the zebrafish ACRC protein, and the second row (4) shows the translated amino acid sequence from the sequenced DNA amplicon. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_ 001013591.1) (Marin Kutnjak, unpublished results).

3.9. Creation of Acrc mutant zebrafish strain carrying C-terminal deletion (ΔC)

To study the role of ACRC in DPC repair and to explore the consequence of disrupted SprT protease domain and protein folding on ACRC activity, in addition to the zebrafish strain carrying enzymatic mutation, a zebrafish strain carrying C-terminal deletion was also created using CRISPR/Cas9 genome editing. sgRNA 2 targeting exon 12 and binding to position 16405-16423 on genomic DNA of zebrafish *acrc* gene (CDS position 1408-1426) was used to create Acrc mutant zebrafish strain carrying C-terminal deletion (ΔC). sgRNA 2 was injected in complex with the Cas9 protein into the yolk of one-cell stage zebrafish embryos. The generated mutated strain of zebrafish was reared and crossed with wild-type individuals after reaching sexual maturity to identify the individuals that passes the acquired mutation to the offspring (founders). High-resolution melting (HRM) analysis was used to distinguish embryos carrying a mutation from those without mutations. Embryos that were HRM positive (T_m different from WT embryos) were sequenced to determine the type of mutation. Individuals carrying a mutation in the *acrc* gene were crossed with each other to obtain an F1 generation. Individuals in the F1 generation were genotyped by analyzing genomic DNA from tail fin tissue using PCR and sequencing for the purpose of finding homozygous individuals (male and female) for the target mutation. Homozygous individuals were crossed to obtain homozygous F2 progeny (Figure 16). At the protein level, ΔC zebrafish strain bears C-terminal deletion of amino acid 473. -END (protein is 586 amino acid long) (Figures 35 and 36).

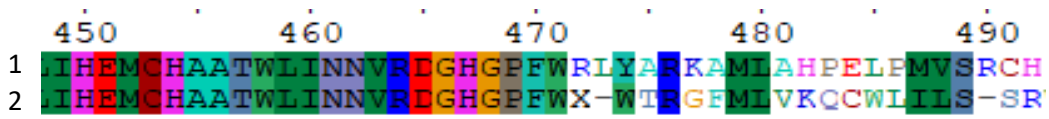


Figure 35. Acrc mutation at a protein level in Acrc mutant zebrafish strain carrying C-terminal deletion (ΔC). Premature stop codon is introduced at position 473 resulting with a frameshift mutation in C-terminal part of the protein (after protease core and HEXXH motif). The first row of the figure (1) shows the consensus amino acid sequence of the zebrafish Acrc protein, and the second row (4) shows the translated amino acid sequence containing frameshift mutation in C-terminal part of the protein. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_ 001013591.1).

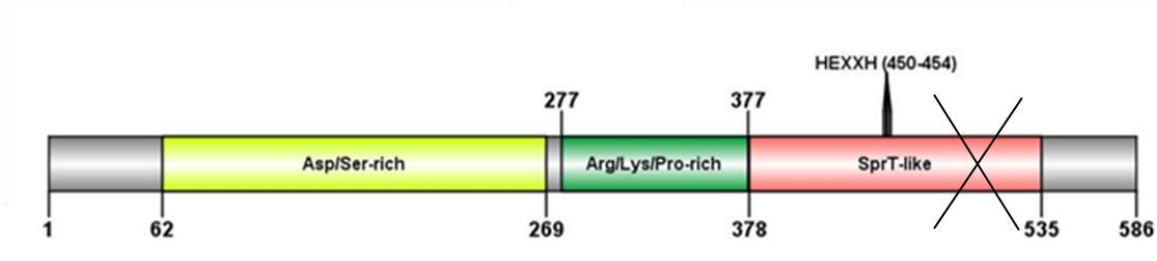


Figure 36. Toplogy of the zebrafish Acrc protein with the indicated C-terminal deletion (ΔC) mutation. Acrc protein has a large N-terminal intrinsically disordered region (IDR),

encompassing Asp/Ser-rich and Arg/Lys/Pro-rich regions and an SprT-like domain in the C-terminal part of the protein containing the HEXXH protease core.

3.9.1. Phenotyping of Acrc mutant zebrafish strain carrying C-terminal deletion (ΔC)

Heterozygous embryos which inherited the mutant allele from the mother dye within the first 12 hours post fertilization (Figure 37). This phenotype is similar to the recently reported phenotypes of the Acrc mutant zebrafish strain (Bhargava et al, 2020) and homozygous embryos carrying enzymatic mutation, which would imply that deletion of specifically Sprt domain C-terminal part is responsible for the adverse phenotype. On the contrary, authors previously hypothesized that possibly intrinsically disordered N-terminal part (IDR) is responsible for the embryonic lethality. It remains unclear why specifically the mothers' mutant allele causes the phenotype.

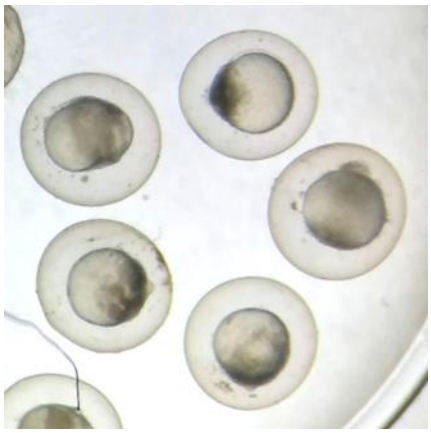


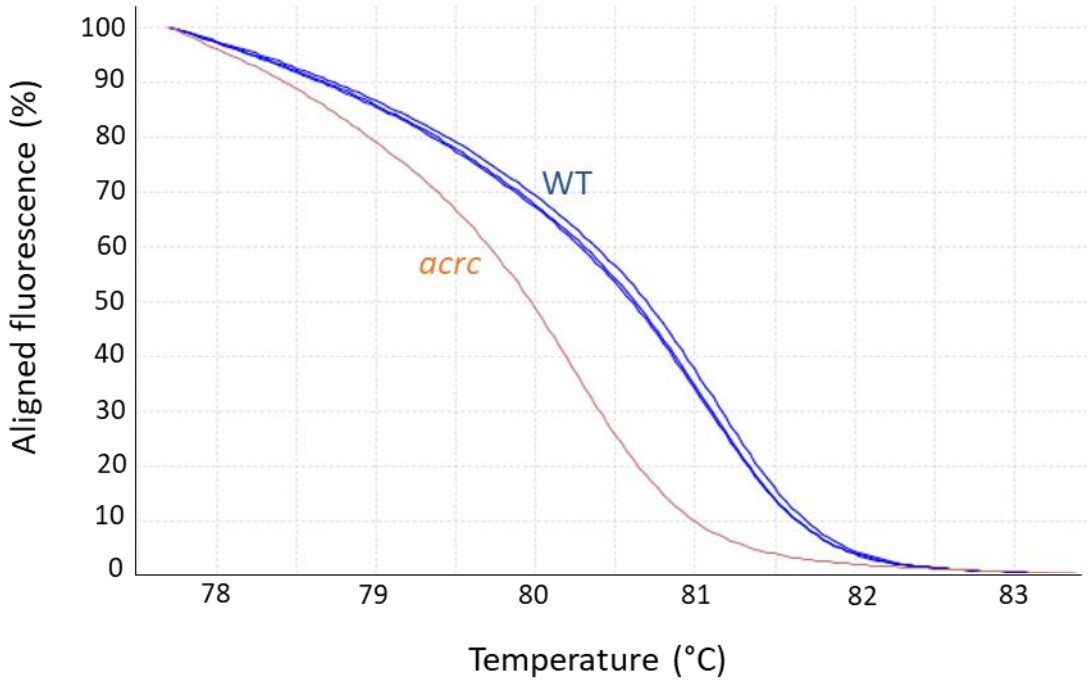
Figure 37. Representative pictures of Acrc mutant zebrafish embryos carrying C-terminal deletion (ΔC) (1 day post fertilization, n = 62). Heterozygous embryos which inherited the mutant allele from the mother and therefore lack maternally deposited Acrc protein die 12 hours post fertilization.

3.9.2. Genotyping of Acrc mutant zebrafish strain carrying C-terminal deletion (ΔC) F0 generation

Embryos obtained by crossing F0 *acrc* mutant female fish and WT male fish were screened for the presence of mutations by HRM analysis to identify female founder. Embryos were collected 6 hpf (hours post fertilization) while they were still viable. *Acrc* mutant F0 female had one HRM positive embryo that differed in T_m from WT embryos (Figure 38 A and Table 3.10). Out of 10 embryos collected from crossing F0 *acrc* mutant female fish and WT male fish, melting curves analysed by HRM software showed that 9 embryos were WT and 1 embryo carried a mutation. The presence of mutation was also confirmed by cloning and sequencing. We have sequenced embryo 1 (E1) in which two different mutations were detected: a 19 nt deletion in exon 12 at position 16402-16420 and 14 nt insertion at position 16407 on genomic DNA (Figure 38 B). A 19 nt deletion results in a frameshift mutation and a

premature stop codon at position 529 at the protein level, while 14 nt insertion results with frameshift mutation and premature stop codon at position 473 (Figure 38 B). Both mutations are located in the C-terminal part of the Acrc protein, including also a part of SPRTN-like domain (Figure 39). This F0 *acrc* mutant female fish was crossed with WT male fish to obtain the F1 generation.

(A)



(B)

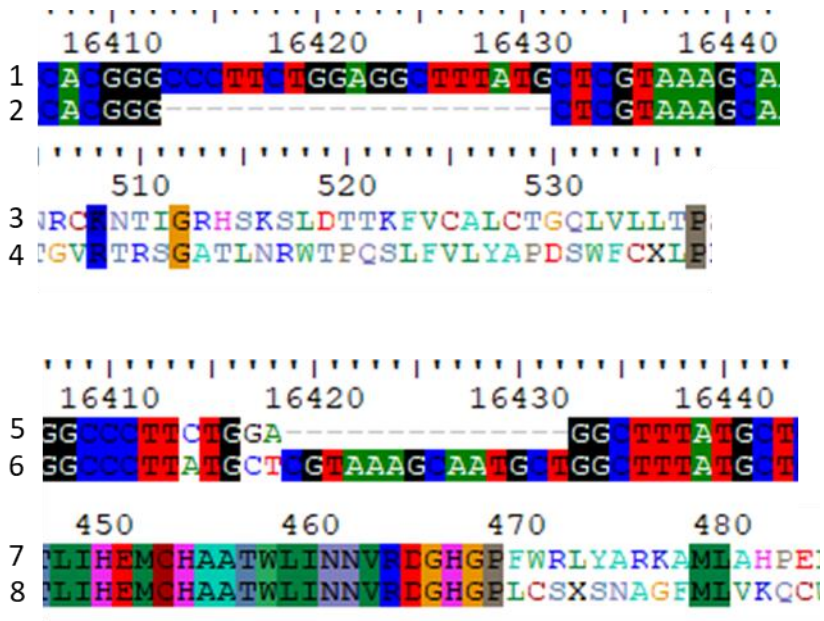


Figure 38. A) Genotyping of F0 *acrc* mutant female fish using HRM analysis. Aligned melt curves are shown determined by HRM analysis in high-resolution melting software for WT embryos and progeny from *acrc* -injected fish crossed with WT fish. The samples that the software recognized as different genomic variants are marked in blue (WT) or orange (*acrc* mutation). (B) Genotyping of F0 *acrc* mutant female fish by cloning and sequencing. Aligned nucleotide sequence (above images) and amino acid sequence (bottom images) for the sequenced DNA amplicon of the E1 embryo sample are shown. The first row of the figures above (1 and 5) shows the consensus nucleotide sequence of the *acrc* zebrafish gene, and the second row (2 and 6) shows the nucleotide sequence of the sequenced DNA amplicon. The first row of the figures below (3 and 7) shows the consensus amino acid sequence of the zebrafish Acrc protein, and the second row (4 and 8) shows the translated amino acid sequence from the sequenced DNA amplicon. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_956765.1).

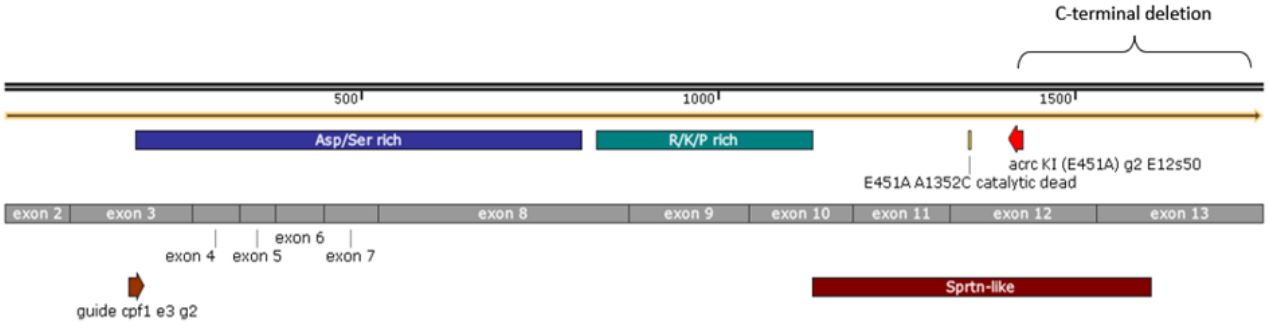


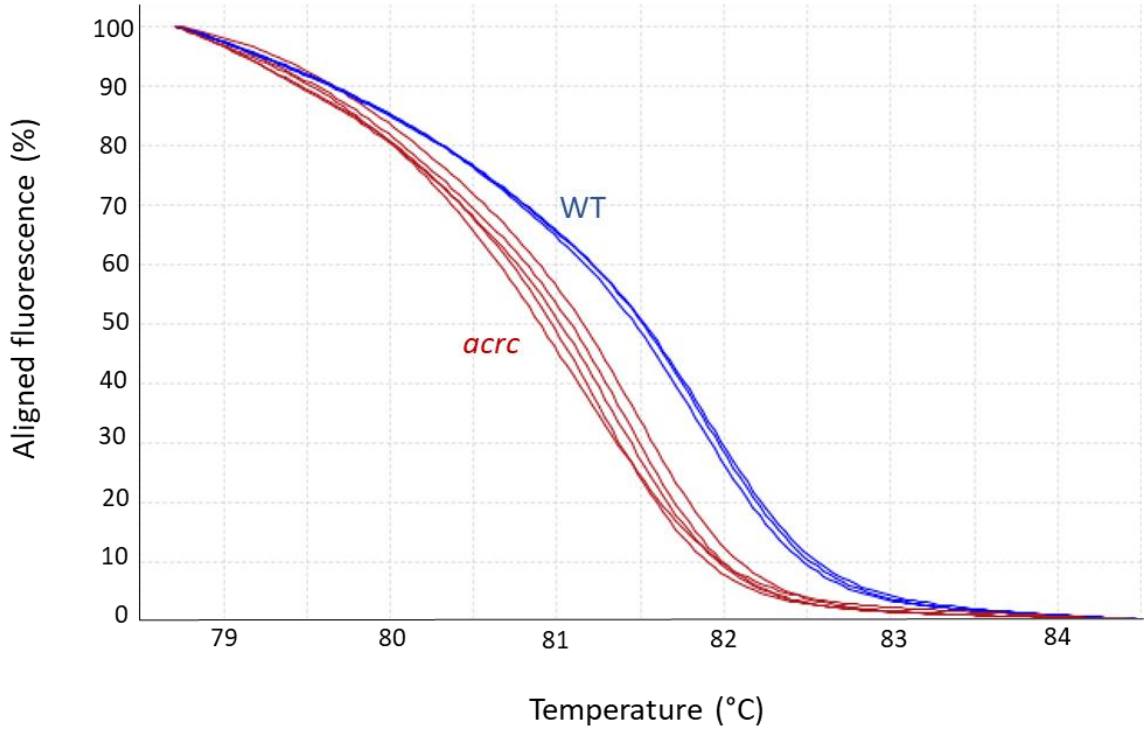
Figure 39. Structure of zebrafish Acrc protein. Deletion of the C-terminal part of the protein involves deletion of part of the Acrc Sprt domain.

Table 3.10. Melting temperatures (°C) determined by HRM analysis in high-resolution melting software for WT and *acrc* mutant embryos. T_m - melting temperature, WT - wild type embryo sample, *acrc* E1 - embryo obtained by crossing the F0 *acrc* mutant female fish with a wild-type male fish.

Sample	**T_m (°C)**
WT1	81,1
WT2	81,1
WT3	81,0
acrc E1	80,2

Embryos obtained by crossing F0 *acrc* mutant male fish and WT female fish were screened for the presence of mutations by HRM analysis to identify male founder. Progeny from *acrc* mutant males was viable without any apparent phenotypes. *acrc*-injected F0 male had five HRM positive embryos that differed in T_m from WT embryos (Figure 40 A and Table 3.11). Out of 10 embryos collected from crossing F0 *acrc* mutant male fish and WT female fish, melting curves analysed by HRM software showed that 5 embryos were potentially WT and 5 embryos potentially carried a mutation. The presence of mutation was also confirmed by cloning and sequencing. We have sequenced embryo 1 (E1) in which insertion of a 1 nt in exon 12 at position 16408 in the genomic DNA was detected, resulting in a frameshift mutation and premature stop codon at position 477 at the protein level (Figure 40 B). F0 male founder was crossed with F0 *acrc* mutant female fish to obtain the F1 generation.

(A)



(B)

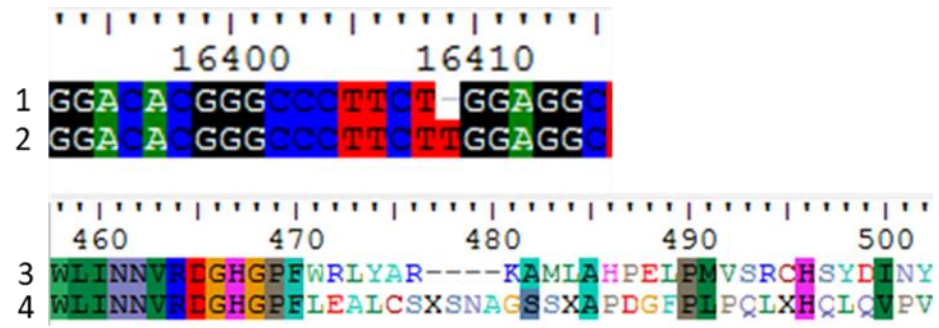


Figure 40. A) Genotyping of F0 *acrc* mutant male fish using HRM analysis. Aligned melt curves are shown determined by HRM analysis in high-resolution melting software for WT embryos and progeny from *acrc* -injected fish crossed with WT fish. The samples that the software recognized as different genomic variants are marked in blue (WT) or red (*acrc* mutation). (B) Genotyping of F0 *acrc* mutant male fish by cloning and sequencing. Aligned nucleotide sequence (above image) and amino acid sequence (bottom image) for the sequenced DNA amplicon of the E1 embryo sample are shown. The first row of the figure above (1) shows the consensus nucleotide sequence of the ACRC zebrafish gene, and the second row (2) shows the nucleotide sequence of the sequenced DNA amplicon. The first row of the figure below (3) shows the consensus amino acid sequence of the zebrafish Acrc protein, and the second row (4) shows the translated amino acid sequence from the sequenced DNA amplicon. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_956765.1).

Table 3.11. Melting temperatures (°C) determined by HRM analysis in high-resolution melting software for WT and *acrc* mutant embryos. T_m - melting temperature, WT - wild type embryo sample, *acrc* E1-E5 - embryos obtained by crossing the F0 *acrc* mutant male fish with a wild-type female fish.

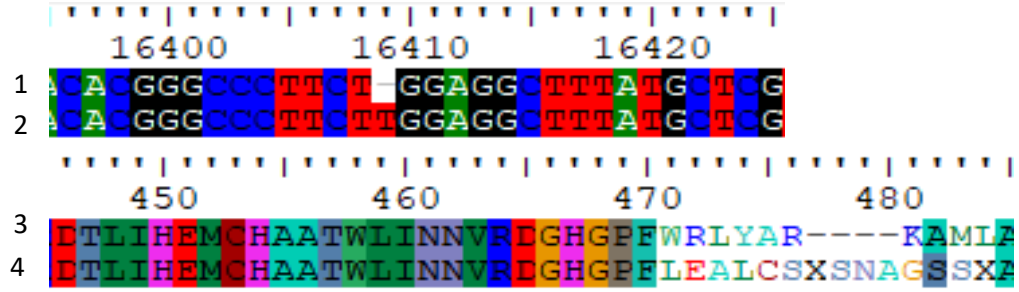
Sample	**T_m (°C)**
WT1	81,9
WT2	81,9
WT3	81,8
acrc E1	81,1
acrc E2	81,2
acrc E3	81,3
acrc E4	81,4
acrc E5	81,4

3.9.3. Genotyping of Acrc mutant zebrafish strain carrying C-terminal deletion (ΔC) F1 generation

F1 generation was created by crossing F0 *acrc* mutant male fish with F0 *acrc* mutant female fish. Since embryos deficient in maternal Acrc protein are unviable, rescue experiment (injection of WT ACRC mRNA) was performed so that the F1 generation could be raised. After 3 months, individuals of F1 generation were screened for the presence of mutations by analysis of DNA from fin tail tissue. Out of 18 tested fish, one homozygous female and male were identified. Homozygous female had insertion of a 1 nt in exon 12 at position 16408 in the genomic DNA, resulting in a frameshift mutation and premature stop codon at position

477 at the protein level on both alleles (Figure 41 A). Homozygous male had same mutation as female on one allele (insertion of a 1 nt in exon 12 at position 16408 in the genomic DNA) and insertion of 11 nt at position 16410 in the genomic DNA on other allele, resulting in a frameshift mutation and premature stop codon at position 472 at the protein level (Figure 41 B) (Cecile Otten, unpublished results). Both mutations are located in the C-terminal part of the ACRC protein, including also a part of SPRTN-like domain. F1 homozygous male and female were crossed to obtain 100% homozygous F2 generation ($ACRC^{\Delta C/\Delta C}$). Embryos of F2 generation were injected with rescue construct to enable their survival. F2 homozygous individuals carrying insertion of a 1 nt (results in a stop codon at position 477) on both alleles or carrying insertion of a 1 nt on one allele and insertion of 11 nt (results in a stop codon at position 472) on other allele were raised to adulthood.

(A)



(B)

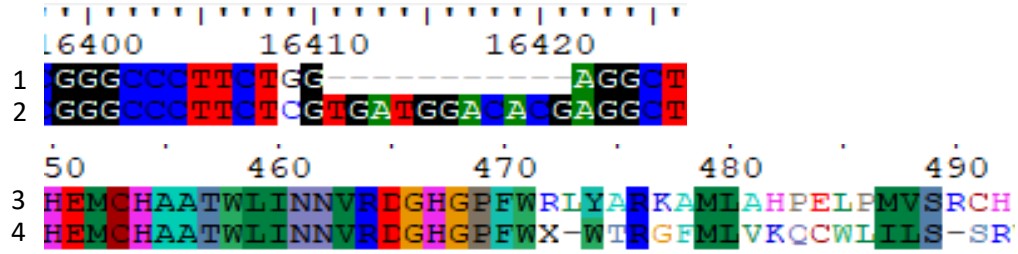


Figure 41. (A) Genotyping of F1 *acrc* homozygous female fish by cloning and sequencing. Aligned nucleotide sequence (above image) and amino acid sequence (bottom image) for the sequenced DNA isolated from tail fin tissue of F1 *acrc* homozygous female fish are shown. The first row of the figure above (1) shows the consensus nucleotide sequence of the *acrc* zebrafish gene, and the second row (2) shows the nucleotide sequence of the sequenced DNA amplicon. The first row of the figure below (3) shows the consensus amino acid sequence of the zebrafish Acrc protein, and the second row (4) shows the translated amino acid sequence from the sequenced DNA amplicon. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_956765.1). (B) Genotyping of F1 *acrc* homozygous male fish by cloning and sequencing. Aligned nucleotide sequence (above image)

and amino acid sequence (bottom image) for the sequenced DNA isolated from tail fin tissue of F1 *acrc* homozygous male fish are shown. The first row of the figure above (1) shows the consensus nucleotide sequence of the *acrc* zebrafish gene, and the second row (2) shows the nucleotide sequence of the sequenced DNA amplicon. The first row of the figure below (3) shows the consensus amino acid sequence of the zebrafish Acrc protein, and the second row (4) shows the translated amino acid sequence from the sequenced DNA amplicon. The sequences were aligned in BioEdit. Consensus sequences were taken from the Ensembl database (NP_956765.1) (Cecile Otten, unpublished results).

3.10. Acrc protein levels in zebrafish Acrc mutant lines: ΔEMCH and ΔC

Acrc protein levels were determined in homozygous embryos and adults of both mutant strains to verify that Acrc protein is present in strain carrying enzymatic mutation (ΔEMCH) and to investigate if the levels are disturbed in strain carrying C-terminal deletion (ΔC). The offspring of homozygous ΔEMCH and ΔC mothers cannot survive without injection of a WT ACRC mRNA, indicating that Acrc protein is essential for early zebrafish embryonic development. As a control, WT embryos were collected 6 hpf (hours post fertilization) and 3 dpf (days post fertilization). ΔEMCH embryos were obtained by crossing F2 homozygous male and female individuals. 40 embryos were collected and lysed 6 hours post fertilization while they were still viable. To characterize ΔC zebrafish strain, tail fin tissues from 10 homozygous adults were lysed and used for analysis. In WT and ΔEMCH embryos, there is a monomeric Acrc protein of the expected size (76 kDa). In additon, posttranslationally modified Acrc protein is present above monomeric form in WT and ΔEMCH embryos (Figure 42). The expected size for the truncated Acrc protein with the ΔC mutation is 52 kDa. In ΔC adults, there is a significant decrease in amount of monomeric Acrc protein that is a slighty bigger than expected (around 60 kDa), without posttranslationally modified forms (Figure 31). The results show that a small amount of the truncated Acrc protein is translated in the Acrc ΔC zebrafish strain, whereas a full-length protein is present in ΔEMCH strain. As a housekeeping gene, tubulin (50 kDa) was detected using a specific antibody.

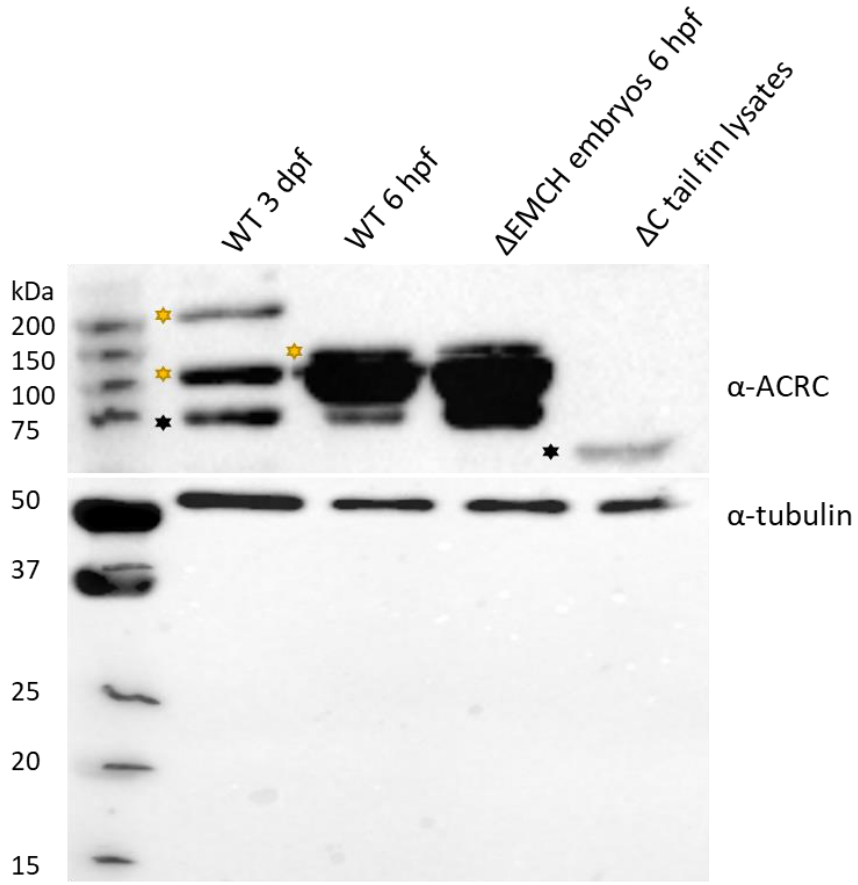


Figure 42. Acrc protein levels in zebrafish Acrc mutant lines (ΔEMCH and ΔC) determined with specific antibody and western blot analysis (n = 40 ΔC embryos and n = 10 adult ΔEMCH tail fin tissues). A monomeric form of Acrc is detected (*) as well as posttranslationally modified Acrc protein (*). As a loading control, specific antibody detecting tubulin was used.

3.11. Rescue experiments in embryos carrying C-terminal deletion (ΔC)

For the C-terminal deletion mutants which exhibit a lethal phenotype when the mutant allele is inherited from the mother, we have developed a system to distinguish which motif, domains, and residues in ACRC are responsible for the lethal phenotype (Cecile Otten, IRB, unpublished results). The mRNA coding for ACRC WT protein was injected into one-cell stage mutant embryos and complete recovery of the phenotype was observed (Figure 43). Rescued embryos were WT-like, they could be raised and were even fertile. After injection of ACRC mRNA with a mutation in putative protease site, specifically E451 catalytic glutamate, the adverse phenotype remained. This observation proves that the catalytic function of ACRC is crucial for embryo survival during embryonic development. Indeed, it implies that ACRC is an active protease, which is to date, the first proof of its protease activity. Given the similarities of SPRTN and ACRC proteases, we wanted to test whether ACRC deficiency can be compensated by the expression of SPRTN protease. Injection of SPRTN WT mRNA into ΔC ACRC embryos failed to rescue the lethal phenotype (Figure 43).

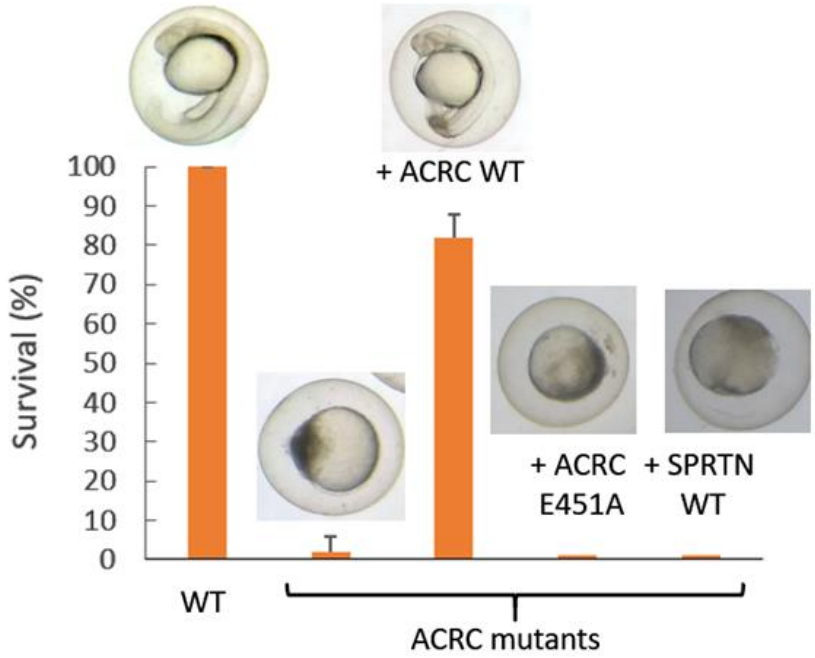


Figure 43. Acrc mutant zebrafish strain carrying C-terminal deletion shows embryonic lethal maternal effect. In a rescue experiment after injection of WT ACRC mRNA viability and survival are fully restored, while the injection of ACRC E451A mRNA or SPRTN WT mRNA could not rescue the phenotype (two independent experiments are shown, n = 50) (Cecile Otten, unpublished results).

After establishing the ACRC rescue system in zebrafish embryos, we set out to identify which domains are important for the ACRC function during embryonic development. 6 different ACRC mRNA were synthesized and injected into WT and mutant embryos at the one-cell stage: WT, E451A mutant, C-terminal deletion mutant, SprT domain mutant, RKP (arginine/lysine/proline-rich)– deletion domain mutant, and IDR (intrinsically disordered region) -deletion mutant. Besides the construct containing WT ACRC, the construct containing deletion of a small part of C-terminus downstream of the Sprt domain (different mutation of C-terminus present in the zebrafish strain carrying C-terminal deletion) was the only ACRC construct tested that could rescue the deadly mutant phenotype, and even this construct caused abnormalities in 40% of injected embryos (Figure 44). For all mutant constructs, except the full-length ACRC protein with point mutation E451A, it remains to be determined if the mutant protein is present in the embryos.

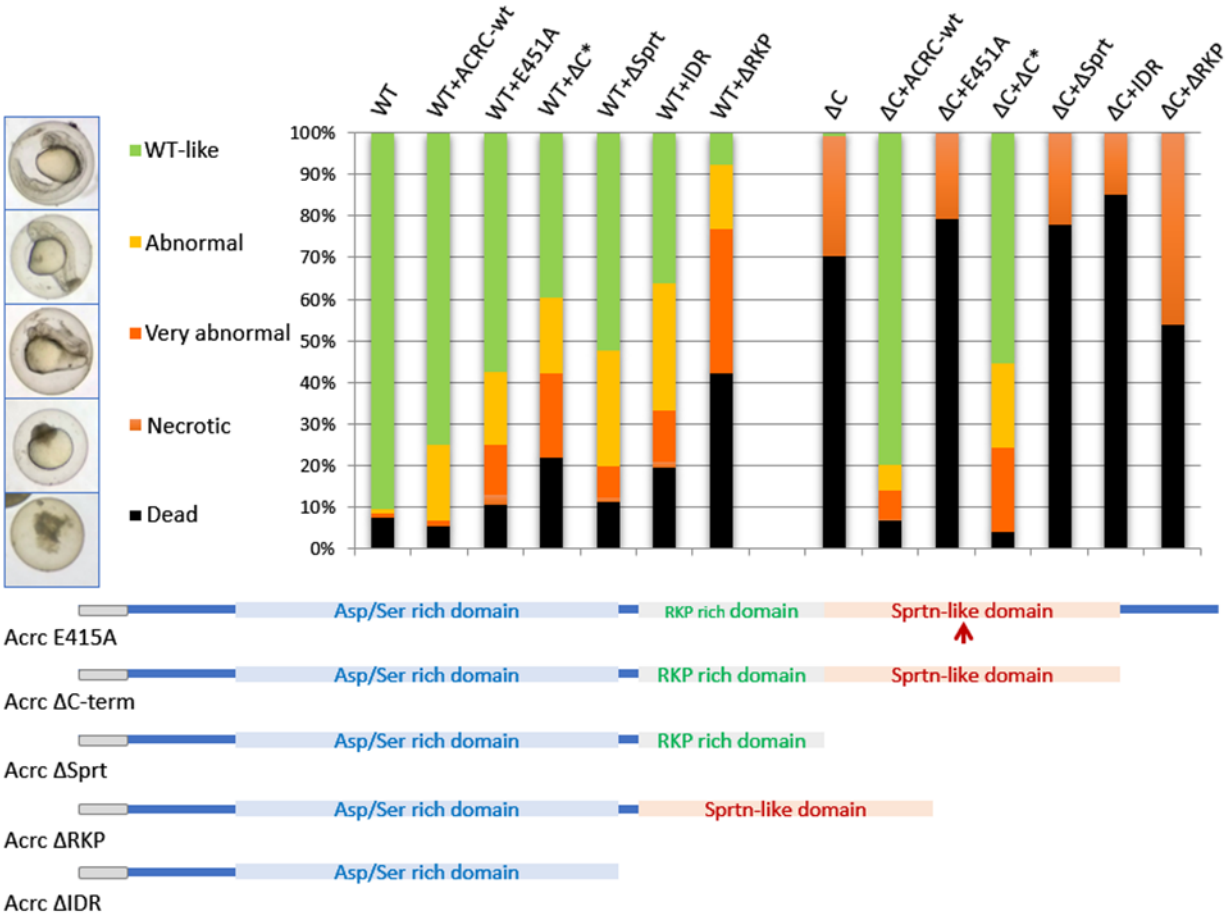
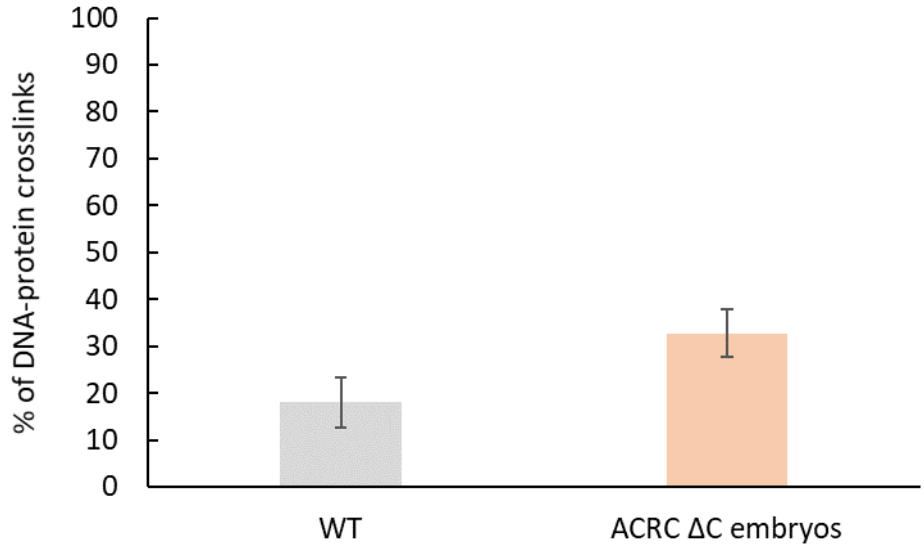


Figure 44. Phenotyping of ACRC$^{\Delta C/\Delta C}$ mutants after injections with different ACRC mRNAs (*ΔC designates the rescue construct depicted in the schematics, it is different mutation of C-terminus present in the zebrafish ACRC $^{\Delta C/\Delta C}$ strain) (Cecile Otten, unpublished results).

3.12. DPC levels in Acrc deficient zebrafish embryos

Total DPC levels in ACRC mutant embryos of ΔEMCH and ΔC strains was measured using SDS/KCl precipitation assay (Marin Kutnjak, unpublished results). F0 *acrc* ΔC female fish was crossed with WT male fish and embryos were collected at 6 hpf (n = 100 embryos per sample). ACRC ΔC mutant embryos accumulate 1.8-fold more DPCs than WT embryos (Figure 45 A). F1 heterozygous *acrc* ΔEMCH male and female fish were crossed and 10 embryos were collected at 3 dpf. The expected ratio of progeny obtain by crossing two heterozygous individuals was 25% of WT, 50% of heterozygous, and 25% of homozygous embryos for the ΔEMCH mutation. WT and ΔEMCH embryos were also treated with formaldehyde (5 mM, 1h). WT, ΔEMCH, and WT embryos treated with formaldehyde (WT+FA) showed similar amount of total DPCs. ΔEMCH embryos treated with formaldehyde have 2.2-fold more DPCs than WT embryos and WT embryos treated with formaldehyde (Figure 45 B).

(A)



(B)

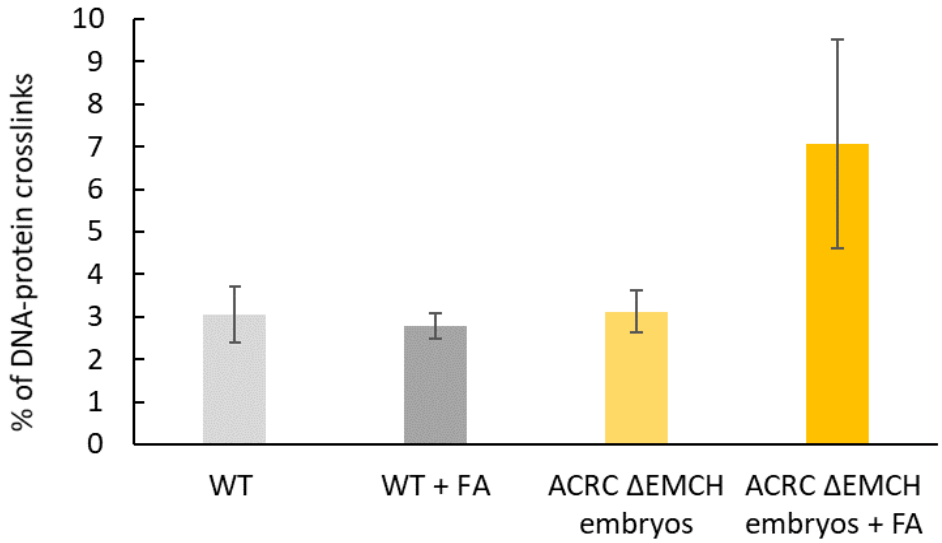


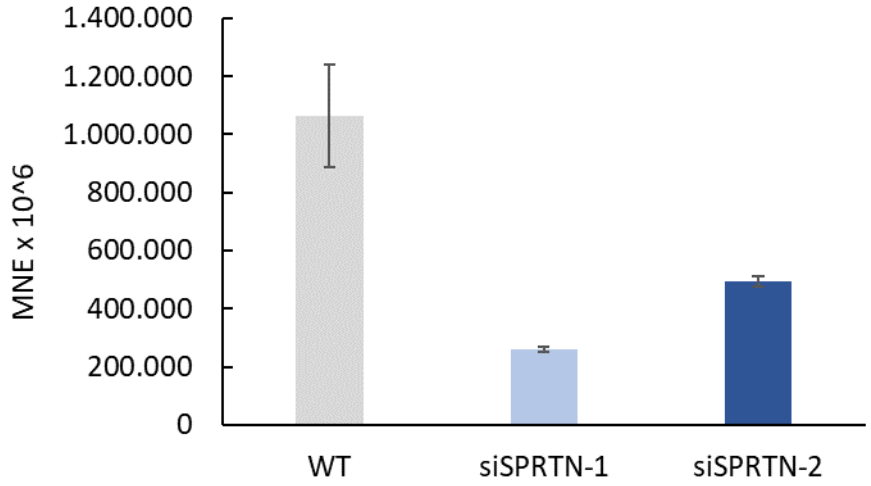
Figure 45. A) Total DPC levels in embryos carrying C-terminal deletion (ΔC) at 6 hours post fertilization (6 hpf) (n = 100 embryos per sample). B) Total DPC levels in in embryos carrying enzymatic mutation (ΔEMCH) at 3 days post fertilization (3 dpf) (n=10 embryos per condition) before and after exposure to formaldehyde (FA). DPCs were isolated using SDS/KCl precipitation assay. The percentage of DNA-protein crosslinks is shown as the mean of the three independent biological replicates ± standard errors of mean (SEM) (Marin Kutnjak, unpublished results).

3.13. Optimization of *XPA, SPRTN* and *ACRC* silencing in human cells

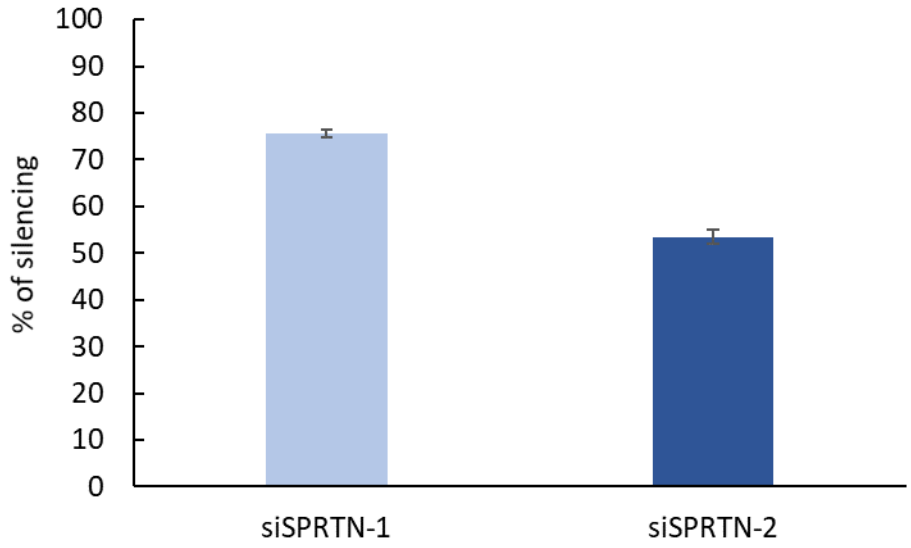
Silencing of *XPA, SPRTN* and *ACRC* by transfection of small interfering RNA (siRNA) with the commercially available liposomal transfection reagent DharmaFECT 1 was optimized in RPE1 (Retinal Pigment Ephitilial-1) and HEK293T (Human Embryonic Kidney 293) cells. HEK293T cells (ATCC, CRL -1573) were chosen because of their short amplification time (< 24 h) and high transfection efficiency (Tom et al, 2008). They express a mutant variant of the SV40 large T antigen and due to transfection with adenoviruses, they exhibit cytogenetic instability. The RPE-1 cell line was derived from a normal human retinal pigment epithelial cell line (RPE-340) and immortalised by the human telomerase reverse transcriptase (hTERT) subunit (Bodnar et al, 1998). RPE1 cells have a stable and normal karyotype and they were selected for our study because they represent a non-transformed alternative to cancer cell lines and have low endogenous levels of DNA damage and DPCs. Two different siRNAs for each gene were tested with different incubation times. Silencing efficiency was determined by qPCR.

SPRTN siRNAs were tested in RPE1 cells at a 5 nM concentration after 48h and 72h incubation. After 48h, siSPRTN-1 decreased *SPRTN* gene expression by 75% compared with WT non-transfected cells, whereas siSPRTN-2 was less efficient in comparison and decreased *SPRTN* expression by 53% (Figure 46 A and B). Both siRNAs targeting *SPRTN* were more efficient after 72h incubation. After 72h, siSPRTN-1 decreased *SPRTN* expression by 89% and siSPRTN-2 by 78% (Figure 46 C and D).

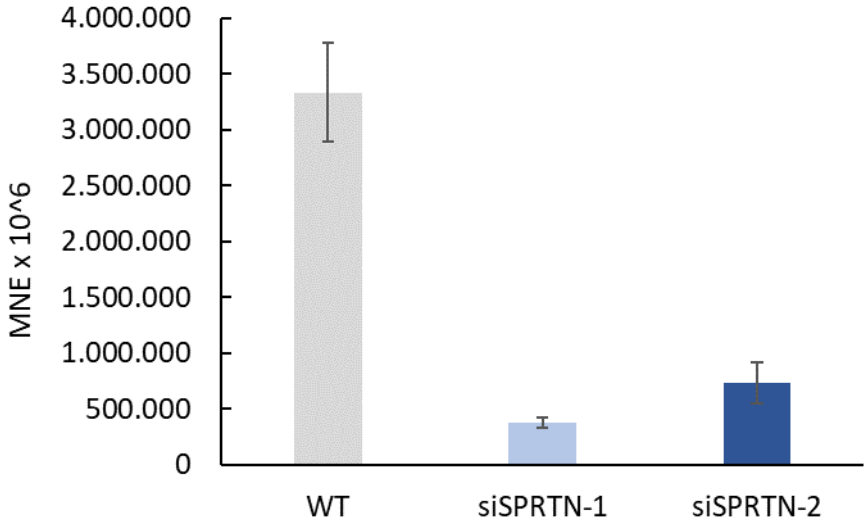
(A)


1.400.000
1.200.000
1.000.000
800.000
600.000
400.000
200.000
0
MNE x 10^6
WT
siSPRTN-1
siSPRTN-2

(B)


100
90
80
70
60
50
40
30
20
10
0
% of silencing
siSPRTN-1
siSPRTN-2

(C)



(D)

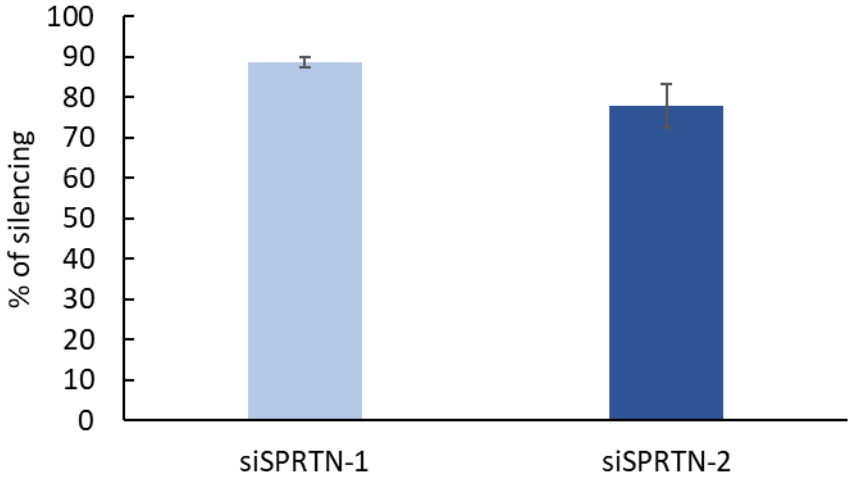
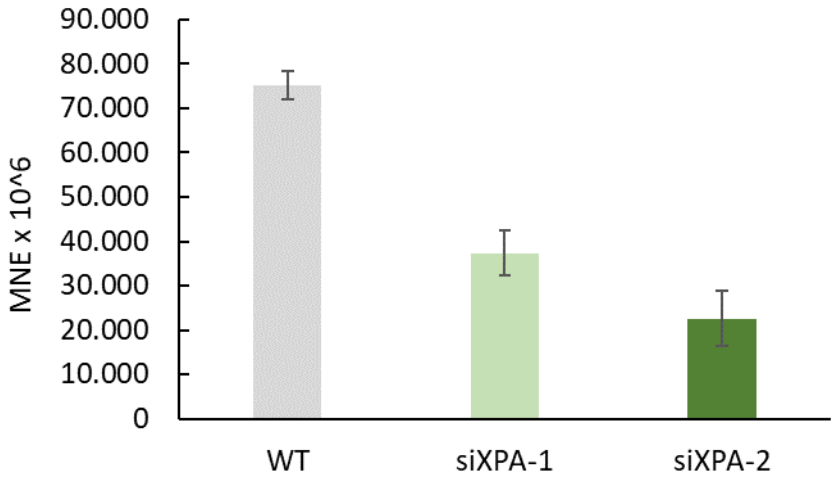


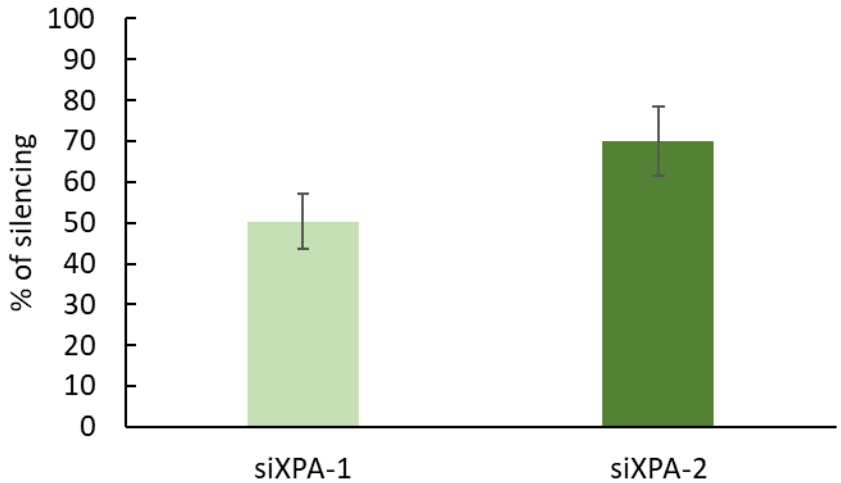
Figure 46. SPRTN silencing by transfection of small interfering RNAs (siRNAs) at 5 nM concentration in RPE1 cells. (A) SPRTN expression after 48h was determined by qPCR in non-transfected WT RPE1 cells and after transfection with 2 different small interfering RNAs (siRNAs). MNE stands for mean normalized expression normalized to the housekeeping gene ATP synthase mitochondrial F1 complex, O subunit (*HsATP50*). (B) Percentage (%) of SPRTN silencing compared to WT sample after 48h calculated from MNE values from (A). (C) SPRTN silencing by transfection of small interfering RNAs (siRNAs) in RPE1 cells after 72h. SPRTN expression was determined by qPCR in non-transfected WT RPE1 cells and after transfection of 2 different small interfering RNAs (siRNAs). MNE stands for mean normalized expression normalized to the housekeeping gene *HsATP50*. (D) Percentage (%) of SPRTN silencing compared to WT sample calculated from (C). Experiment was run in two independent biological replicates. Data represents mean ± standard errors of mean (SEM).

Two different siRNAs targeting *XPA* were tested in RPE1 cells at a 5 nM concentration after 48h and 72h incubation. After 48h, siXPA-1 decreased *XPA* expression by 50% compared with WT cells, whereas siXPA-2 decreased *XPA* expression by 70% (Figure 47 A and B). siXPA-1 is more efficient after 72h incubation, with a decrease in expression by 55%, while siXPA-2 was less efficient with a decrease in expression by 51% (Figure 47 C and D).

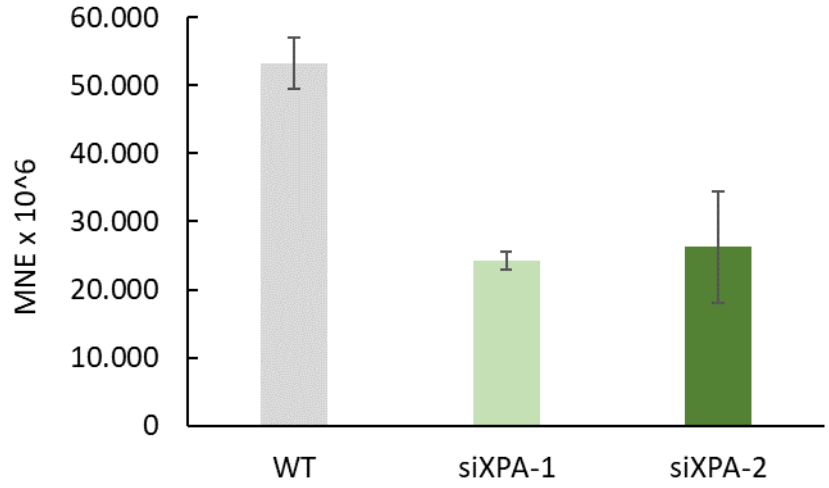
(A)



(B)



(C)



(D)

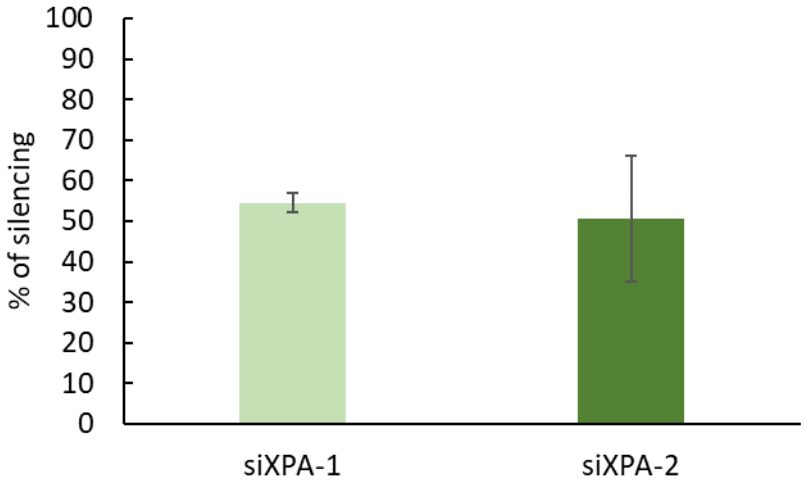
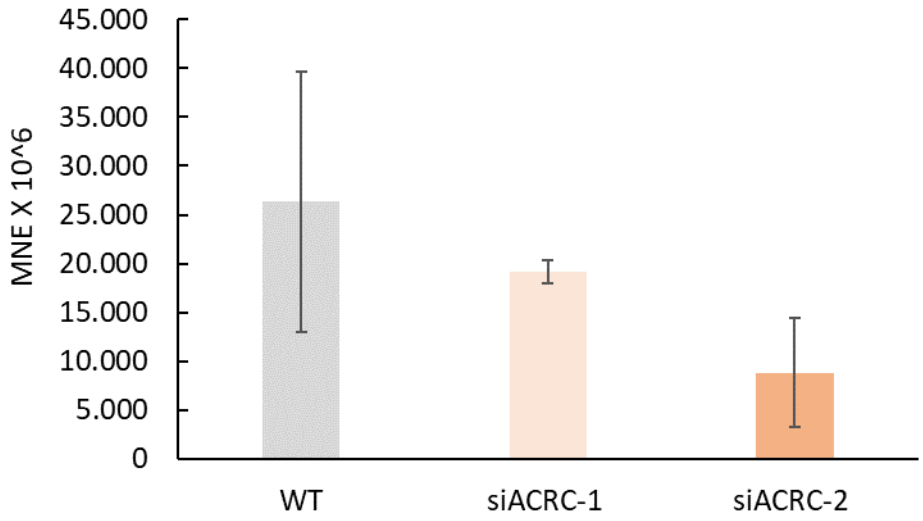


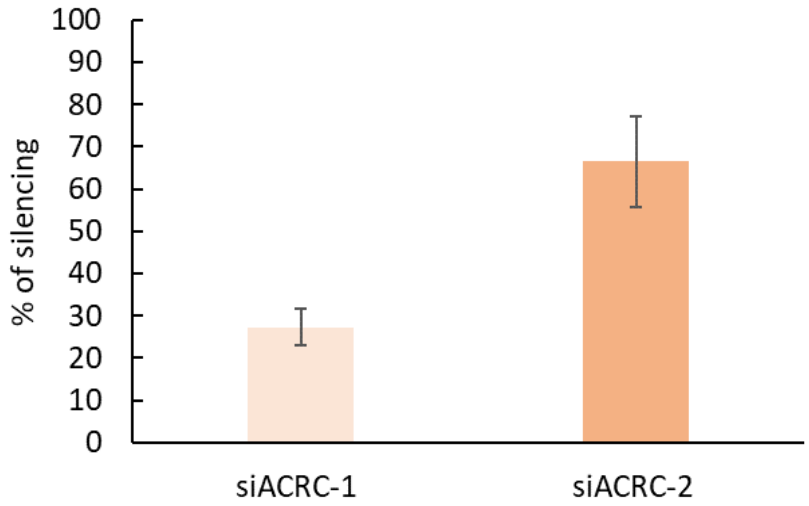
Figure 47. XPA silencing by transfection of small interfering RNAs (siRNAs) at 5 nM concentration in RPE1 cells. (A) XPA expression after 48h was determined by qPCR in non-transfected WT RPE1 cells and after transfection with 2 different small interfering RNAs (siRNAs). MNE stands for mean normalized expression normalized to the housekeeping gene *HsATP50*. (B) Percentage (%) of XPA silencing compared to WT sample after 48h calculated from MNE values from (A). (C) XPA silencing by transfection of small interfering RNAs (siRNAs) in RPE1 cells after 72h. XPA expression was determined by qPCR in non-transfected WT RPE1 cells and after transfection of 2 different small interfering RNAs (siRNAs). MNE stands for mean normalized expression normalized to the housekeeping gene *HsATP50*. (D) Percentage (%) of XPA silencing compared to WT sample after 72h calculated from (C). Experiment was run in two independent biological replicates. Data represents mean ± standard errors of mean (SEM).

ACRC siRNAs were tested in RPE1 cells at a 10 nM concentration after 48h and 72h incubation. After 48h, siACRC-1 decreased ACRC gene expression by 27% compared with WT non-transfected cells, whereas siACRC-2 was more efficient in comparison, reducing ACRC expression by 66% (Figure 48 A and B). siACRC-1 was more efficient after 72h incubation with a 64% reduction in expression, while siACRC-2 was less efficient, reducing ACRC expression by 38% (Figure 48 C and D).

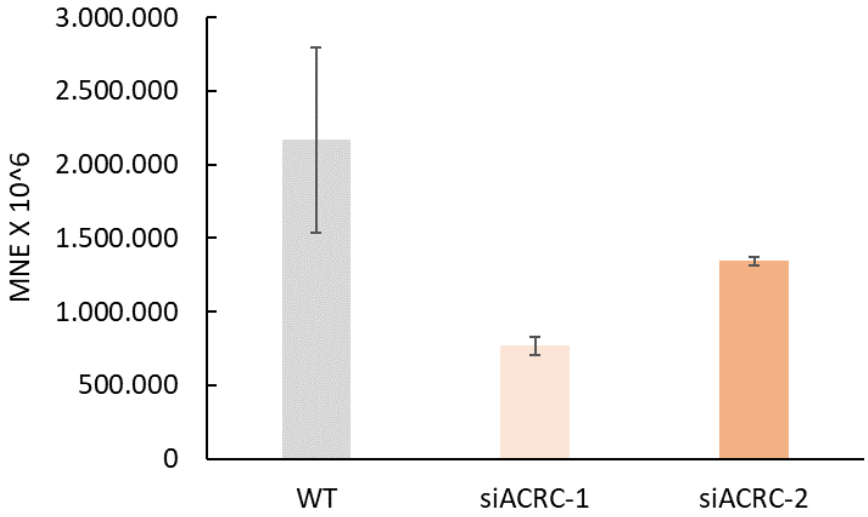
(A)



(B)



(C)



(D)

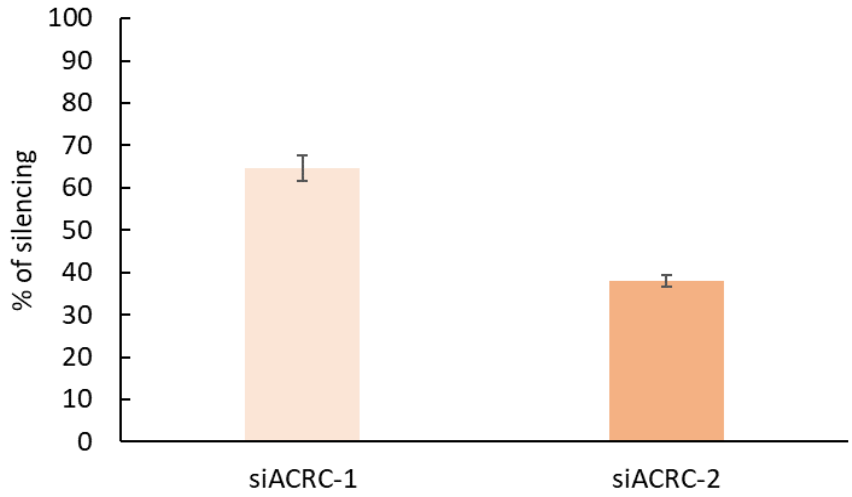
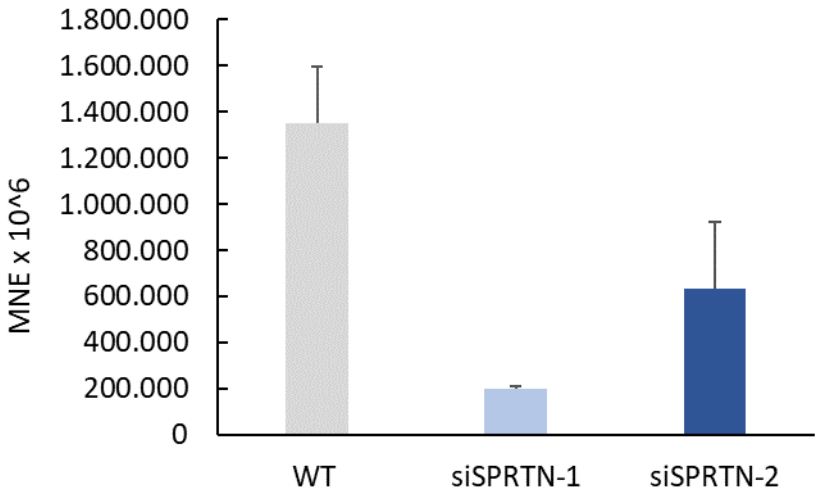


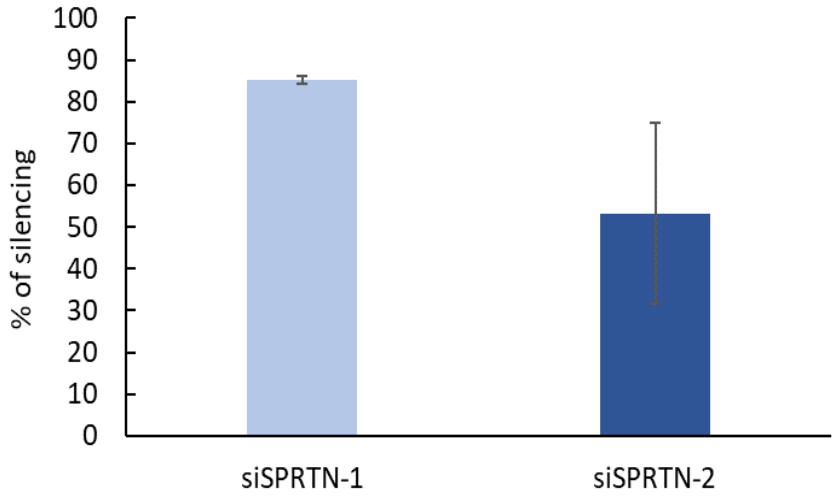
Figure 48. ACRC silencing by transfection of small interfering RNAs (siRNAs) at 10 nM concentration in RPE1 cells. (A) ACRC expression after 48h was determined by qPCR in non-transfected WT RPE1 cells and after transfection of 2 different small interfering RNAs (siRNAs). MNE stands for mean normalized expression normalized to the housekeeping gene *HsATP50*. (B) Percentage (%) of ACRC silencing compared to WT sample after 48h calculated from MNE values from (A). (C) ACRC silencing by transfection of small interfering RNAs (siRNAs) in RPE1 cells after 72h. ACRC expression was determined by qPCR in non-transfected WT RPE1 cells and after transfection of 2 different small interfering RNAs (siRNAs). MNE stands for mean normalized expression normalized to the housekeeping gene *HsATP50*. (D) Percentage (%) of ACRC silencing compared to WT sample calculated from (C). Experiment was run in two independent biological replicates. Data represents mean ± standard errors of mean (SEM).

SPRTN siRNAs were also tested in HEK293T cells at a 5 nM concentration after 48h and 72h incubation. After 48h, siSPRTN-1 decreased *SPRTN* expression by 85% compared to WT cells, while siSPRTN-2 decreased *SPRTN* expression by 53% (Figure 49 A and B). After 72h incubation, siSPRTN-1 was equally effective as at 48h incubation and reduced *SPRTN* expression by 85%, while siSPRTN-2 was more effective than after 48h, reducing expression by 76% (Figure 49 C and D).

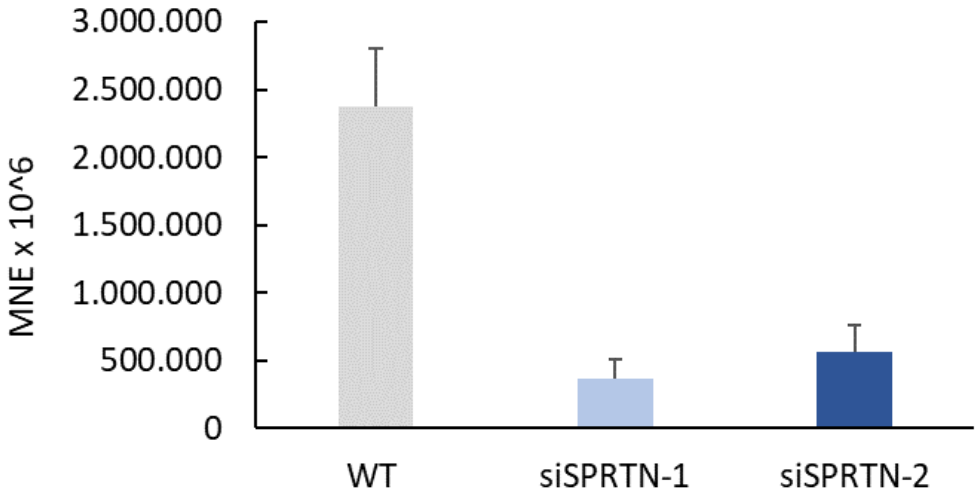
(A)



(B)



(C)



(D)

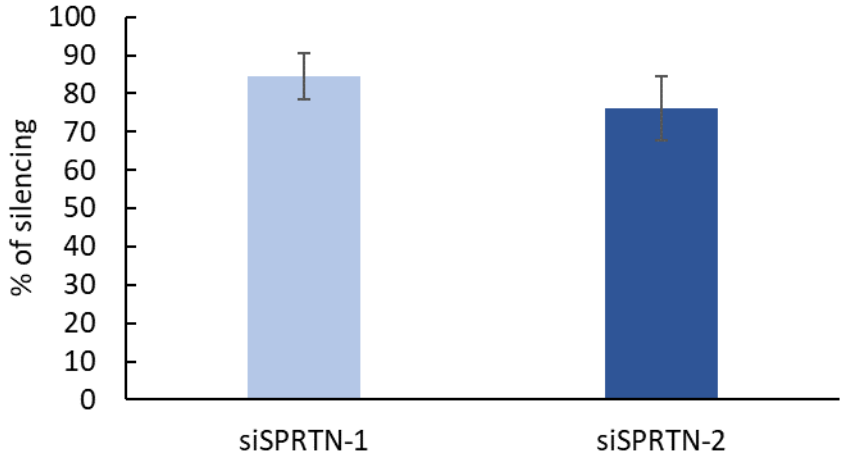
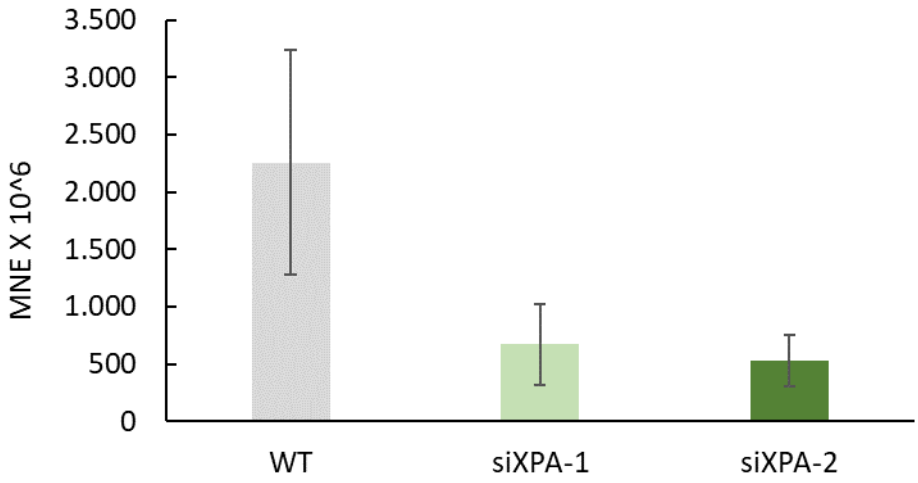


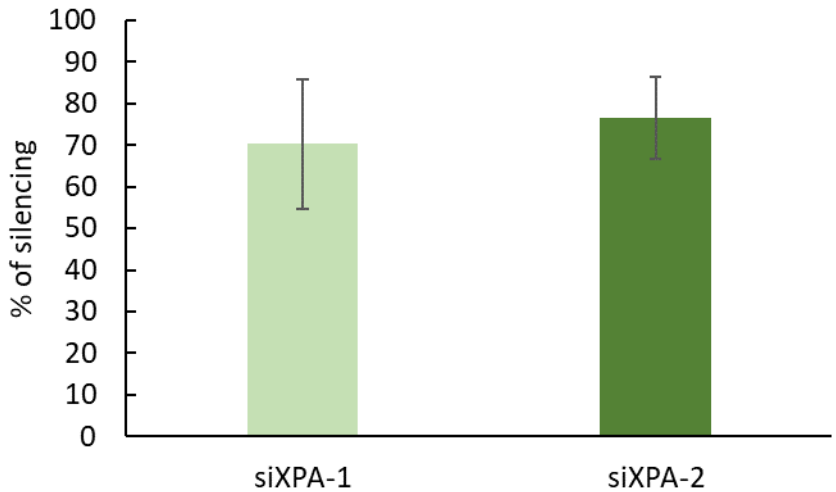
Figure 49. SPRTN silencing by transfection of small interfering RNAs (siRNAs) at 5 nM concentration in HEK293T cells. (A) SPRTN expression after 48h was determined by qPCR in non-transfected WT HEK293T cells and after transfection with 2 different small interfering RNAs (siRNAs). MNE stands for mean normalized expression normalized to the housekeeping gene *HsATP50*. (B) Percentage (%) of SPRTN silencing compared to WT sample after 48h calculated from MNE values from (A). (C) SPRTN silencing by transfection of small interfering RNAs (siRNAs) in HEK293T cells after 72h. SPRTN expression was determined by qPCR in non-transfected WT HEK293T cells and after transfection of 2 different small interfering RNAs (siRNAs). MNE stands for mean normalized expression normalized to the housekeeping gene H*sATP50*. (D) Percentage (%) of SPRTN silencing compared to WT sample after 72h calculated from (C). Experiment was run in two independent biological replicates. Data represents mean ± standard errors of mean (SEM).

XPA silencing with two different siRNAs was also tested in HEK293T cells at a 5 nM concentration after 48h and 72h incubation. After 48h, siXPA-1 decreased *XPA* expression by 70% compared to WT cells, whereas siXPA-2 decreased expression by 76% (Figure 50 A and B). After 72h, both siRNAs were less efficient: siXPA-1 reduced expression by 50% and siXPA-2 by 54% (Figure 50 C and D).

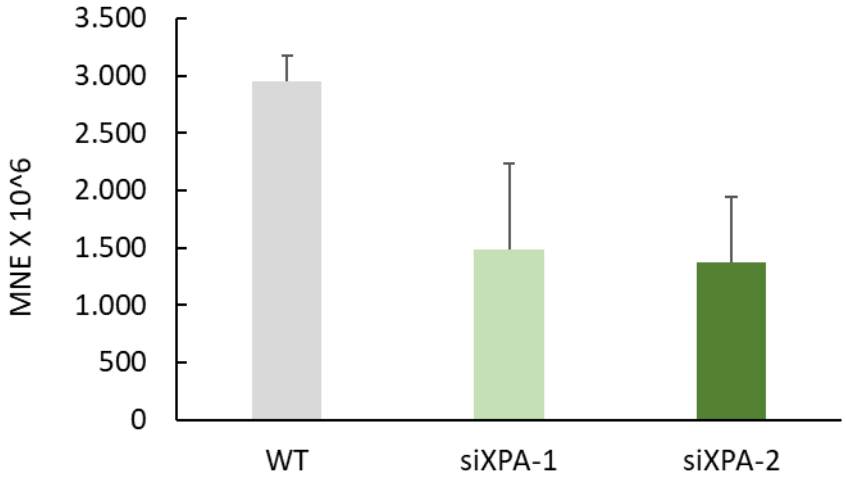
(A)



(B)



(C)



(D)

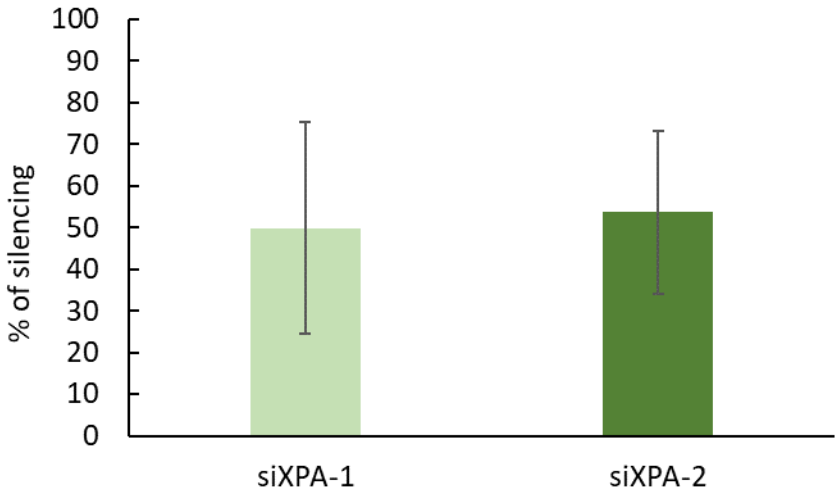
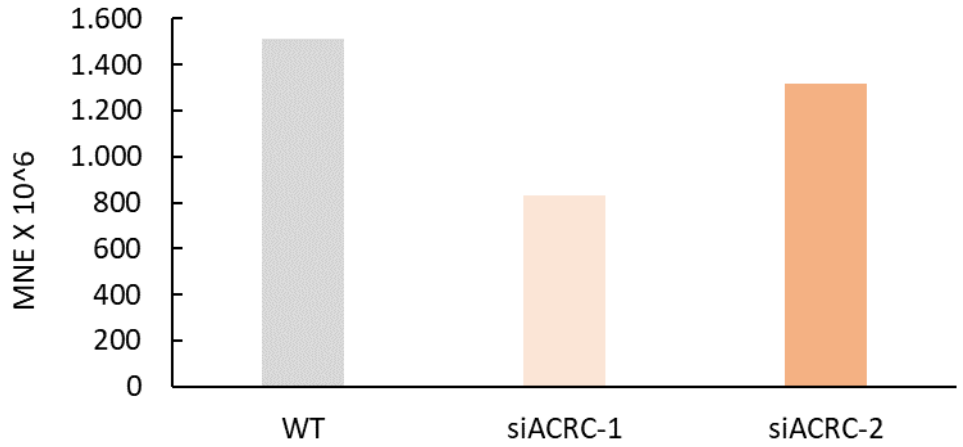


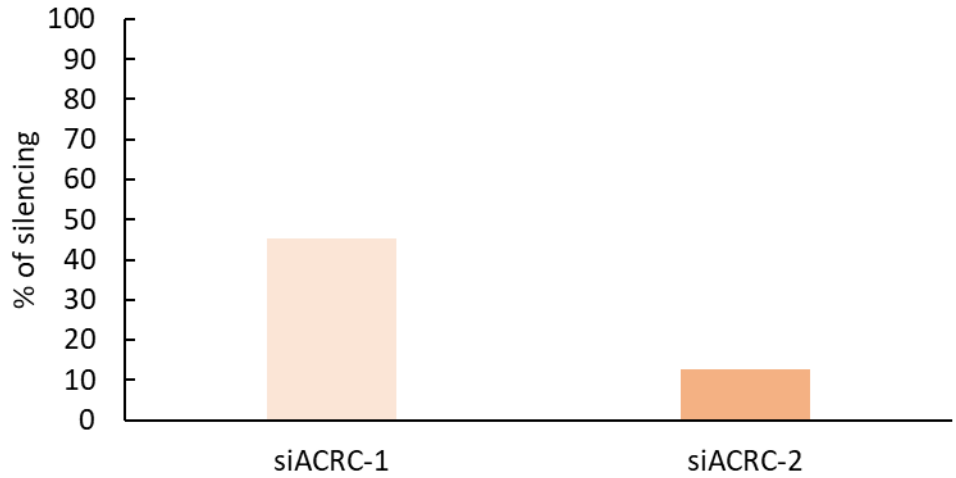
Figure 50. XPA silencing by transfection of small interfering RNAs (siRNAs) at 5 nM concentration in HEK293T cells. (A) XPA expression after 48h was determined by qPCR in non-transfected WT HEK293T cells and after transfection of 2 different small interfering RNAs (siRNAs). MNE stands for mean normalized expression normalized to the housekeeping gene *HsATP50*. (B) Percentage (%) of XPA silencing compared to WT sample after 48h calculated from MNE values from (A). (C) XPA silencing by transfection of small interfering RNAs (siRNAs) in HEK293T cells after 72h. XPA expression was determined by qPCR in non-transfected WT HEK293T cells and after transfection of 2 different small interfering RNAs (siRNAs). MNE stands for mean normalized expression normalized to the housekeeping gene *HsATP50*. (D) Percentage (%) of XPA silencing compared to WT sample after 72h calculated from (C). Experiment was run in two independent biological replicates. Data represents mean ± standard errors of mean (SEM).

ACRC siRNAs were also tested in HEK293T cells at a 10 nM concentration after 48h and 72h incubation. After 48h, siACRC-1 decreased *ACRC* expression by 45% compared with WT cells, while siACRC-2 decreased *ACRC* expression by only 12% (Figure 51 A and B). After 72h, both siRNAs were more efficient. siACRC-1 decreased *ACRC* expression by 50%, while siACRC-2 decreased by 64% (Figure 51 C and D).

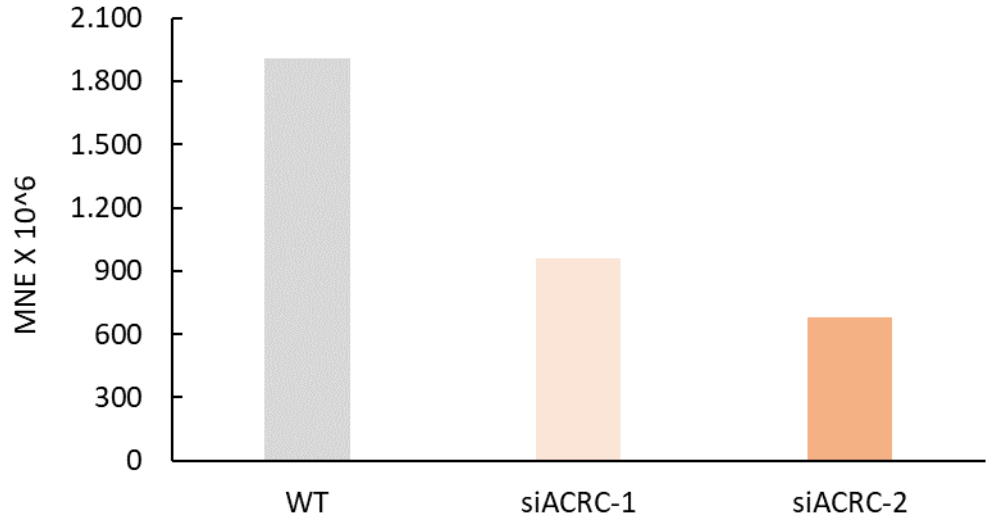
(A)



(B)



(C)



(D)

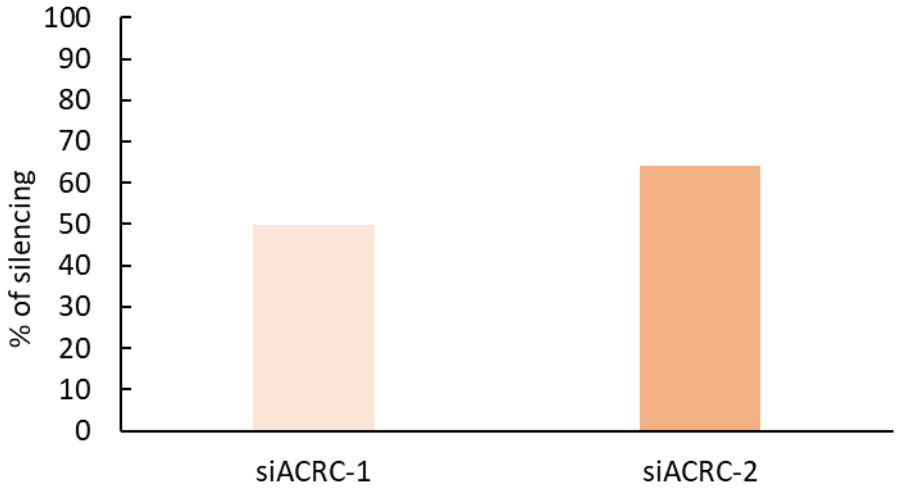
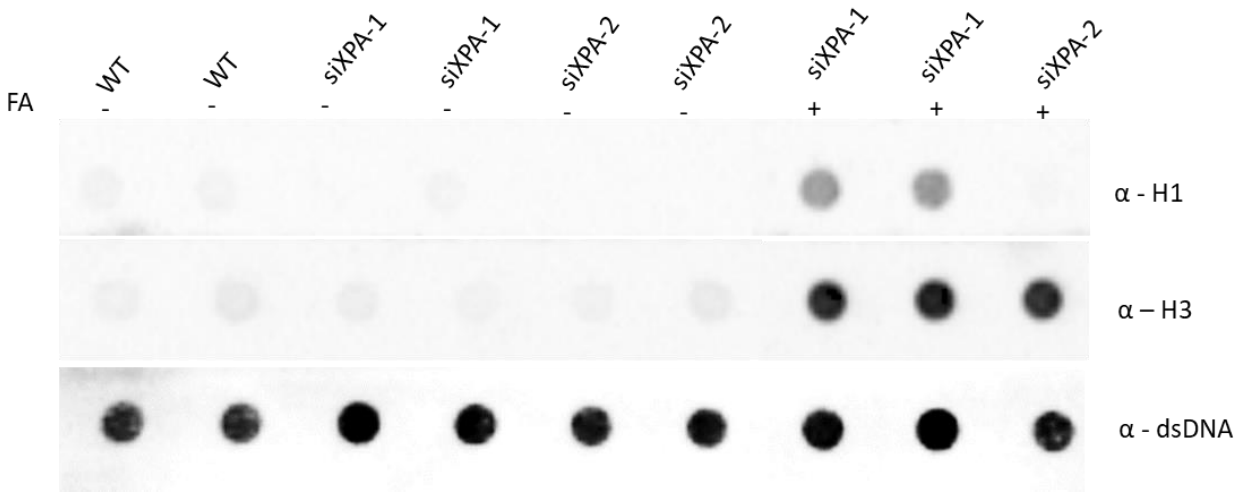


Figure 51. ACRC silencing by transfection of small interfering RNAs (siRNAs) at 10 nM concentration in HEK293T cells. ACRC expression after 48h was determined by qPCR in non-transfected WT HEK293T cells and after transfection of 2 different small interfering RNAs (siRNAs). MNE stands for mean normalized expression normalized to the housekeeping gene *HsATP50*. (B) Percentage (%) of ACRC silencing compared to WT sample calculated from MNE values from (A). (C) ACRC silencing by transfection of small interfering RNAs (siRNAs) in HEK293T cells after 72h. ACRC expression was determined by qPCR in non-transfected WT HEK293T cells and after transfection of 2 different small interfering RNAs (siRNAs). MNE stands for mean normalized expression normalized to the housekeeping gene *HsATP50*. (D) Percentage (%) of ACRC silencing compared to WT sample calculated from (C). Experiment was run in biological monoplicates.

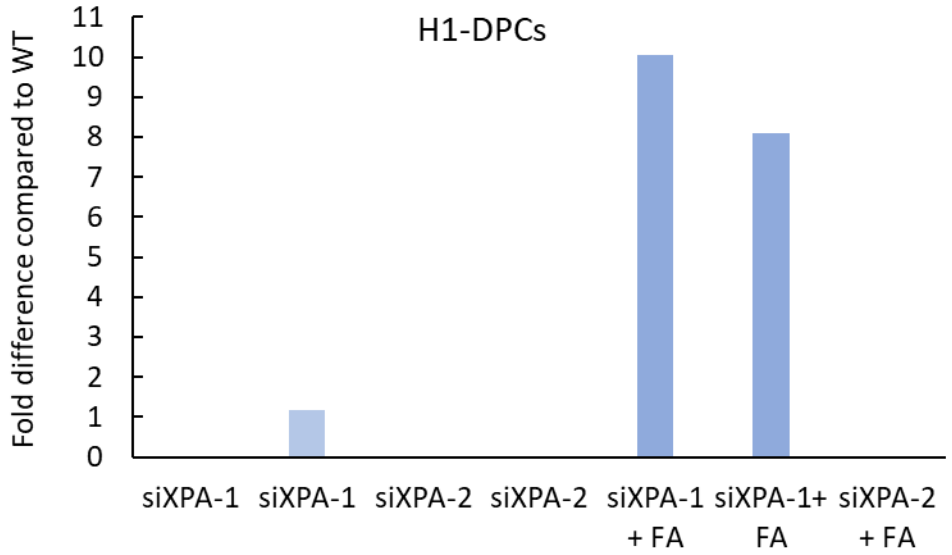
3.14. Histone 1 (H1) and Histone 3 (H3) DPC levels in XPA deficient RPE1 cells

To detect specific DPCs in XPA deficient RPE1 cells, XPA was silenced by transfection of XPA siRNA-1 and siRNA-2 using previously optimized conditions (Figure 47). Both siRNAs were used at 5 nM concentrations and cells were collected after 72h. Cells were treated with formaldehyde, a known histone DPC inducer (10 mM, 20 min) and DPCs were isolated by the RADAR assay, followed by slot blot analysis and immunostaining with protein specific antibodies to detect specifically histone H1 and H3 DPCs. Previously published studies have shown that the bacterial NER complex is able to repair DPCs *in vitro* up to 16 kDa and up to 11 kDa *in vivo*. The upper size limit for DPCs repair by mammalian NER complex *in vitro* is around 8 kDa, which is significantly less than that for bacterial NER. Histones are small proteins, with the molecular weight of individual histones ranging from 11 to 22 kDa. Therefore, it was proposed that histone DPCs could be repaired by the NER pathway (Nakano et al, 2009; Stingele et al, 2017; Dinant et al, 2012). Histones H1 and H3 were detected with a specific antibody, and as a loading control, the amount of double-stranded DNA (dsDNA) was visualized with a specific antibody (Figure 52 A). The amount of dsDNA was similar in each sample, meaning that samples are comparable to each other and that same amount of every sample was loaded on the membrane for histone H1 and H3 detection. XPA silencing in RPE1 cells did not cause an increase in the amount of H1-DPCs and H3-DPCs under physiological conditions (non-treated cells). However, after treatment with a potent DPC inducer, formaldehyde, H1 and H3 DPC levels significantly increase if XPA is silenced (Figure 52 A-C). Specifically, in cells treated with siXPA-1 and formaldehyde there is 8-10 times more H1-DPCs compared to non-treated WT cells (Figure 52 B), while cells treated with siXPA-1 or siXPA-2 and formaldehyde accumulate 14 - 16 times more H3-DPCs than WT cells (Figure 52 C).

(A)



(B)



(C)

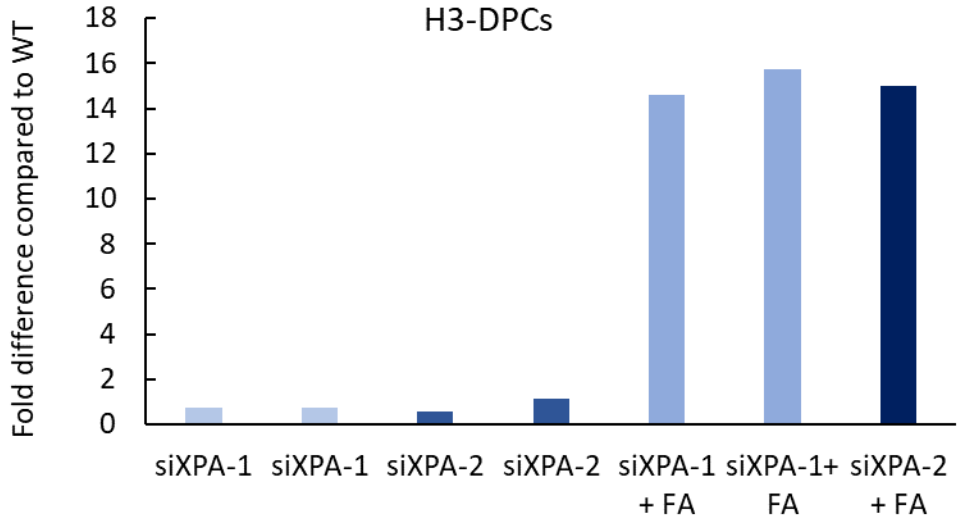


Figure 52. (A) Detection histone H3 -DPCs, and H1- DPCs by slot analysis following DPC isolation by RADAR assay in RPE1 cells after XPA silencing by transfection of small interfering RNAs (siRNAs). siXPA 1 and 2 were used in 5 nM concentration and incubated for 72h. H3, H1 and dsDNA were detected with specific antibodies. (B) Quantification of H1-DPCs and (C) H3-DPCs from (A) using ImageJ. Amount of DPCs is expressed as a fold difference compared to WT sample.

4. Discussion

4. 1. Characterization of zebrafish XPA protein

Conserved XPA orthologs are found in vertebrates and invertebrates, emphasizing the importance of the XPA protein and NER pathway in DNA damage repair (Figure 12). Our phylogenetic analysis of XPA orthologs which resulted in the first comprehensive phylogenetic tree ranging from bacteria, algae and yeast to metazoans, showed that bacteria do not have XPA orthologs. This finding strengthens previously reported observations that bacteria do not share direct gene orthology of NER counterparts with mammals (Truglio et al., 2006). However, bacteria do possess UvrABC genes which are a part of the DNA repair pathway analogous to eukaryotic NER (Truglio *et al.*, 2006) thus again emphasizing the crucial role of this repair pathway. In addition to bacteria and mammals where NER was mainly studied, it has been shown that complete NER complex is present in *Caenorhabditis elegans* and *Drosophila melanogaster* (Lans and Vermeulen, 2011).

Xpa gene in eukaryotes shows no homology to bacterial damage recognition protein UvrA, a possible functional analog in bacterial NER pathway (Rademakers *et al.*, 2003). Homologs of the bacterial UvrABC genes and eukaryotic genes such as XPB, XPD and XPF have also been found in a large number of archaea (White and Allers, 2018). However, the XPA gene was not found in archaea nor in unicellular eukaryote *Plasmodium,* which goes in line with our phylogenetic analysis. In addition to published results, we have shown that distant orthologs of XPA are found in yeast and green algae (Figure 12). Intrestingly, we and others have shown that *xpa* genes are present with only one representative gene (one-to-one orthology) in all analyzed groups, including zebrafish, which indicates high degree of conservation of XPA function (Figure 12). Considering that teleost fish underwent whole genome duplication (WGD) event through the course of evolution (Taylor *et al.*, 2003), zebrafish often has duplicated genes in comparison to other vertebrates. However, we have confirmed that zebrafish has one *xpa* ortholog and due to one-to-one orthology of the *Xpa* between humans and zebrafish, we suggest that zebrafish is a convenient model to study the role of XPA in DPC repair. Moreover, conserved synteny analysis performed in this study showed that gene environment of the zebrafish Xpa protein is relatively conserved (Figure 13), while other studies have shown that XPA protein is conserved across species at sequence and structure level (Barve *et al.*, 2021).

4.2. Creation and optimization of cellular and animal models to study the role of XPA in DPC repair

Historically, NER was mostly studied in mammals – from rodents to humans. Several NER-deficient cell lines as well as *Xeroderma pigmentosum* patient-derived cell lines have been used to investigate the function of human NER pathway (de Boer and Hoeijmakers, 2000). *In vitro* studies in human cell lines have shown that small DPCs (8-16 kDa) can be repaired by NER, while the involvement of NER in DPC repair at the organismal level is still unknown. That is why we have set out to investigate the role of NER pathway in DPC repair using zebrafish animal model. Among the various proteins involved in the NER pathway, we chose to mutate the *xpa* gene because it is specifically involved in the NER pathway and has no other cellular

functions (Pulzová, Ward and Chovanec, 2020). Also, based on the phenotype of XPA-deficient mice, we hypothesized that mutation of *xpa* would not be embryonic lethal in zebrafish.

Human XPA is a relatively small protein of 273 amino acids (31 kDa), while in zebrafish it is a protein of 549 amino acids (62 kDa) that has no enzymatic activity but interacts with many other NER proteins and serves as a scaffold. The XPA protein contains three domains: a central globular domain with a zinc-finger motif and disordered N- and C-terminal domains. The N-terminus has a nuclear localization signal (NLS). XPA binds DNA through the DNA-binding domain, which includes both the central globular core domain and part of the C-terminal domain (Sugitani et al., 2016). The flexible C-terminus is important for the interaction of XPA with other NER components. We have selected a sgRNA that targets exon 6 of the zebrafish *xpa* gene because introducing premature stop codon into the DNA-binding domain would impair XPA scaffold function and thus impair NER.

In order to introduce frameshift mutation in exon 6 of zebrafish *xpa* we have used CRISPR/Cas9 system (Figure 16). sgRNA was designed and injected into the yolk of one-cell stage embryos together with Cas9 protein. Cas9 protein leads to DNA double-strand break, which is repaired mainly by non-homologous end joining (NHEJ). NHEJ is an error-prone pathway that often leads to insertions or deletions at the break sites. During DSB repair, insertion or deletion of different nucleotide sequences may occur, and this event is specific to break site within each individual cell. Therefore, different mutations arise in each cell leading to the mosaicism of embryos during the development and eventually to mosaic adults (Mehravar et al., 2019). Although the sgRNA/Cas9 complex is injected at the single cell stage, it takes time for the complex to enter the nucleus, initiate the break and for NHEJ machinery to reach and repair the break site. In the meantime, the zygote starts dividing and the complex does not reach all the cells during the early development. Also considering that NHEJ does not always introduce indels and repairs the DNA without errors, some cells will not have the change in the sequence of target gene (wild-type sequence). High-resolution melting (HRM) analysis has been used to genotype the germline of F0 adults, because it easily and rapidly distinguishes individuals carrying a mutation from WTs, that differ in T_m values. However, due to the limitations of the method, small deletions (1-2 bp) do not cause a shift in T_m value and are therefore not visible. Therefore, embryos that were HRM positive were sequenced to determine the type of mutation and to ensure that the mutation leads to protein disruption (the number of deleted bases must not be divisible by 3).

Morpholino antisense oligomers are used for efficient gene silencing in a number of model organisms, including zebrafish. They are chemically synthesized DNA analogs that are injected into embryos at the one-cell stage, where they bind complementary target mRNAs and prevent their translation or alter splicing (Nan and Zhang, 2018). Morpholinos are commonly used in zebrafish as knock-down tool because of their ease of administration and high efficacy during zebrafish embryonic and larval development. Zebrafish embryogenesis is complete at 2 dpf (days post fertilization) when the majority of organ systems develop. Morpholinos are stable after microinjection into embryos up to 3 dpf, but the dilution by cell growth and division reduces their activity and efficacy. To silence *xpa* and *sprtn* in zebrafish embryos, splice-blocking morpholinos were designed to inhibit mRNA splicing, resulting in a frameshift mutation and degradation of Xpa and Sprtn protein (Figures 21 and 24). Both translation-

blocking and splice-blocking types of morpholinos inhibit translation of zygotic transcripts, whereas splice-blocking morpholinos are ineffective on maternally inherited transcripts which are already mature (Moulton, 2017). While there are no published studies of using morpholinos targeting zebrafish *xpa* gene, morpholino-mediated *sprtn* silencing in zebrafish was performed by Lessel et al (2014). Embryos injected with *xpa* morpholino showed no phenotypic changes, and PCR reaction showed morpholino to be 100% efficient (Figure 22). Lessel et al reported that morpholino-mediated depletion of *sprtn* in zebrafish embryos caused accumulation of DNA damage which led to severe phenotypic defects, and early mortality at 10 hpf (Lessel et al., 2014). However, in our hands, morfolino used by Lessel et al caused complete mortality of the embryos within 24 hpf and therefore we have designed and optimized two other morpholino probes (Cecile Otten, unpublished results). The splice morpholino for *sprtn* was of high efficiency (79%) and caused mild phenotypic changes including shorter body and mild tail curvature (Figure 26), similar to the second morpholino tested (translation blocking) (Cecile Otten, personal communication). Therefore, our current hypotheses is that published morpholino (Lessel *et al.*, 2014) most probably had unspecific effects in addition to silencing *sprtn*, which remains to be proven in future experiments.

In addition to achieving efficient silencing of *xpa* and *sprtn* in zebrafish embryos (Cecile Otten, unpublished results), we have successfully optimized *XPA*, *ACRC* and *SPRTN* gene silencing in stable human cell lines RPE1 (Retinal Pigment Ephitilial-1) and HEK293T (Human Embryonic Kidney 293) cells. These results are first of a kind for *XPA* and *ACRC* genes, while *SPRTN* silencing was previously published (Vaz *et al.*, 2016; Fielden *et al.*, 2020; Ruggiano *et al.*, 2021a). The RPE-1 cell line was chosen because it is a non-transformed alternative to cancer cell lines and has low endogenous levels of DNA damage and DPCs. Specifically, the RPE-1 cell line is derived from a normal human retinal pigment epithelial cells (RPE-340) and immortalised by the human telomerase reverse transcriptase (hTERT) subunit (Bodnar et al., 1998). RPE1 cells have a stable and normal karyotype with a modal chromosome number of 46 and are commonly used to study physiological processes in human cell cultures (Hindul et al., 2022). HEK293T cells (ATCC, CRL -1573) were chosen because of their short amplification time (< 24 h), high transfection efficiency (Tom, Bisson and Durocher, 2008) and very high protein yields. Despite the historically long-term productive exploitation, the origin, phenotype, karyotype, and tumorigenicity of HEK293T cells are still debated. Cytogenetic analysis showed that the HEK293T cell line is pseudotriploid with heterogeneous and unstable karyotype. The mean chromosome number and chromosome aberrations differ between cells and its derivatives as well as between cells from the different cell banks or laboratories. HEK293T cells are tumorigenic and have increased number of DSB in normal physiological conditions (Stepanenko and Dmitrenko, 2015). Therefore, silencing efficiency differs among different cell lines used.

4.3. The role of XPA in DPC repair

In this thesis, the RADAR method was adapted and optimized for the isolation of DPCs from zebrafish embryos based on previously published protocols (Kiianitsa and Maizels, 2013; Vaz *et al.*, 2016; Bhargava *et al.*, 2020). Bhargava et al (2020) first performed RADAR assay for DPC isolation from zebrafish embryos. We have further adapted the method to reduce variability between experiments (Ivan Anticevic, unpublished results). To investigate whether NER

pathway and SPRTN protease are involved in the same repair pathway, we simultaneously injected zebrafish embryos with *sprtn* and *xpa* morpholino, after which DPCs were isolated by RADAR assay and visualized by silver staining. Our results showed that zebrafish embryos deficient in Xpa protein had 1.8-fold higher accumulation of total DPCs compared with WT embryos (Figure 27 B). Surprisingly, attenuation of *sprtn* did not cause an increase in total DPC levels (1.1-fold compared with WT). Considering that there are no published results on measuring DPC levels in zebrafish embryos after *sprtn* silencing, we have no point of reference from the scientific community. However, considering that we have shown that Acrc protease is highly expressed and crucial in the early embryonic development, it is possible that Acrc is compensating for Sprtn after silencing. A second possibility is that *sprtn* morpholino was not efficient enough in this experiment or there are maternally-deposited Sprtn mRNA and protein that function in DPC repair. In human cell lines, previous studies have shown that SPRTN depletion in HeLa, HEK293T and T24 cells results in a 2- to 5-fold increase in the total DPCs (Vaz et al, 2016). However, stable cell lines and embryos are completely different systems. Cell lines are heterogeneous due to the mixture of cells at different cell cycle phases, while the early zebrafish embryo undergoes many rounds of rapid cell divisions between 3.7 and 6 hpf to the 1000-cell stage (Laue *et al.*, 2019).

However, embryos deficient in both Xpa and Sprtn experienced a strong accumulation of total DPCs, which increased 4.1-fold compared to WT embryos, suggesting that NER pathway and Sprtn act independently of each other in the repair of general cellular DPCs (Figure 27 B). However, this hypothesis will have to be confirmed by additional experiments. It has been proposed that DPCs are first degraded by replication-coupled DPC proteases to shorter peptides that are then removed by NER, but this model hasn't been studied *in vivo* ((Vaz, Popovic and Ramadan, 2017; Pachva et al., 2020). Alternatively, the NER could be active on small DPCs throughout the cell cycle (Nakano et al., 2009). We and others envisaged that covalently bound protein is proteolytically degraded to a small peptide by the SPRTN protease or by the proteasome, while the damaged DNA component is repaired by NER (Pachva et al., 2020). During replication arrest, DPCs could be digested by SPRTN protease, after which the peptide residue could be bypassed by TLS DNA synthesis. Because TLS often cause mutations, proteolysis could be followed by NER, which represents error-free repair. In replication-independent repair, 26S proteasome-mediated degradation followed by removal of the peptide residue could occur with the help of the NER pathway. However, these models of the interplay between different repair pathways are hypothetical and have not been studied yet. Given the additive effect of DPC accumulation in the absence of both NER pathway and Sprtn protease, our results suggest that NER pathway *in vivo* repairs DPCs independently of SPRTN.

A general DPC inducer, formaldehyde increased DPC levels by 75% (1.8 fold) in 2 days old WT embryos (Figure 27 B). This is the first study to measure total DPC levels in fomaldehyde exposed zebrafish embryos. For comparison, in Hela cell lines FA exposure causes a huge accumulation of general or specific DPCs (Vaz et al., 2016). Embryos deficient in either Xpa or Sprtn and treated with FA showed a slight increase in total DPCs (1.2-fold compared with WT). However, embryos deficient in both proteins and treated with FA had 2.2-fold more total DPCs than WT embryos (Figure 27 B) and 1.3-fold more than WT+FA. A potent DPC-inducing agent formaldehyde also induces DNA interstrand crosslinks (ICLs) (Ide *et al.*, 2018), and due

to subsequent SSB and DSB induction activates DNA damage signaling pathways other than DPC repair. It seems that during embryonic development under physiological (non-treated) conditions, Sprtn and Xpa are very important for DPC removal, while upon acute exposure to formaldehyde their deficiency only slightly affects the DPC repair. When comparing the level of high, medium and low molecular weight DPCs, we have observed similar patterns with the HMW, MMW and LMW accumulation as with the total DPC levels (Fig 27 C-E). It is of note that Xpa deficiency caused similarly mild or slightly stronger accumulation of LMW (2.8- fold) and MMW (1.8-fold) in comparison to Sprtn-deficient embryos (1.1. and 0.9-fold) (Figure 27 D and E). In addition, embryos deficient in both Sprtn and Xpa strongly accumulated LMW DPCs (9.5-fold), followed by MMW and HMW DPCs (Figure 17 C-E). More experiments are needed to conclude whether our hypothesis that NER is involved in the removal of SMW *in vivo* is correct. Also, specific DPCs, mainly histones DPCs should be quantified in further experiments on zebrafish embryos.

In human stable cell line, RPE1 cells we have observed a very strong increase in histones DPCs, after *XPA* silencing and FA treatment, namely H1 and H3 DPCs (Figure 52). DNA-binding proteins (histones, transcription factors, DNA metabolizing enzymes) are often covalently crosslinked to DNA. The basic chromatin structure makes histones prime targets of DPC-inducing agents, leading to the formation of histone DPCs (Solomon and Varshavsky, 1985). Formaldehyde is a general DPC inducer that covalently traps proteins of different sizes, and it has been shown that both exogenously and endogenously present aldehydes can induce the formation of histone DPCs in cells (Kuykendall and Bogdanffy, 1992). Interestingly, in physiological (non-treated) conditions, XPA deficiency did not cause an increase in the amount of H1-DPCs and H3-DPCs (Figure 52), but when the system was challenged by high dose of FA (10 mM, 20 min), histone DPCs levels significantly increased indicating that NER pathway is crucial for H1 and H3 DPC removal in cells. Silencing efficiencies of both siXPA used in this experiment were analyzed by qPCR which showed that siXPA-1 silenced XPA expression by 65% and siXPA-2 by 75%. Our results are in line with the data from bacterial systems and human stable cell lines where NER was found to remove small DPCs of up to 14 kDa and 11 kDa, respectively (Minko, Zou and Lloyd, 2001). More specifically, the role of mammalian NER in DPC repair was previously investigated *in vitro* using cell-free mammalian extracts and defined DPC substrates. The mammalian NER complex was found to perform incisions for crosslinked proteins of 4-12 amino acids (0.57 - 1.5 kDa) but was inefficient in repairing DPCs containing T4 endonuclease (16 kDa) and DNMT1-DPCs (37 kDa). The incision efficiency was also insignificant for the 11 kDa protein. These observations suggest that the upper size limit for DPCs repair by mammalian NER *in vitro* is around 8 kDa.

It was initially proposed that histone DPCs could be repaired by the NER pathway without prior proteolysis (Sczepanski *et al.*, 2010). Although individual histones are very small (11-13 kDa), they form larger complexes that can sterically inhibit the binding of NER factors to DNA (Ide et al, 2011). In addition, the *Xeroderma pigmentosum* F (XPF) nuclease has been shown to require prior proteolytic processing of the crosslinked protein before it can be removed (Stingele, Bellelli and Boulton, 2017). Chromatin remodeling at a DPC site is also essential to enable binding of NER machinery, indicating that degradation of the bulky protein is needed prior to repair of the DNA component by the NER pathway (Nakano *et al.*, 2007). Although histone DPCs are abundant in cells, it is believed that there is no specialized DNA repair pathway for these complexes and cells probably use different DNA repair and protein

degradation mechanisms to repair them (Pachva *et al.*, 2020). Our results strongly indicate that NER complex is involved in repair of histone DPCs, however, the involvement of other factors remains to be further explored.

4.4. Characterization of zebrafish ACRC protein

ACRC orthologs are found in archea and eukarya, while no orthologs have been found in prokaryotes. More specifically, ACRC protein is conserved with one-to-one orthology in all invertebrate and vertebrate species (Figure 7) which would suggests an essential role of this protein for cellular functions (Carmell et al., 2016; Fielden et al., 2018). Our hypothesis that ACRC is a nuclear protease with a role in DPC repair was based on the similarity of its Sprt domain with SPRTN protease (Vaz, Popovic and Ramadan, 2017). Indeed, we have previously shown that ACRC protein family is evolutionarily close to the SPRT family of proteases, whereas it is more distant from the WLM family of proteases, which includes the yeast protease Wss1 (Figure 7).

The conserved synteny analysis performed in this study confirmed that zebrafish *acrc* gene is syntenic with human gene with relatively conserved gene environment (Figure 14). Zebrafish do not have sex chromosomes and sex determination has been linked to certain genes spreadout on different chromosomes, although this process is also strongly influenced by environmental factors (Pradhan and Olsson, 2016). Considering conserved synteny between human and zebrafish *ACRC* gene environment and that human *ACRC* is located on chromosome X, it is possible that gene cluster on zebrafish chromosome 14, where *acrc* is located, is involved in sex determination in zebrafish. So far, studies showed the presence of a sex associated regions on chromosomes 3, 4, 5, and 16, which contain genes linked to zebrafish sex differentiation, such as *cytochrome P450, family 21, subfamily A, polypeptide 2 (cyp21a2), hydroxysteroid (17-a) dehydrogenase-1 (hsd17b19)*, and *patched domain containing 3 (ptchd3)* (Anderson et al., 2012). Specifically, chromosome 4 was suggested to be the candidate zebrafish sex chromosome due to the suppressed recombination which is noticed in sex chromosomes of other species (Pradhan and Olsson, 2016). However, additional studies are needed to determine if identified gene cluster around *acrc* on zebrafish chromosome 14 is involved in sex determination in zebrafish.

We have found that zebrafish *acrc* is predominantly expressed in testes and ovaries (Figure 15 A). Although a tissue expression analysis has not so far been reported in a single vertebrate species, high expression of ACRC was reported in mouse testis (Carmell *et al.*, 2016) and in germ cells of *C. elegans*, *Drosophila*, and zebrafish (Carmell *et al.*, 2016; Bhargava *et al.*, 2020). Indeed, data from human expression atlas show that ACRC is highly expressed at both mRNA and protein level in human testis, while expression at the protein level is moderate in lung. The highest expression of the ACRC in humans is in testis, followed by brain, liver, intestine, kidney, and finally the ovaries which shows the lowest level of ACRC expression (https://www.proteinatlas.org/). In contrast to previous hypotheses that ACRC plays a critical role only in germ tissues (Bhargava *et al.*, 2020; Dokshin *et al.*, 2020) we found moderate and high expression at the mRNA level in most tissues, suggesting that ACRC also acts in low replicative tissues (such as kidney and intestine). In summary, we and others propose that ACRC is important for the maintenance of genome integrity, especially in cells whose genomes will be passed to the next generation. Considering that ACRC expression is enriched

during meiosis in single-celled eukaryotes (such as fungi *Schizosaccharomyces pombe* and green algae *Chlamydomonas)* and in germ cells of vertebrates, ACRC may protect cells from DNA damage occurring during meiosis, a process that involves extensive, programmed DNA damage (Carmell et al., 2016; Bhargava et al., 2020; Dokshin et al., 2020).

Considering that ACRC is evolutionary close to SPRTN and has a similar Sprt domain to SPRTN protease which is crucial for DPC repair we have compared the expression of these two genes in zebrafish adult tissues and during the embryonic development for the purpose of getting more insights into the role of ACRC. Similarly to *acrc*, *sprtn* is most highly expressed in zebrafish gonads (Figure 15 B). The only available data of SPRTN expression are from human expression atlas, which shows high expression at the transcriptional level in human testis, whereas expression at the protein level is high in brain, skin, and bone marrow, low in testis and colon, and undetectable in other tissues (https://www.proteinatlas.org/). While *acrc* is mainly expressed in the ovaries of adult zebrafish, *sprtn* is more highly expressed in the testis. *Acrc* is expressed more than *sprtn* in the ovaries by a factor of seven, whereas *sprtn* is expressed seven times higher in the testis. Considering the published data and that both genes show very high expression in zebrafish gonads, we propose that they both protect germ cells from DPC-induced DNA damage.

Interestingly, during the zebrafish development which is a highly replicative environment, we have found that *acrc* is consistently more expressed than *sprtn* until 72 hpf (Figure 15 C). The early embryo relies exclusively on maternal RNA transcripts and proteins, and the zygotic genome is not being transcribed until maternal to zygotic transition (4 hpf). Until 6 hpf, both zygotic and maternal transcripts are present, while at 6 hpf maternal transcripts are gone (Laue *et al.*, 2019), while the protein product main remain for several more hours. At the beginning of zygotic transcription, microRNAs are transcribed, which causes maternal transcripts to degrade. During the period of the first 10 divisions (3 hpf), embryos reach the 1000-cell stage through the rapid cell division (Laue et al., 2019). The maternal transcripts detected until 6hpf of both proteases are very highly expressed, suggesting that they both act during replication. Maternal transcripts of *acrc* are more dominant than *sprtn* (4.3-fold at 1 hpf and 4.9-fold at 4 hpf), suggesting that *acrc* has a crucial role in this period of development. In the highly replicative environment, up to 6 hpf, there is likely an increased requirement for DNA damage repair pathways, including DPC repair. Replication fidelity is critical at this stage and any unrepaired DNA damage would be lethal for the embryo. Expression of zygotic *acrc* begins at 12 hpf and is stable until 24 hpf, when expression begins to decline towards the 72 hpf. It is important to note, that despite the decrease in expression in comparison to early stages, *acrc* expression is still at very high level ($MNE*10^6$= 4401±1366) (Figure 15 C), indicating its importance in also later developmental stages.

A similar expression pattern was observed for *sprtn*, although it is less expressed than *acrc* at all stages of embryonic development. Acrc and Sprtn could have the same role or could be equally important during early zebrafish development. *Sprtn* mRNA levels are highest at 1 hpf, which also corresponds to maternal mRNA for Sprtn protein, whereas zygotic *sprtn* expression begins 12 hpf and remains at the same level until 72 hpf. According to MNE threshold values, *sprtn* is highly expressed in most of the time points observed, with exception at 1 hpf were expression is very high and at 24 hpf were expression is moderate (Figure 15 C).

4.5. Creation and optimization of cellular and animal models to study the role of ACRC in DPC repair

ACRC is a SprT domain-containing protein and a second potential DPC protease in higher eukaryotes, besides SPRTN. To investigate the role of ACRC in DPC repair, two zebrafish deficient strains were created using the CRISPR/Cas9 system. sgRNA 1 and 2 were designed and injected into the yolk of one-cell stage embryos together with Cas9 protein (Figure 16). HRM analysis was used to distinguish embryos carring a mutation from those without mutations. Embryos that were HRM positive (T_m different from WT embryos) were sequenced to determine the type of mutation. ACRC has an active site similar to that of SPRTN and contains a HEXXH motif (Vaz et al., 2016). However, considering that IDR domain of ACRC is possibly involved in cellular processes other than DPC repair, we have decided to mutate its protease core in zebrafish in order to follow the consequences of its impaired enzymatic activity, presumably specific for DPC repair. We identified the predicted active site residue (E451) within the conserved active site in zebrafish *acrc*, and this residue was targeted to create an enzymatic dead version of Acrc. The strain carrying *acrc* enzymatic mutation (ΔEMCH) has a deletion of the catalytic glutamate E451 and the following three amino acids including the Zn-bearing histidine H454, which together form the HEXXH motif (Figures 28 and 29). In addition, we have created strain carrying C-terminal deletion ΔC which has a premature stop codon at position 473 that results in a frameshift mutation in the C-terminal part of the protein downstream of the active protease site and the HEXXH motif, but includes a part of the SprT-like domain (Figures 35 and 36). This strain was created to explore the consequence of disrupted SprT protease domain on Acrc activity in DPC repair.

We found that the offspring of Acrc-deficient homozygous mothers of both strains have a maternal lethality phenotype regardless of the paternal genome (WT or Acrc homozygous/heterozygous mutant). Embryos survive to 6 hpf and die between 6 and 12 hpf. As already described, maternal transcripts are known to be present in embryos during the first 6 hours of embryonic development, whereas proteins encoded by these mRNAs are likely to be present for longer (Laue *et al.*, 2019). mRNA transcription from the zygotic genome begins at 2.8 hours of development and reaches its highest level at 6 h (Laue *et al.*, 2019). However, based on the results of our expression analysis by qPCR and observed mortality phenotype, we hypothesize that expression of the *acrc* gene from the embryo genome begins at approximately 12 hours after fertilization and remains at the same level until 24 hours, after which expression begins to decrease, while still remaining at very high levels (MNE*10^6= 8771±6033) (Figure 15 C). The amount of Acrc mRNA is highest 1h after fertilization (Figure 15 C), which corresponds to the mRNA inherited from the mother through the cytoplasm of the egg cell, since embryos do not yet have gene expression arising from their own genome. Embryos inherit mRNA for Acrc from the mother, from which Acrc is translated into the functional protein. All these data suggest that 6 hours old embryos do not have their own expression of Acrc, but their development depends on maternally inherited Acrc mRNA. Because the Acrc protein can be inherited exclusively through the oocyte cytoplasm, embryo development and phenotype are independent of the paternal genome. Embryos obtained from homozygous mothers inherit the mutated ACRC allele, and because only mutant Acrc mRNA is present in the cytoplasm of the oocyte, which is also inherited, all embryos inherit

non-functional Acrc through the maternal effect and have a 100% lethality phenotype (Figures 30 and 37).

The result of our Western blot analysis to determine Acrc protein levels in homozygous embryos showed that the same post-translational modification is present in all samples at a size of 100 kDa, while two other post-translationally modified variants of Acrc are present: (1) at 6 hpf in both WT and ΔEMCH embryos (between 100-150 kDa) and (2) in 3 days old WT embryos at around 200 kDa (Figure 42). This is the first observation of postranslationally modified forms of Acrc protein *in vivo* which adds another layer of complexity to the regulation of DPC proteases. We hypothesize that ubiquitin and/or SUMO molecules constitute the post-translational modifications (PTMs) on ACRC considering similar observation for SPRTN protease (Stingele, Habermann and Jentsch, 2015; Lopez-Mosqueda et al., 2016). In human cells, SPRTN is monoubiquitinated, which prevents his access to chromatin. Upon DPC induction, ubiquitin-specific protease 7 (USP7) deubiquitinates SPRTN, regulating its chromatin localization and recruitment to DPC lesions (Perry et al., 2021). However, the PTMs of ACRC and their role in regulation and function of ACRC will be the subject of future studies.

Beside developed zebrafish model to study the role of ACRC in DPC repair, we have also successfully optimized ACRC silencing by transfection of small interfering RNA (siRNA) in RPE-1 and HEK293T cells. These results are first of a kind considering ACRC was not silenced so far in human cells and will be a valuable tool to study the function of ACRC in DPC repair in cells.

4.6. The role of ACRC in DPC repair

Our results showed that *acrc* is expressed 2-5 times more than *sprtn* during early zebrafish development, suggesting that Acrc may be the dominant protease for the resolution of DPCs during embryonic development (Figure 15 C). Indeed, studies in invertebrates showed a role of Acrc in DPC repair in the germline and during embryonic development (Borgermann *et al.*, 2019; Bhargava *et al.*, 2020; Dokshin *et al.*, 2020). So far, the only study that connected Acrc and DPC repair directly is done by Bhargava et al (2020) who measured total DPC levels in Acrc deficient zebrafish embryos and *Drosophila* ovaries and found very mild increase in total levels in ACRC mutants. In order to prove our hypothesis that catalytic function of Acrc is crucial for DPC repair in zebrafish embryos and that Acrc is indeed an active protease, we have measured total DPC levels in ΔEMCH zebrafish strain. Specifically, after DPC induction with formaldehyde in 3 days old progeny from ΔEMCH heterozygous parents, DPCs accumulated 2.2-fold in Acrc mutant progeny in comparison to WT embryos (Figure 45 B), which is an indirect proof that Acrc is indeed a protease and a direct proof that catalytic function of Acrc is needed for DPC repair. Analyzed mutant progeny was a mixture of homozygotes (25%), heterozygotes (50 %) and WTs (25%). In physiological conditions, without formldehyde treatement, DPC levels were not increased in the progeny from ΔEMCH heterozygous parents (Figure 45 B). Most probably, the effect is not visible given that progeny contained 25% WT embryos and 50% heterozygous embryos while 25% of homozygous embryos died after 6 hpf. This hypotheses cannot be tested in 3 day old homozygous embryos because they do not survive 6 hpf (Figure 30). These observations are consistent with rescue experiments in which embryos injected with mRNA for Acrc protein with the E451A mutation

did not survive (Figure 43). It is likely that Acrc protein in WT embryos repairs DPCs formed by formaldehyde treatment, so that they do not show increased DPC levels compared with untreated embryos. In contrast, homozygotes and heterozygotes for ΔEMCH mutation had impaired repair of DPCs due to catalytic mutation in the Acrc protease core resulting in the increased DPC levels after formaldehyde treatment.

To further investigate the role of Acrc Sprt domain in DPC repair, we have used the ΔC zebrafish strain. A progeny from female Acrc ΔC F0 founder fish was crossed with a WT male. Embryos were collected 6 hpf because they exhibited 100% mortality between 6 hpf and 12 hpf, regardless of male genotype (Figure 37). Based on the observed phenotype, we hypothesize that the F0 females gave homozygous oocytes for the ΔC mutation, which could be confirmed by oocyte genotyping. We suggest that embryos inherited the mutant *acrc* allele from the female, while they inherited the WT allele from the male, making them heterozygous for ΔC mutation. Compared with WT embryos, ΔC embryos without formaldehyde treatment had 1.8 times more DPCs (Figure 45 A). Considering that western blot analysis showed negligible levels of truncated Acrc ΔC protein, we hypotheses that DPC accumulation at 6 hpf is due to lack of functional Acrc protein. Embryos at 6 hpf of age do not have their own expression of Acrc protein, but their development depends on maternal Acrc mRNA which bares ΔC mutation, while the paternal WT allele does not come into play at this stage (Laue et al, 2019). According to the mortality phenotype, all embryos inherit a non-functional Acrc protein by maternal effect (Laue et al., 2019). Given that truncated Acrc is present at very low levels and that truncated protein carries catalytic glutamate at position 451 (stop codon is at position 473), it remains possible that DPC repair via Acrc is active to some extent in these embryos, and therefore DPC accumulation is not even higher than 1.8 fold than in WT embryos. We have tried to challenge the system with formaldehyde at 6 hpf, but in three independent experiments we could not observe increase in DPC levels in WT embryos at 6 hpf (Marin Kutnjak, IRB, personal communication), and thus we did not have a positive control for DPC measurements in this setup at 6 hpf.

Bhargava and co-workers created homozygous mutant zebrafish strain containing premature stop codon at the beginning of *acrc* gene and found that the offspring of mutant mothers exhibited various morphological defects, including chromosomal instability, asynchronous mitotic divisions, and tangled chromosomes, eventually resulting in cell death. Similar phenotypes were observed in mutant Acrc flies, worms, and zebrafish (Bhargava et al., 2020), suggesting that Acrc has conserved functions throughout the animal kingdom. In comparison, our results from both mutant zebrafish strains further explains the cause of the observed phenotypes. Specifically, catalytic function and disruption of Acrc SprT domain are responsible for the early embryonic cell death (Figures 30 and 37). When crossing two Acrc heterozygous fish carrying mutation, the expected ratio of progeny is 25% WT, 50% heterozygous and 25% homozygous embryos. Although 25% of the embryos were homozygous and both *acrc* alleles were mutated on their chromosomes, 100% of these embryos survived because their mother had a WT *acrc* allele and a mutant *acrc* allele. Both WT and mutant *acrc* mRNAs for Acrc were present in the cytoplasm of the oocyte of this females, and the embryos inherited both. Because they inherited WT mRNAs for *acrc* that were translated into functional Acrc protein, the embryos survived early embryonic development. Although the functional Acrc protein cannot be translated from the zygotic

mRNAs because both alleles in the progeny are mutated, these embryos survive later without any adverse phenotype or defect, indicating that the Acrc protein is a crucial limiting factor during early embryonic development. This result also suggests that enzymatic activity and E451 catalytic glutamate are important during early zebrafish development, which was further proven by recue experiments which allow us to distinguish which motif, domains, and residues in Acrc are responsible for the lethal phenotype (Cecile Otten, IRB, unpublished results). As excpected, the mRNA encoding the Acrc WT protein resulted in complete recovery of the mortality phenotype. After injection of *acrc* mRNA with a mutation in the putative protease site, E451 catalytic glutamate, the adverse phenotype persisted (Figure 43). This observation proves that the catalytic function of Acrc is critical for embryo survival during embryonic development. It also implies that Acrc is an active protease, which is the first demonstration of its protease activity to date. Bhargava et al (2020) showed that the SprT domain is in fruit fly essential for preventing DNA damage in the germline, but that the enzymatic activity of Acrc is not essential for the proper regulation of chromosome segregation and cell cycle during early embryogenesis. They created an Acrc enzymatic dead mutant fruit fly, which demonstrated that the SprT enzymatic domain is needed to prevent the induction of DNA damage during oogenesis in *Drosophila*, indicating the role of the SprT domain in maintaining the genomic integrity of germ cells in fruit fly. The enzymatic Acrc mutant transgene partially rescued the maternal-lethality phenotype of Acrc mutant fruit fly by decreasing chromosome bridges and other mitotic defects. This indicates that Acrc enzymatic activity is not essential for the proper chromosome segregation and cell cycle during early embryogenesis in fruit fly (Bhargava et al., 2020), while in vertebrates the role of Acrc protease activity in chromosome segregation remains to be investigated.

In zebrafish adults carrying C-terminal deletion, the truncated Acrc protein is negligibly expressed (Figure 33). Considering that ΔC mutated protein (amino acids 473- 586) includes part of the structured SprT domain (containing amino acids 378 - 535) and overall lacks 113 amino acids (19%), most probably this truncation and disruption of SprT domain at position 473 leads to unstable conformation of Acrc protein and subsequent premature protein degradation due to impaired protein folding (Ciechanover and Kwon, 2017). We set out to investigate this effect futher and found that embryos injected with *acrc* mRNAs in which the C-terminus without SprT domain (amino acids 535-586) was deleted showed no mortality phenotype, indicating that the C-terminus itself is not important for early zebrafish development, but that specifically the SprT domain C-terminal part is essential (Figure 44). In addition, while determining Acrc protein levels in our mutant strains, we have noticed that embryos or adults of both mutant Acrc strains have post-translationally modified Acrc protein (Figure 42).

Besides the importance of Sprt domain and its catalytic function that we have demonstrated, an unstructured IDR domain in the N-terminal half of the protein was investigated by Bhargava et al (2020). They have shown that in flies and worms IDR domain promotes chromosomal stability during early embryogenesis of invertebrates. ACRC mutant *C. elegans* containing only the IDR domain and lacking Sprt domain does not exhibit a reduction in average brood size over the first eight generations, indicating that germ cell functions, in addition to Sprt domain, also depend on the IDR. However, after 15 generations a subset of

the population started to exhibit sterility phenotype, indicating that both the catalytic domain and the C-terminus of the Acrc protein are required to prevent genome instability across generations. In summary, these observations show that both Sprt domain and IDR are required to ensure genome integrity (Bhargava *et al.*, 2020). In addition to described results from *C.elegans* model, one study in mouse showed that knock-out of ACRC in mice impairs male fertility (Carmell *et al.*, 2016). Interestingly, mouse is the only known species of eukayra with ACRC which lacks Sprt domain, while IDR is intact (Carmell *et al.*, 2016), which indicates that in mice, IDR effects male fertility. Recent study in humans and mice shed light on the mechanism of action of the IDR domain and showed that ACRC protects the spermatogonial stem cells (SSC) pool, suggesting that the ACRC IDR domain has a crucial role during mammalian spermatogenesis (Ribeiro and Crossan, 2022). ACRC deficient mice were unable to maintain the SSC pool through the lifetime, leading to reduction in sperm production. Indeed, similar phenotype was observed in several azoospermic men carrying mutations in the ACRC locus (Hardy *et al.*, 2021). During our breeding experiments, we observed that homozygous males carrying enzymatic mutation with intact IDR domain are fertile, while males carrying C-terminal deletion with also inact IDR domain, but reduced Acrc protein levels, have reduced fertility.

Furthermore, in our rescue experiments, Acrc mRNAs lacking the IDR domain failed to rescue the mortality phenotype (Figure 44) which indicates that IDR is crucial for early development of vertebrates. However, the stability of protein forming from mRNA lacking IDR remains to be tested in embryos using western blot in order to confirm these results. Considering that the mouse ACRC WT is similiar in topology to zebrafish Acrc-IDR construct and consists only of the N- terminal part of the protein, without the SprT domain and the C-terminal tail, we expect the protein to be stable. Acrc mRNAs lacking the RKP domain also failed to rescue the mortality phenotype, indicating important role of this domain during development, however stability of protein forming from injected mRNA needs to also be tested (Figure 44). Given the similarities of the SPRTN and ACRC protease cores, we wanted to test whether ACRC deficiency could be compensated by expression of the SPRTN protease. Injection of SPRTN WT mRNA failed to rescue the lethal phenotype, indicating that the role of Acrc protein during early zebrafish development cannot be replaced by SPRTN protease (Figure 43) (Cecile Otten, IRB, unpublished results). However, it remains to test SPRTN protein levels in zebrafish embryos after the injection in order to confirm that our mRNA SPRTN WT construct is producing sufficient amounts of stable SPRTN WT protein. We are currently developing a zebrafish Sprtn antibody, since it is not commercially available.

In this thesis, we have shown that Acrc is a protease involved in DPC repair *in vivo* in zebrafish animal model, while others have shown that ACRC associates with the replication machinery, based on its ability to immunoprecipitate components of the MCM complex, and that ACRC deficient germ cells and human germ cell tumors suffer from replicative stress (Bhargava *et al.*, 2020). Considering that SPRTN protease is also active in DPC repair specifically during replication (Vaz *et al.*, 2016), the question arises as to why the cells evolved two different proteins with possibly redundant functions. One possible explanation, at least in germ cells could be that germ cells have an increased DPC load in comparison to somatic cells and are subjected to extensive epigenetic reprogramming, including processes such as histone demethylation, in which DPC-inducing agents are produced as a by-products (Stingele, Bellelli

and Boulton, 2017). In addition, enzymatic DPCs are generated during meiotic DSB induction when the Spo11 endonuclease becomes covalently attached to DNA at the break site, and germ cells (particularly oocytes) experience extended periods of cell cycle arrest during which they accumulate DPCs (Bhargava *et al.*, 2020). Another explanation in relation to highly replicative environment during early vertebrate development is that fast divisions increase the indicence of replisome encountering physiological DPCs in the cell. In order to ensure high fidelity during replication, two proteases might be needed to deal with DPC replication blocks and ensure the embryo development and survival.

5. Conclusions

This study provides the first comprehensive study of the DPC repair mechanisms in zebrafish (*Danio rerio*) as an important vertebrate model. It revealed similarities between zebrafish and human orthologs of ACRC and XPA proteins, characterized the importance of ACRC during early zebrafish embryonic development, discovered that catalytic function of ACRC is crucial for early embryonic development, and showed the first indications that ACRC and NER are involved in DPC repair at the organismal level. The most important conclusions of this study are:

1. Zebrafish Xpa protein shares one-to-one orthology with human XPA protein;

2. Zebrafish *xpa* and *acrc* genes are syntenic to the human genes;

3. Considering high expression of *acrc* and *sprtn* in zebrafish gonads, we propose that they protect germ cells from DPC-induced DNA damage. *Acrc* is expressed more than *sprtn* during early zebrafish development, indicating that *Acrc* may be the dominant protease for resolution of DPCs during embryonic development;

4. *Acrc* is expressed throughout zebrafish tissues and is not restricted only to gonads, thus indicating that, as opposed to current opinion in the field, Acrc most probably has cellular functions other than the protection of germ cells;

5. Disruption of *Acrc* in zebrafish results in a maternal-effect lethality phenotype, suggesting that maternally inherited *Acrc* is essential for early zebrafish embryonic development. Our results showed that disruption of specifically catalytic function (E451 deletion) and *Acrc* SprT domain are responsible for the early embryonic cell death;

6. Double Xpa and Sprtn deficiency showed additive effect on DPC levels in zebrafish embryos, suggesting that, contrary to the hypothesis and opinion in the field, Xpa and Sprtn may act independently in the repair of general cellular DPCs;

7. NER pathway is involved in repair of DNA-histone crosslinks in human cells: XPA deficient cells accumulate histone H1-DPCs and H3-DPCs;

8. E451 catalytic glutamate of Acrc is needed for DPC repair, which is also an indirect proof of Acrc proteolytic activity, while the structured SprT domain is important for protein stability and activity;

9. Zebrafish Acrc mutant strains created within this thesis can be used for characterization of DPC repair in adult tissues in follow-up studies.

6. References

Anderson, J.L. *et al.* (2012) "Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics," *PLoS ONE*, 7(7). doi:10.1371/journal.pone.0040701.

Aparicio, T. *et al.* (2016) "MRN, CtIP, and BRCA1 mediate repair of topoisomerase II-DNA adducts," *Journal of Cell Biology*, 212(4), pp. 399–408. doi:10.1083/jcb.201504005.

Barve, A. *et al.* (2021) "DNA Repair Repertoire of the Enigmatic Hydra," *Frontiers in Genetics*, 12(April). doi:10.3389/fgene.2021.670695.

Bhargava, V. *et al.* (2020) "GCNA Preserves Genome Integrity and Fertility Across Species," *Developmental Cell*, 52(1), pp. 38-52.e10. doi:10.1016/j.devcel.2019.11.007.

Bodnar, A.G. *et al.* (1998) "Extension of life-span by introduction of telomerase into normal human cells," *Science*, 279(5349), pp. 349–352. doi:10.1126/science.279.5349.349.

de Boer, J. and Hoeijmakers, J.H.J. (2000) *Nucleotide excision repair and human syndromes*, *Carcinogenesis*.

Borgermann, N. *et al.* (2019) " SUMO ylation promotes protective responses to DNA -protein crosslinks ," *The EMBO Journal*, 38(8). doi:10.15252/embj.2019101496.

Bradford, M.M. (1976) *A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding*, *ANALYTICAL BIOCHEMISTRY*.

Bradford, Y.M. *et al.* (2017) "Zebrafish models of human disease: Gaining insight into human disease at ZFIN," *ILAR Journal*, 58(1), pp. 4–16. doi:10.1093/ilar/ilw040.

Burger, A. *et al.* (2016) "Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes," *Development (Cambridge)*, 143(11), pp. 2025–2037. doi:10.1242/dev.134809.

Carmell, M.A. *et al.* (2016) "A widely employed germ cell marker is an ancient disordered protein with reproductive functions in diverse eukaryotes," *eLife*, 5(OCTOBER2016), pp. 1–25. doi:10.7554/eLife.19993.

Centore, R.C. *et al.* (2012) "Spartan/C1orf124, a Reader of PCNA Ubiquitylation and a Regulator of UV-Induced DNA Damage Response," *Molecular Cell*, 46(5), pp. 625–635. doi:10.1016/j.molcel.2012.05.020.

Charpentier, E. and Marraffini, L.A. (2014) "Harnessing CRISPR-Cas9 immunity for genetic engineering," *Current Opinion in Microbiology*. Elsevier Ltd, pp. 114–119. doi:10.1016/j.mib.2014.07.001.

Chatterjee, N. and Walker, G.C. (2017) "Mechanisms of DNA damage, repair, and mutagenesis," *Environmental and Molecular Mutagenesis*, 58(5), pp. 235–263. doi:10.1002/em.22087.

Chesner, L.N. and Campbell, C. (2018) "A quantitative PCR-based assay reveals that nucleotide excision repair plays a predominant role in the removal of DNA-protein crosslinks from plasmids transfected into mammalian cells," *DNA Repair*, 62, pp. 18–27. doi:10.1016/j.dnarep.2018.01.004.

Ciechanover, A. and Kwon, Y.T. (2017) "Protein quality control by molecular chaperones in neurodegeneration," *Frontiers in Neuroscience*, 11(APR), pp. 1–18. doi:10.3389/fnins.2017.00185.

Cimprich, K.A. and Cortez, D. (2008) “ATR: An essential regulator of genome integrity,” *Nature Reviews Molecular Cell Biology*, pp. 616–627. doi:10.1038/nrm2450.

Davis, E.J. *et al.* (2012) “DVC1 (C1orf124) recruits the p97 protein segregase to sites of DNA damage,” *Nature Structural and Molecular Biology*, 19(11), pp. 1093–1100. doi:10.1038/nsmb.2394.

Doetschman, T. and Georgieva, T. (2017) “Gene Editing with CRISPR/Cas9 RNA-Directed Nuclease,” *Circulation Research*, 120(5), pp. 876–894. doi:10.1161/CIRCRESAHA.116.309727.

Dokshin, G.A. *et al.* (2020) “GCNA Interacts with Spartan and Topoisomerase II to Regulate Genome Stability,” *Developmental Cell*, 52(1), pp. 53-68.e6. doi:10.1016/j.devcel.2019.11.006.

Duan, M. *et al.* (2021) “Transcription-coupled nucleotide excision repair: New insights revealed by genomic approaches,” *DNA Repair*, 103(April), p. 103126. doi:10.1016/j.dnarep.2021.103126.

Duxin, J.P. *et al.* (2014) “Repair of a DNA-protein crosslink by replication-coupled proteolysis,” *Cell*, 159(2), pp. 346–357. doi:10.1016/j.cell.2014.09.024.

Fielden, J. *et al.* (2018) “DNA protein crosslink proteolysis repair: From yeast to premature ageing and cancer in humans,” *DNA Repair*, 71(August), pp. 198–204. doi:10.1016/j.dnarep.2018.08.025.

Fielden, J. *et al.* (2020) “TEX264 coordinates p97- and SPRTN-mediated resolution of topoisomerase 1-DNA adducts,” *Nature Communications*, 11(1), pp. 1–16. doi:10.1038/s41467-020-15000-w.

Gagnon, J.A. *et al.* (2014) “Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs,” *PLoS ONE*, 9(5), pp. 5–12. doi:10.1371/journal.pone.0098186.

Ghodke, P.P. *et al.* (2021) “Enzymatic bypass of an N6-deoxyadenosine DNA–ethylene dibromide–peptide cross-link by translesion DNA polymerases,” *Journal of Biological Chemistry*, 296. doi:10.1016/j.jbc.2021.100444.

Guindon, S. and Gascuel, O. (2003) “A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood,” *Systematic Biology*, 52(5), pp. 696–704. doi:10.1080/10635150390235520.

Guo, T. *et al.* (2018) “Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9-mediated genome editing,” *Genome Biology*, 19(1), pp. 1–20. doi:10.1186/s13059-018-1518-x.

Hakem, R. (2008) “DNA-damage repair; the good, the bad, and the ugly,” *EMBO Journal*, 27(4), pp. 589–605. doi:10.1038/emboj.2008.15.

Hardy, J.J. *et al.* (2021) “Variants in GCNA, X-linked germ-cell genome integrity gene, identified in men with primary spermatogenic failure,” *Human Genetics*, 140(8), pp. 1169–1182. doi:10.1007/s00439-021-02287-y.

Harper, J.W. and Elledge, S.J. (2007) “The DNA Damage Response: Ten Years After,” *Molecular Cell*. Cell Press, pp. 739–745. doi:10.1016/j.molcel.2007.11.015.

Hindul, N.L. *et al.* (2022) “Construction of a human hTERT RPE-1 cell line with inducible Cre for editing of endogenous genes,” *Biology Open*, 11(2). doi:10.1242/bio.059056.

Hoa, N.N. *et al.* (2016) “Mre11 Is Essential for the Removal of Lethal Topoisomerase 2 Covalent Cleavage Complexes,” *Molecular Cell*, 64(3), pp. 580–592. doi:10.1016/j.molcel.2016.10.011.

Hoshijima, K., Jurynec, M.J. and Grunwald, D.J. (2016) “Precise Editing of the Zebrafish Genome Made Simple and Efficient,” *Developmental Cell*, 36(6), pp. 654–667. doi:10.1016/j.devcel.2016.02.015.

Howe, K. *et al.* (2013) “The zebrafish reference genome sequence and its relationship to the human genome,” *Nature*, 496(7446), pp. 498–503. doi:10.1038/nature12111.

Hruscha, A. *et al.* (2013) “Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish,” *Development (Cambridge)*, 140(24), pp. 4982–4987. doi:10.1242/dev.099085.

Hryhorowicz, M. *et al.* (2017) “CRISPR/Cas9 Immune System as a Tool for Genome Engineering,” *Archivum Immunologiae et Therapiae Experimentalis*, 65(3), pp. 233–240. doi:10.1007/s00005-016-0427-5.

Hsieh-Feng, V. and Yang, Y. (2020) “Efficient expression of multiple guide RNAs for CRISPR/Cas genome editing,” *aBIOTECH*, 1(2), pp. 123–134. doi:10.1007/s42994-019-00014-w.

Hsu, P.D., Lander, E.S. and Zhang, F. (2014) “Development and applications of CRISPR-Cas9 for genome engineering,” *Cell*. Elsevier B.V., pp. 1262–1278. doi:10.1016/j.cell.2014.05.010.

Hwang, W.Y. *et al.* (2013) “Efficient genome editing in zebrafish using a CRISPR-Cas system,” *Nature Biotechnology*, 31(3), pp. 227–229. doi:10.1038/nbt.2501.

Ide, H. *et al.* (2018) “DNA–protein cross-links: Formidable challenges to maintaining genome integrity,” *DNA Repair*, 71, pp. 190–197. doi:10.1016/j.dnarep.2018.08.024.

Jackson, S.P. and Bartek, J. (2009) “The DNA-damage response in human biology and disease,” *Nature*, 461(7267), pp. 1071–1078. doi:10.1038/nature08467.

Jao, L.E., Wente, S.R. and Chen, W. (2013) “Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system,” *Proceedings of the National Academy of Sciences of the United States of America*, 110(34), pp. 13904–13909. doi:10.1073/pnas.1308335110.

Jiricny, J. (2013) “Postreplicative mismatch repair,” *Cold Spring Harbor Perspectives in Biology*, pp. 1–23. doi:10.1101/cshperspect.a012633.

Khadempar, S. *et al.* (2019) “CRISPR–Cas9 in genome editing: Its function and medical applications,” *Journal of Cellular Physiology*, 234(5), pp. 5751–5761. doi:10.1002/jcp.27476.

Kiianitsa, K. and Maizels, N. (2013) “A rapid and sensitive assay for DNA-protein covalent complexes in living cells,” *Nucleic Acids Research*, 41(9), pp. 1–7. doi:10.1093/nar/gkt171.

Klages-Mundt, N.L. and Li, L. (2017) “Formation and repair of DNA-protein crosslink damage,” *Science China Life Sciences*, 60(10), pp. 1065–1076. doi:10.1007/s11427-017-9183-4.

Kong, E.Y., Cheng, S.H. and Yu, K.N. (2016) “Zebrafish as an in vivo model to assess epigenetic effects of ionizing radiation,” *International Journal of Molecular Sciences*, 17(12), pp. 1–21. doi:10.3390/ijms17122108.

Kusakabe, M. *et al.* (2019) “Mechanism and regulation of DNA damage recognition in nucleotide excision repair,” *Genes and Environment*, 41(1). doi:10.1186/s41021-019-0119-6.

Kuykendall, J.R. and Bogdanffy, M.S. (1992) *Efficiency of DNA-histone crosslinking induced by saturated and unsaturated aldehydes in vitro*, *Mutation Research*.

Lans, H. and Vermeulen, W. (2011) “ Nucleotide Excision Repair in Caenorhabditis elegans ,” *Molecular Biology International*, 2011, pp. 1–12. doi:10.4061/2011/542795.

Laue, K. *et al.* (2019) "The maternal to zygotic transition regulates genome-wide heterochromatin establishment in the zebrafish embryo," *Nature Communications*, 10(1). doi:10.1038/s41467-019-09582-3.

Lawrence, C. (2007) "The husbandry of zebrafish (Danio rerio): A review," *Aquaculture*, 269(1–4), pp. 1–20. doi:10.1016/j.aquaculture.2007.04.077.

Lehmann, J. *et al.* (2017) "XPF knockout via CRISPR/Cas9 reveals that ERCC1 is retained in the cytoplasm without its heterodimer partner XPF," *Cellular and Molecular Life Sciences*, 74(11), pp. 2081–2094. doi:10.1007/s00018-017-2455-7.

Lessel, D. *et al.* (2014) "Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features," *Nature Genetics*, 46(11), pp. 1239–1244. doi:10.1038/ng.3103.

Li, G.M. (2008) "Mechanisms and functions of DNA mismatch repair," *Cell Research*, 18(1), pp. 85–98. doi:10.1038/cr.2007.115.

Li, X. and Heyer, W.-D. (2008) "Homologous recombination in DNA reapir and DNA," *Cell Research*, 18(1), pp. 99–113. doi:10.1038/cr.2008.1.Homologous.

Li, Y. *et al.* (2021) "Progress in gene-editing technology of Zebrafish," *Biomolecules*, 11(9). doi:10.3390/biom11091300.

Lin, J. *et al.* (2016) "A clinically relevant in vivo zebrafish model of human multiple myeloma to study preclinical therapeutic efficacy," *Blood*, 128(2), pp. 249–252. doi:10.1182/blood-2016-03-704460.

Liu, M. *et al.* (2019) "Methodologies for improving HDR efficiency," *Frontiers in Genetics*, 10(JAN), pp. 1–9. doi:10.3389/fgene.2018.00691.

Lončar, J. *et al.* (2016) "The first characterization of multidrug and toxin extrusion (MATE/SLC47) proteins in zebrafish (Danio rerio)," *Scientific Reports*, 6(February), pp. 1–15. doi:10.1038/srep28937.

Lopez-Mosqueda, J. *et al.* (2016) "SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks," *eLife*, 5(NOVEMBER2016), pp. 1–19. doi:10.7554/eLife.21491.

Louis, A., Muffato, M. and Crollius, H.R. (2013) "Genomicus: Five genome browsers for comparative genomics in eukaryota," *Nucleic Acids Research*, 41(D1), pp. 700–705. doi:10.1093/nar/gks1156.

Ma, Y., Zhang, L. and Huang, X. (2014) "Genome modification by CRISPR/Cas9," *FEBS Journal*, 281(23), pp. 5186–5193. doi:10.1111/febs.13110.

Marteijn, J.A. *et al.* (2014) "Understanding nucleotide excision repair and its roles in cancer and ageing," *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 465–481. doi:10.1038/nrm3822.

Martin, L.J. (2008) "DNA damage and repair: Relevance to mechanisms of neurodegeneration," *Journal of Neuropathology and Experimental Neurology*, 67(5), pp. 377–387. doi:10.1097/NEN.0b013e31816ff780.

Maskey, R.S. *et al.* (2014) "Spartan deficiency causes genomic instability and progeroid phenotypes," *Nature Communications*, 5. doi:10.1038/ncomms6744.

Mehravar, M. *et al.* (2019) "Mosaicism in CRISPR/Cas9-mediated genome editing," *Developmental Biology*, 445(2), pp. 156–162. doi:10.1016/j.ydbio.2018.10.008.

Minko, I.G., Zou, Y. and Lloyd, R.S. (2001) *Incision of DNA-protein crosslinks by UvrABC nuclease suggests a potential repair pathway involving nucleotide excision repair*. Available at: www.pnas.orgcgidoi10.1073pnas.042700399.

Mórocz, M. *et al.* (2017) "DNA-dependent protease activity of human Spartan facilitates replication of DNA-protein crosslink-containing DNA," *Nucleic Acids Research*, 45(6), pp. 3172–3188. doi:10.1093/nar/gkw1315.

Moulton, J.D. (2017) "Using morpholinos to control gene expression," *Current Protocols in Nucleic Acid Chemistry*, 2017(March), pp. 4.30.1-4.30.29. doi:10.1002/cpnc.21.

Nakamura, J. and Nakamura, M. (2020) "DNA-protein crosslink formation by endogenous aldehydes and AP sites," *DNA Repair*, 88(February), p. 102806. doi:10.1016/j.dnarep.2020.102806.

Nakano, T. *et al.* (2007) "Nucleotide Excision Repair and Homologous Recombination Systems Commit Differentially to the Repair of DNA-Protein Crosslinks," *Molecular Cell*, 28(1), pp. 147–158. doi:10.1016/j.molcel.2007.07.029.

Nakano, T. *et al.* (2009) "Homologous recombination but not nucleotide excision repair plays a pivotal role in tolerance of DNA-protein cross-links in mammalian cells," *Journal of Biological Chemistry*, 284(40), pp. 27065–27076. doi:10.1074/jbc.M109.019174.

Nan, Y. and Zhang, Y.J. (2018) "Antisense phosphorodiamidate morpholino oligomers as novel antiviral compounds," *Frontiers in Microbiology*, 9(APR), pp. 1–15. doi:10.3389/fmicb.2018.00750.

Nolte, D. *et al.* (2001) "ACRC codes for a novel nuclear protein with unusual acidic repeat tract and maps to DYT3 (dystonia parkinsonism) critical interval in Xq13.1," *Neurogenetics*, 3(4), pp. 207–213. doi:10.1007/s100480100120.

Oleinick, N.L. *et al.* (1987) "The formation, identification, and significance of DNA-protein cross-links in mammalian cells," *British Journal of Cancer*, 55(SUPPL. 8), pp. 135–140.

Pachva, M.C. *et al.* (2020) "DNA-Histone Cross-Links: Formation and Repair," *Frontiers in Cell and Developmental Biology*, 8(December), pp. 1–14. doi:10.3389/fcell.2020.607045.

Patterson-Fortin, J. and D'Andrea, A.D. (2020) "Exploiting the microhomology-mediated end-joining pathway in cancer therapy," *Cancer Research*. American Association for Cancer Research Inc., pp. 4593–4600. doi:10.1158/0008-5472.CAN-20-1672.

Pawelczak, K.S. *et al.* (2018) "Modulating DNA Repair Pathways to Improve Precision Genome Engineering," *ACS Chemical Biology*, 13(2), pp. 389–396. doi:10.1021/acschembio.7b00777.

Perry, M. *et al.* (2021) "USP11 mediates repair of DNA–protein cross-links by deubiquitinating SPRTN metalloprotease," *Journal of Biological Chemistry*, 296, p. 100396. doi:10.1016/j.jbc.2021.100396.

Pommier, Y. *et al.* (2006) "Repair of Topoisomerase I-Mediated DNA Damage," *Progress in Nucleic Acid Research and Molecular Biology*, pp. 179–229. doi:10.1016/S0079-6603(06)81005-6.

Pradhan, A. and Olsson, P.-E. (2016) "Regulation of zebrafish gonadal sex differentiation," *AIMS Molecular Science*, 3(4), pp. 567–584. doi:10.3934/molsci.2016.4.567.

Prasad, R. *et al.* (2014) "Suicidal cross-linking of PARP-1 to AP site intermediates in cells undergoing base excision repair," *Nucleic Acids Research*, 42(10), pp. 6337–6351. doi:10.1093/nar/gku288.

Prill, K. and Dawson, J.F. (2020) “Homology-Directed Repair in Zebrafish: Witchcraft and Wizardry?,” *Frontiers in Molecular Biosciences*, 7(December), pp. 1–9. doi:10.3389/fmolb.2020.595474.

Pulzová, L.B., Ward, T.A. and Chovanec, M. (2020) “XPA: DNA repair protein of significant clinical importance,” *International Journal of Molecular Sciences*, 21(6). doi:10.3390/ijms21062182.

Rademakers, S. *et al.* (2003) “Xeroderma Pigmentosum Group A Protein Loads as a Separate Factor onto DNA Lesions,” *Molecular and Cellular Biology*, 23(16), pp. 5755–5767. doi:10.1128/mcb.23.16.5755-5767.2003.

Rahman Khan, F. and Sulaiman Alhewairini, S. (2019) “ Zebrafish (Danio rerio) as a Model Organism ,” in *Current Trends in Cancer Management*. IntechOpen. doi:10.5772/intechopen.81517.

Ribas, L. *et al.* (2017) “Heat-induced masculinization in domesticated zebrafish is family-specific & yields a set of different gonadal transcriptomes,” *Proceedings of the National Academy of Sciences of the United States of America*, 114(6), pp. E941–E950. doi:10.1073/pnas.1609411114.

Ribeiro, J. and Crossan, G.P. (2022) “GCNA is a histone binding protein required for spermatogonial stem cell maintenance.”

Ruggiano, A. *et al.* (2021a) “The protease SPRTN and SUMOylation coordinate DNA-protein crosslink repair to prevent genome instability,” *Cell Reports*, 37(10). doi:10.1016/j.celrep.2021.110080.

Ruggiano, A. *et al.* (2021b) “The protease SPRTN and SUMOylation coordinate DNA-protein crosslink repair to prevent genome instability,” *Cell Reports*, 37(10). doi:10.1016/j.celrep.2021.110080.

Ruggiano, A. and Ramadan, K. (2021) “DNA–protein crosslink proteases in genome stability,” *Communications Biology*, 4(1), pp. 1–11. doi:10.1038/s42003-020-01539-3.

Schärer, O.D. (2013) “Nucleotide excision repair in Eukaryotes,” *Cold Spring Harbor Perspectives in Biology*, 5(10), pp. 1–19. doi:10.1101/cshperspect.a012609.

Schellenberg, M.J. *et al.* (2017) “ZATT (ZNF451)–mediated resolution of topoisomerase 2 DNA-protein cross-links,” *Science*, 357(6358), pp. 1412–1416. doi:10.1126/science.aam6468.

Sczepanski, J.T. *et al.* (2010) “Rapid DNA-protein cross-linking and strand scission by an abasic site in a nucleosome core particle,” *PNAS*, 107(52), pp. 22475–22480. doi:10.1073/pnas.1012860108/-/DCSupplemental.

Sedgwick, B. *et al.* (2007) “Repair of alkylated DNA: Recent advances,” *DNA Repair*, 6(4), pp. 429–442. doi:10.1016/j.dnarep.2006.10.005.

Shin, J., Chen, J. and Solnica-Krezel, L. (2014) “Efficient homologous recombination-mediated genome engineering in zebrafish using TALE nucleases,” *Development (Cambridge, England)*, 141(19), pp. 3807–3818. doi:10.1242/dev.108019.

Solomon, M.J. and Varshavsky, A. (1985) “Formaldehyde-mediated DNA-protein crosslinking: A probe for in vivo chromatin structures,” *Proceedings of the National Academy of Sciences of the United States of America*, 82(19), pp. 6470–6474. doi:10.1073/pnas.82.19.6470.

Stepanenko, A.A. and Dmitrenko, V. v. (2015) “HEK293 in cell biology and cancer research: Phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution,” *Gene*. Elsevier B.V., pp. 182–190. doi:10.1016/j.gene.2015.05.065.

Stingele, J. *et al.* (2014) “A DNA-dependent protease involved in DNA-protein crosslink repair,” *Cell*, 158(2), pp. 327–338. doi:10.1016/j.cell.2014.04.053.

Stingele, J., Bellelli, R. and Boulton, S.J. (2017) “Mechanisms of DNA-protein crosslink repair,” *Nature Reviews Molecular Cell Biology*, 18(9), pp. 563–573. doi:10.1038/nrm.2017.56.

Stingele, J., Habermann, B. and Jentsch, S. (2015) “DNA-protein crosslink repair: Proteases as DNA repair enzymes,” *Trends in Biochemical Sciences*, 40(2), pp. 67–71. doi:10.1016/j.tibs.2014.10.012.

Sugitani *et al.* (2016) “XPA: a key scaffold for human NER,” *DNA Repair (Amst)*, 176(1), pp. 139–148. doi:10.1016/j.dnarep.2016.05.018.XPA.

Symington, L.S. and Gautier, J. (2011) “Double-strand break end resection and repair pathway choice,” *Annual Review of Genetics*, 45, pp. 247–271. doi:10.1146/annurev-genet-110410-132435.

Taylor, J.S. *et al.* (2003) “Genome duplication, a trait shared by 22,000 species of ray-finned fish,” *Genome Research*, 13(3), pp. 382–390. doi:10.1101/gr.640303.

Tiwari, V. and Wilson, D.M. (2019) “DNA Damage and Associated DNA Repair Defects in Disease and Premature Aging,” *American Journal of Human Genetics*, 105(2), pp. 237–257. doi:10.1016/j.ajhg.2019.06.005.

Tom, R., Bisson, L. and Durocher, Y. (2008) “Transfection of adherent HEK293-EBNA1 cells in a six-well plate with branched PEI for production of recombinant proteins,” *Cold Spring Harbor Protocols*, 3(3), pp. 1–5. doi:10.1101/pdb.prot4978.

Tran, M.H. *et al.* (2021) “A more efficient CRISPR-Cas12a variant derived from Lachnospiraceae bacterium MA2020,” *Molecular Therapy - Nucleic Acids*, 24(June), pp. 40–53. doi:10.1016/j.omtn.2021.02.012.

Tretyakova, N.Y., Groehler, A. and Ji, S. (2015) “DNA-Protein Cross-Links: Formation, Structural Identities, and Biological Outcomes,” *Accounts of Chemical Research*, 48(6), pp. 1631–1644. doi:10.1021/acs.accounts.5b00056.

Truglio, J.J. *et al.* (2006) “Structural basis for DNA recognition and processing by UvrB,” *Nature Structural and Molecular Biology*, 13(4), pp. 360–364. doi:10.1038/nsmb1072.

Varshney, G.K., Sood, R. and Burgess, S.M. (2015) “Understanding and Editing the Zebrafish Genome,” *Advances in Genetics*, 92, pp. 1–52. doi:10.1016/bs.adgen.2015.09.002.

Vaz, B. *et al.* (2016) “Metalloprotease SPRTN/DVC1 Orchestrates Replication-Coupled DNA-Protein Crosslink Repair,” *Molecular Cell*, 64(4), pp. 704–719. doi:10.1016/j.molcel.2016.09.032.

Vaz, B., Popovic, M. and Ramadan, K. (2017) “DNA–Protein Crosslink Proteolysis Repair,” *Trends in Biochemical Sciences*, 42(6), pp. 483–495. doi:10.1016/j.tibs.2017.03.005.

Waters, C.A. *et al.* (2014) “Nonhomologous end joining: A good solution for bad ends,” *DNA Repair*, 17, pp. 39–51. doi:10.1016/j.dnarep.2014.02.008.

White, M.F. and Allers, T. (2018) “DNA repair in the archaea-an emerging picture,” *FEMS microbiology reviews*. NLM (Medline), pp. 514–526. doi:10.1093/femsre/fuy020.

Wilson, C.A. *et al.* (2014) “Wild sex in zebrafish: Loss of the natural sex determinant in domesticated strains,” *Genetics*, 198(3), pp. 1291–1308. doi:10.1534/genetics.114.169284.

Yang, H. *et al.* (2020) “Methods favoring homology-directed repair choice in response to crispr/cas9 induced-double strand breaks,” *International Journal of Molecular Sciences*, 21(18), pp. 1–20. doi:10.3390/ijms21186461.

Zhang, H., Xiong, Y. and Chen, J. (2020) “DNA-protein cross-link repair: What do we know now?,” *Cell and Bioscience*, 10(1), pp. 1–10. doi:10.1186/s13578-019-0366-z.

Zhang, J.H. *et al.* (2016) “Optimization of genome editing through CRISPR-Cas9 engineering,” *Bioengineered*, 7(3), pp. 166–174. doi:10.1080/21655979.2016.1189039.

Zhang, Y., Zhang, Z. and Ge, W. (2018) “An efficient platform for generating somatic point mutations with germline transmission in the zebrafish by CRISPR/Cas9-mediated gene editing,” *Journal of Biological Chemistry*, 293(17), pp. 6611–6622. doi:10.1074/jbc.RA117.001080.

Zhang, Y.W. *et al.* (2011) “Poly(ADP-ribose) polymerase and XPF-ERCC1 participate in distinct pathways for the repair of topoisomerase I-induced DNA damage in mammalian cells,” *Nucleic Acids Research*, 39(9), pp. 3607–3620. doi:10.1093/nar/gkq1304.

Zhitkovich, A. and Costa, M. (1992) “A simple, sensitive assay to detect dna-protein cromlinks in intact cells and in vivo,” *Carcinogenesis*, 13(8), pp. 1485–1489. doi:10.1093/carcin/13.8.1485.

Zhu, X. *et al.* (2019) “Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9,” *Nature Structural and Molecular Biology*, 26(8), pp. 679–685. doi:10.1038/s41594-019-0258-2.

7. Summary

DNA is constantly exposed to various harmful endogenous and exogenous factors that can cause formation of different lesions in cells. DPCs are the second most common DNA lesions and occur with a frequency of about 6,000 times daily in the mammalian genome. At the cellular level, impaired DPC repair results in mutations, genomic instability, and/or cell death, while at the organismal level, it can cause tumors, premature aging, and neurodegenerative diseases. Despite the frequency and severe outcomes of DPCs, it is still unknown how DPC repair pathway is regulated and which other factors, besides SPRTN protease are involved, while almost nothing is known of DPC repair mechanism *in vivo.* Based on the previously published studies, it was assumed that NER pathway and ACRC protein could be involved in DPC repair. Therefore, the main goal of our study was to unravel regulation of DPC repair *in vitro* and *in vivo* using human cell lines and zebrafish (*Danio rerio*) and to characterize the role of NER and ACRC in DPC repair. In orther to achieve the main goal, specific goals of this study were characterization of Xpa and Acrc proteins in zebrafish, generation of ACRC and XPA deficient cellular and animal models, and optimization of methods for DPC isolation and detection in zebrafish.

Using methods including phylogenetic and syntenic analysis, tissue expression analysis, CRISPR/CAS9 genome editing, morpholino- and small interfering RNA (siRNA)-mediated gene silencing, we provided first characterization of zebrafish Xpa and Acrc proteins and developed models to study their role in DPC repair.

Phylogenetic analysis revealed conserved XPA orthologs in vertebrates and invertebrates. Zebrafish has one *xpa* gene which showed one-to-one orthology with human *XPA*. Synteny analysis showed that *xpa* is located on chromosome 1 in zebrafish and that gene environment is relatively conserved in comparison to human ortholog. Zebrafish *acrc* is positioned on the chromosome 14 and is syntenic to human *ACRC* gene.

Expression analysis showed that zebrafish *acrc* is predominantly expressed in testis and ovaries and that is more expressed than *sprtn* during early zebrafish embryonic development. Therefore, we suggest that ACRC is the dominant protease that maintains germ cells genome integrity and repairs DPCs during zebrafish embryonic development.

DPC isolation by RADAR assay from zebrafish embryos showed additive effect on DPC accumulation in the absence of both NER pathway and Sprtn protease, indicating that NER pathway *in vivo* repairs DPCs independently of SPRTN. However, this hypothesis will have to be confirmed by additional experiments. Embryos deficient in both Sprtn and Xpa strongly accumulated low molecular weight DPCs, suggesting that both, NER and SPRTN are involved in their removal *in vivo*. XPA deficient human cells treated with DPC-inducing agent formaldehyde had significantly increased levels of histone DPCs, indicating that NER pathway is crucial for H1 and H3 DPC removal in human cells.

Zebrafish Acrc mutants were successfully generated to study the role of Acrc in DPC repair. Embryos carrying Acrc enzymatic mutation and treated with formaldehyde had increased levels of total DPCs, which proves that catalytic function of Acrc is needed for DPC repair and that Acrc is indeed an active protease. Embryos carrying C-terminal deletion also accumulated total DPCs, even in the absence of the treatment, due to the lack of functional Acrc protein.

Due to the mortality phenotype observed in heterozygous embryos carrying enzymatic or C-terminal deletion mutation, we also showed that enzymatic activity and Acrc SprT domain are important during early zebrafish development. This was further confirmed by rescue experiments in embryos which showed that Acrc E451 catalytic glutamate, SprT domain and IDR domain are crucial for early development of vertebrates.

8. Sažetak

DNA je konstantno izložena raznim štetnim endogenim i egzogenim čimbenicima koji mogu uzrokovati nastanak različitih lezija u stanicama. DPC-evi su druge najčešće DNA lezije i javljaju se s učestalošću od oko 6000 puta dnevno u genomu sisavaca. Na staničnoj razini, narušeni popravak DPC-eva rezultira mutacijama, genomskom nestabilnošću i/ili staničnom smrti, dok na razini organizma može uzrokovati tumore, preuranjeno starenje i neurodegenerativne bolesti. Unatoč učestalosti i posljedicama nastanka DPC-eva, još uvijek je nepoznato kako je reguliran put popravka i koji su drugi čimbenici uključeni, osim SPRTN proteaze, dok se o mehanizmu popravka DPC-eva *in vivo* gotovo ništa ne zna. Na temelju prethodno objavljenih istraživanja, pretpostavljeno je da NER put i ACRC protein mogu biti uključeni u popravak DPC-eva. Stoga je glavni cilj našeg istraživanja bio otkriti regulaciju popravka DPC-eva *in vitro* i *in vivo* koristeći ljudske stanične linije i zebricu (*Danio rerio*) te karakterizirati ulogu NER-a i ACRC-a u popravku DPC-eva. Da bi se postigao glavni cilj, specifični ciljevi ovog istraživanja bili su karakterizacija Xpa i Acrc proteina u zebrici, stvaranje staničnih i životinjskih modela s utišanim ACRC i XPA genima te optimizacija metoda za izolaciju i detekciju DPC-eva u zebrice.

Koristeći metode koje uključuju filogenetsku i sintenijsku analizu, analizu tkivne ekspresije, CRISPR/CAS9 editiranje genoma, utišavanje gena posredovano morfolino oligonukleotidima i malim interferirajućim RNA (siRNA), pružili smo prvu karakterizaciju Xpa i Acrc proteina zebrice i razvili modele za proučavanje njihove uloge u popravaku DPC-eva.

Filogenetskom analizom otkriveno je postojanje konzerviranih XPA ortologa u kralježnjaka i beskralježnjaka. Zebrica ima jedinstven *xpa* gen koji je pokazao ortologiju jedan-ka-jedan s ljudskim XPA genom. Sintenijska analiza je pokazala da se *xpa* gen kod zebrice nalazi na kromosomu 1 i da je genska okolina relativno očuvana u usporedbi s ljudskim ortologom. *Acrc* se kod zebrice nalazi na kromosomu 14 i pokazuje sintenijski odnos sa ljudskim *ACRC* genom.

Ekspresijska analiza je pokazala da je *acrc* kod zebrice dominatno eksprimiran u testisima i jajnicima i da je eksprimiran više od *sprtn* proteaze tijekom ranog embrionalnog razvoja zebrice. Stoga pretpostavljamo da je ACRC dominantna proteaza koja održava integritet genoma germinativnih stanica i popravlja DPC-eve tijekom embrionalnog razvoja zebrice.

Izolacija DPC-eva RADAR metodom iz embrija zebrice pokazala je aditivni učinak u akumulaciji DPC-eva u odsutnosti NER puta i Sprtn proteaze, što ukazuje na to da NER put *in vivo* popravlja DPC-eve neovisno o SPRTN proteazi. Međutim, ova hipoteza će trebati biti potvrđena dodatnim eksperimentima. Embriji deficijentni za Sprtn i Xpa protein pojačano su akumulirali DPC-eve niske molekularne mase, što sugerira da su i NER put i SPRTN uključeni u njihov popravak *in vivo*. Ljudske stanice s utišanim *XPA*, tretirane s DPC-inducirajućim agensom formaldehidom, imale su značajno povećane razine histonskih DPC-eva, što pokazuje da je NER put ključan za uklanjanje H1 i H3 DPC-eva u ljudskim stanicama.

Acrc mutanti zebrice uspješno su stvoreni u svrhu proučavanja uloge Acrc proteina u popravku DPC-eva. Embriji koji sadrže Acrc enzimsku mutaciju i tretirani su formaldehidom imali su povećane razine ukupnih DPC-eva, što dokazuje da je katalitička funkcija Acrc-a potrebna za popravak DPC-eva i da je Acrc doista aktivna proteaza. Embriji koji nose deleciju C-kraja Acrc proteina također su pojačano akumulirali ukupne DPC-eve, čak i u odsutnosti tretmana, zbog nedostatka funkcionalnog Acrc proteina.

S obziorm na fenotip mortaliteta uočen kod heterozigotnih embrija koji nose enzimsku ili C-terminalnu deleciju, pokazali smo da su enzimska aktivnost i SprT domena Acrc proteina važne tijekom ranog razvoja zebrice. To su dodatno potvrdili eksperimenti spašavanja koji su pokazali da su E451 katalitički glutamat, SprT domena i IDR domena Acrc proteina ključni za rani razvoj kralježnjaka.

9.Curriculum vitae

First name: Christine
Family name: Supina Pavić
Date of birth: January 22th 1993
Place of birth: Zagreb, Croatia
Nationality: Croatian
Present position: Ph.D. student, Laboratory for Molecular Ecotoxicology, Division for Marine and Environmental Research, Ruder Boskovic Institute, Zagreb, Croatia.

Education:

- PhD - Postgraduate Interdisciplinary Doctoral Study of "Molecular Biosciences", joint study of the J. J. Strossmayer University of Osijek, Ruđer Bošković Institute, Zagreb and University of Dubrovnik – in progress
- Master's degree in Molecular biology, Faculty of science, University of Zagreb, 2018
- Bachelors's degree in Molecular biology, Faculty of science, University of Zagreb, 2015

Projects:

- *Deciphering DNA-Protein crosslink repair in vivo using zebrafish model.* Croatian science foundation project (2018. - 2023.), Project leader Marta Popović, PhD.

Platform presentations at scientific conferences:

1. FEBS3+ conference „From molecules to living systems", 2-5 September, 2018, Siófok, Hungary, The role of Nucleotide Excision Repair (NER) pathway in the repair of DNA-protein crosslinks (DPCs)

2. HDIR-5 meeting "Translating Science to Medicine – Targets and Therapeutics", 8-10 November, 2018, Zagreb, Croatia, The role of Nucleotide Excision Repair (NER) pathway in the repair of DNA-protein crosslinks (DPCs)

3. Mediterranean Institute for Life Sciences „Advances in Biomedical Research III", 17-21 June, 2019, Split, Croatia, The role of proteases in the repair of DNA-protein crosslinks

4. The 11th European Zebrafish Meeting, June 29 - July 2, 2020, virtual meeting, The role of ACRC/GCNA in the repair of DNA-protein crosslinks

5. 45th FEBS Congress „Molecules of life: towards new horizons", virtual meeting, July 3 - July 8, 2021, The role of ACRC/GCNA in the repair of DNA-protein crosslinks.

Publications:

1. Stojanović, N., Dekanić, A., Paradžik, M., Majhen, D., Ferenčak, K., Ruščic, J., Bardak, I., Supina, C., Tomičić, M., Christmann, M., Osmak, M. & Ambriović-Ristov, A. (2018) Differential effects of integrin αv knockdown and cilengitide on sensitisation of triple-negative breast

cancer and melanoma cells to microtubule poisons. Molecular pharmacology, 94 (6), 1334-1351.

Supina, C., Otten, C., Anticevic I., Kutnjak M., Popovic, M. (2022) The role of ACRC protease in DNA protein crosslink repair. In preparation.

Supina, C., Otten, C., Anticevic I., Kutnjak M., Popovic, M. (2022) The role of NER pathway in DNA protein crosslink repair *in vivo* using zebrafish model. In preparation.

Memberships:

- Croatian society for biochemistry and molecular biology
- Croatian biophysical society